



**HAL**  
open science

# Analyse métagénomique d'échantillons de carnivores du Pléistocène supérieur et de leur alimentation

Pauline Palacio

► **To cite this version:**

Pauline Palacio. Analyse métagénomique d'échantillons de carnivores du Pléistocène supérieur et de leur alimentation. Biologie moléculaire. Université Paris-Saclay, 2015. Français. NNT : 2015SACLS259 . tel-01666035

**HAL Id: tel-01666035**

**<https://theses.hal.science/tel-01666035>**

Submitted on 18 Dec 2017

**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.

THÈSE DE DOCTORAT  
DE  
L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE A  
L'UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE N°577  
STRUCTURE ET DYNAMIQUE DES SYSTÈMES VIVANTS  
Spécialité : Sciences de la Vie et de la Santé

Par

Pauline PALACIO

ANALYSE MÉTAGÉNOMIQUE D'ÉCHANTILLONS DE  
CARNIVORES DU PLÉISTOCÈNE SUPÉRIEUR  
ET DE LEUR ALIMENTATION

Thèse présentée et soutenue à Gif-sur-Yvette, le 17 décembre 2015

Composition du Jury :

Pierre Capy

Véronique Barriel

Ludovic Orlando

Myriam Boudadi-Maligne

Jean-Marc Elalouf

Professeur (Université Paris-Saclay)

Maître de Conférences (UMR7207, MNHN)

Professeur (Université de Copenhague)

Chargée de Recherche (UMR5199, CNRS)

Directeur de Recherche (CEA)

Président du Jury

Rapporteuse

Rapporteur

Examinatrice

Directeur de thèse







À Bernard Laude,

Ne pas pouvoir partager ces recherches avec toi, mon grand-père, fut une réelle tristesse,



## *Remerciements*

Je tiens tout d'abord à remercier mon directeur de thèse Jean-Marc Elalouf, pour m'avoir accueillie dans son équipe. Vous m'avez formée avec compétence et une grande patience à l'étude de l'ADN ancien qui était pour moi une grande inconnue. Vous m'avez également fait découvrir les merveilles de la Grotte Chauvet : ce fut un moment unique que je n'oublierai jamais. Merci pour ces quatre années durant lesquelles j'ai tant appris.

Je tiens à remercier les membres de mon jury pour accepter de juger mon travail. Merci à Véronique Barriel et Ludovic Orlando d'avoir accepté de juger mon travail en tant que rapporteurs. Je remercie Pierre Capy pour avoir pris le temps d'évaluer mon travail. Je remercie également Myriam Boudadi-Maligne pour avoir accepté d'apporter son expertise de paléontologue à mon travail.

Je tiens à remercier Marie-Claude pour toutes les connaissances en bio-informatique qu'elle m'a apportées et pour le temps qu'elle m'a gentiment consacré. Je regrette de ne pas avoir travaillé plus tôt avec toi.

Merci également à Céline pour les précieux conseils et toute l'aide qu'elle a pu apporter sur le projet Chauvet, sans jamais compter son temps.

Mes remerciements se tournent vers Arnaud Martel pour son aide qui m'a été très précieuse pour tous les scripts informatiques, ainsi que pour sa grande disponibilité et sa gentillesse.

J'adresse un grand remerciement à Michel Philippe et à Frédéric Maksud pour les très nombreux échantillons qu'ils nous ont gracieusement fournis et qui m'ont permis d'effectuer la majorité de mes recherches.

Je tiens à remercier Myriam et JB pour m'avoir accueillie sur le site de Maldidier et m'avoir appris à fouiller (même si, au début, mon carré ressemblait plus à « Verdun » qu'à autre chose ...). Ce fut une expérience inoubliable ! Vos connaissances en paléontologie et nos discussions m'ont également été d'une aide précieuse. Merci pour tout !

Merci également à toute l'équipe de fouilleurs de Maldidier pour l'accueil chaleureux qu'ils m'ont réservé.



Merci à Brigitte pour les nombreux goûters et discussions partagés. Ce fut des moments de détente fort agréables. J'espère que nous aurons l'occasion d'en partager d'autres.

Je tiens à adresser une pensée aux personnes de l'équipe de Badminton du CEA qui m'ont permis de me changer les idées tous les mercredis soirs en me faisant courir sans vergogne d'un bout à l'autre du terrain.

Merci à toutes les personnes du bâtiment 144, Linda, Marie, Guillaume, Rosa, Marie-Hélène, Catherine, Monique, Clément, Guillaume, et Mathieu et tous les autres qui ont rendu ces années encore plus agréables.

Merci à Cristelle pour tous les moments de bonne humeur partagés, au café ou ailleurs. Tu étais un rayon de soleil et d'énergie dans les couloirs du bâtiment. Tu me dois toujours une course de chariot d'ailleurs ... Merci également à sa fille Mégane qui m'a permis d'ajouter l'option esthétique à ma thèse.

Comme je te l'ai promis maintes fois au cours de ces 3 ans, je t'érige une statue, Bruno, pour avoir répondu avec une patience d'ange à tous mes coups de fils désespérés sur la bio-informatique et le reste, quelle que soit l'heure. Je crois bien que sans toi le python m'aurait avalée toute crue ...

Merci à Camille d'avoir été là pour comprendre mes travaux quand je tentais, sans grand succès, de les expliquer en dehors du travail. Merci également pour tes relectures même si tu as refusé de comprendre mon amour des points virgules !

Véronique, te connaître et partager le laboratoire avec toi au cours de ces quatre années a été une aventure formidable. Nos discussions qu'elles aient été légères, sérieuses ou encore scientifiques ont été des grands moments de plaisir. Sans ton soutien de « mère » toujours présente et compréhensive, je ne sais pas où je serais en ce moment ... Roulée en boule sous mon bureau probablement !

Papa, Maman, je ne sais quoi vous dire pour vous remercier de votre soutien sans faille, de vos conseils et de votre patience. Merci Papa d'avoir rendu les choses claires en parfois seulement quelques mots. Merci Maman pour les longues heures de relectures et corrections,

de paragraphes plus ou moins obscurs (les dégradations de l'ADN ancien entre-autres ... ). Vous avez été là pour moi comme vous l'avez toujours été, et jamais je ne pourrai assez vous en remercier.

Enfin, j'adresse un immense merci à la boule d'énergie positive qui partage ma vie. Romain, tu as su, sans tomber dans la compassion aveugle, dédramatiser avec talent mes angoisses et mes coups de blues dans les moments difficiles. Tu as toujours été là pour moi, parfois dur mais toujours vrai. Merci d'être là, merci d'être toi !



## Table des matières

<b>INTRODUCTION GÉNÉRALE .....</b>	<b>20</b>
------------------------------------	-----------

### **PARTIE 1 : INTRODUCTION**

<b><u>1. L'étude de l'ADN ancien .....</u></b>	<b><u>23</u></b>
--	------------------

1.1. Les premières molécules d'ADN ancien .....	23
1.2. La paléogénétique .....	25
1.3. La naissance de la paléogénomique .....	26
1.4. Les techniques de séquençage haut-débit .....	27
1.4.1. La méthode de séquençage haut-débit 454 .....	27
1.4.2. La méthode de séquençage haut-débit Illumina .....	28
1.4.3. Les autres méthodes de séquençage haut-débit .....	31
1.4.4. Apports et limites des méthodes de séquençage haut-débit pour l'analyse des génomes anciens .....	32
1.5. Les problèmes liés à l'ADN ancien .....	33
1.5.1. La dégradation de l'ADN ancien .....	34
1.5.2. Les lésions bloquantes .....	36
1.5.3. Les dommages subis par l'ADN .....	38
1.5.4. Les molécules inhibitrices .....	40
1.5.5. Les risques de contamination .....	41
1.6. Les échantillons étudiés .....	42
1.6.1. Les tissus mous .....	42
1.6.2. Les os .....	43
1.6.3. Les dents .....	44
1.6.4. Les cheveux et les poils .....	45
1.6.5. Les coprolithes .....	46
1.6.6. Le sol .....	47
1.6.7. Les coquilles d'œufs .....	48

<b><u>2. Les carnivores du Pléistocène étudiés .....</u></b>	<b><u>49</u></b>
--	------------------

2.1. Les loups du Pléistocène .....	49
2.1.1. La famille des <i>Canidae</i> .....	49
2.1.2. Histoire évolutive des <i>Canis</i> durant le Pléistocène en Europe .....	51
2.1.3. Mode de vie des <i>Canis lupus</i> .....	57
2.2. L'hyène des cavernes .....	59
2.2.1. Origine évolutive de l'hyène des cavernes .....	60

2.2.2. Morphologie de l'hyène des cavernes .....	61
2.2.3. Mode de vie de l'hyène des cavernes .....	62

### **3. Les grottes d'où sont issus les échantillons analysés ..... 65**

3.1. Les grottes des gorges de l'Ardèche .....	65
3.1.1. La Grotte Chauvet-Pont d'Arc .....	66
3.1.2. La Grotte des Deux-Ouvertures .....	72
3.1.3. La Grotte-aux-Points .....	75
3.1.4. La Grotte-aux-Ours de Châteaubourg .....	77
3.1.5. Les grottes du site de Soyons .....	79
3.1.6. Le Réseau Salomé .....	80
3.2. Les grottes du Lot .....	80
3.2.1. La Grotte-aux-Ours .....	80
3.2.2. La Grotte de Siréjol .....	81
3.3. La Grotte des Trois-Frères en Ariège .....	83
3.3.1. La découverte de la Grotte des Trois-Frères .....	83
3.3.2. Les représentations picturales de la Grotte des Trois-Frères .....	83

## **PARTIE 2 : MATÉRIEL ET MÉTHODES**

1. Prélèvement des échantillons étudiés .....	89
2. Datation radiocarbone des échantillons .....	90
3. Prévention des contaminations.....	90
4. Extraction d'ADN .....	91
5. Réalisation d'amplifications d'ADN ciblées .....	92
6. Sélection des échantillons .....	94
7. Construction de banques de fragments .....	96
8. Reconstitution des génomes mitochondriaux .....	98
8.1. Assemblage des fragments en contigs .....	98
8.2. Alignement des lectures de la banque .....	98
8.3. Réalisation de réactions de PCR ciblées .....	99
9. Recherche de l'ADN mitochondrial d'autres espèces animales dans les coprolithes .....	99
10. Analyses phylogénétiques .....	100

## **PARTIE 3 : ANALYSES DE COPROLITHES DE CANIDÉS DE LA GROTT CHAUVET-PONT D'ARC**

<b><u>1.Introduction .....</u></b>	<b><u>103</u></b>
<b><u>2.Matériel et Méthodes .....</u></b>	<b><u>105</u></b>
2.1. Alignement des lectures sur le génome nucléaire du chien .....	105
2.2. Analyses des catégories de gènes .....	105
2.3. Amplifications ciblées de gènes nucléaires .....	106
<b><u>3. Résultats et discussion .....</u></b>	<b><u>107</u></b>
3.1. Choix de l'échantillon .....	107
3.2. Datation de l'échantillon .....	109
3.3. Les banques de fragments construites pour l'échantillon CCH3 .....	111
3.4. Génome mitochondrial du canidé de la Grotte Chauvet-Pont d'Arc .....	111
3.4.1. Reconstitution et analyse du génome mitochondrial complet .....	111
3.4.2. Position phylogénétique du génome mitochondrial du canidé .....	121
3.5. Analyses du génome nucléaire .....	123
3.5.1. Analyse globale du génome nucléaire .....	123
3.5.2. Analyse des <i>Gene Ontology</i> .....	124
3.5.3. Analyse ciblée de gènes nucléaires .....	125
3.6. Analyse de l'alimentation du canidé de la grotte Chauvet-Pont d'Arc .....	129
<b><u>4. Conclusion .....</u></b>	<b><u>134</u></b>

## **PARTIE 4 : ANALYSES DE COPROLITHES D'HYÈNE DES CAVERNES**

<b><u>1. Introduction .....</u></b>	<b><u>139</u></b>
<b><u>2. Matériel et méthodes .....</u></b>	<b><u>142</u></b>
2.1. Séquençage Illumina de l'ADN extrait de coprolithes d'hyène des cavernes .....	142
2.2. Reconstitution des génomes mitochondriaux d'hyène des cavernes .....	142
2.3. Reconstitution des génomes mitochondriaux de bovinés .....	143
2.3.1. Bison des steppes .....	143
2.3.2. Boviné mis en évidence dans le coprolithe de la Grotte-aux-Ours .....	145
2.4. Analyse par réactions de PCR de l'ADN mitochondrial du <i>Bison schoetensacki</i> .....	146

<b>3. Résultats et discussion .....</b>	<b>146</b>
3.1. Panorama de l'alimentation et des génomes d'hyène des cavernes .....	146
3.1.1. Diversité génétique d'hyène des cavernes au Pléistocène .....	146
3.1.2. Panorama de l'alimentation de l'hyène des cavernes .....	149
3.2. Reconstitution d'un génome complet de <i>Bison priscus</i> .....	150
3.2.1. Choix de l'échantillon .....	150
3.2.2. Datation de l'échantillon SGE2 .....	153
3.2.3. Caractéristiques de la banque de fragments construite pour SGE2 .....	154
3.2.4. Reconstitution d'un génome mitochondrial complet pour <i>Bison priscus</i> .....	156
3.2.5. Position phylogénétique du génome mitochondrial de <i>Bison priscus</i> .....	159
3.3. Reconstitution d'un génome mitochondrial de boviné à partir du coprolithe de la Grotte-aux-Ours .....	160
3.3.1. Analyse des séquences d'ADN d'origine animale présents dans le coprolithe de la Grotte-aux-Ours .....	160
3.3.2. Reconstitution d'un génome mitochondrial pour le boviné de la Grotte-aux-Ours ....	162
3.3.3. Position phylogénétique du spécimen de la Grotte-aux-Ours .....	163
3.4. Analyse du génome mitochondrial de <i>Bison schoetensacki</i> .....	166
3.4.1. Sélection de l'échantillon .....	166
3.4.2. Datation de l'os de <i>Bison schoetensacki</i> .....	167
3.4.3. Analyse de fragments du génome mitochondrial de <i>Bison schoetensacki</i> .....	168
3.4.4. Position phylogénétique du génome mitochondrial partiel de <i>Bison schoetensacki</i> ...	169
<b>4. Conclusion.....</b>	<b>170</b>
<b>CONCLUSION GÉNÉRALE .....</b>	<b>174</b>
<b>BIBLIOGRAPHIE .....</b>	<b>178</b>
<b>ANNEXES .....</b>	<b>196</b>

## Liste des figures

<b>Figure 1.</b> Construction classique d'une banque de fragments d'ADN ancien par clonage ....	26
<b>Figure 2.</b> Principe de la méthode de séquençage haut-débit 454 .....	28
<b>Figure 3.</b> Schéma de la méthode de construction d'une banque métagénomique .....	29
<b>Figure 4.</b> Schéma de l'amplification par pont .....	30
<b>Figure 5.</b> Principe du séquençage en cours de synthèse lors du séquençage Illumina .....	31
<b>Figure 6.</b> Préservation de l'ADN dans des tissus de mammoths conservés dans le permafrost .....	34
<b>Figure 7.</b> Sites des principales lésions conduisant à la coupure des molécules d'ADN ancien .....	35
<b>Figure 8.</b> Courbe de survie d'une molécule d'ADN de taille supérieure à 100 nucléotides en fonction de la température .....	36
<b>Figure 9.</b> Les différents types de liaisons covalentes empêchant la progression de la Taq Polymérase .....	37
<b>Figure 10.</b> Les transitions de type 2 dues aux désaminations .....	39
<b>Figure 11.</b> Mise en évidence de l'augmentation des transitions de type 2 aux extrémités 5' et 3' des molécules d'ADN ancien .....	39
<b>Figure 12.</b> Gel d'électrophorèse montrant les fragments d'ADN d'amplifiés par réaction de PCR pour trois coprolithes d'hyène des cavernes .....	40
<b>Figure 13.</b> Structure d'un os long .....	43
<b>Figure 14.</b> Structure d'une dent .....	44
<b>Figure 15.</b> Structure d'un follicule pileux .....	45
<b>Figure 16.</b> Coprolithes de canidés de la Grotte Chauvet-Pont d'Arc .....	46
<b>Figure 17.</b> Hémi-mandibule d'un spécimen de <i>Canis etruscus</i> .....	52
<b>Figure 18.</b> Récapitulatif des différentes hypothèses pour le lieu et la date de la domestication du loup .....	57
<b>Figure 19.</b> Sculpture probable d'une hyène des cavernes sur un propulseur en ivoire .....	59
<b>Figure 20.</b> Crâne d'hyène des cavernes.....	62
<b>Figure 21.</b> Carte des grottes où ont été prélevés les échantillons .....	65
<b>Figure 22.</b> Topographie de la Grotte Chauvet-pont d'Arc .....	68
<b>Figure 23.</b> Représentations picturales de la Grotte Chauvet-Pont d'Arc .....	69
<b>Figure 24.</b> Les traces du passage de l'Homme dans la Grotte Chauvet-Pont d'Arc .....	71
<b>Figure 25.</b> Les traces du passage des animaux dans la Grotte Chauvet-Pont d'Arc .....	72
<b>Figure 26.</b> Topographie de la Grotte des Deux-Ouvertures .....	74
<b>Figure 27.</b> Représentation d'un bouquetin dans la Grotte des Deux-Ouvertures .....	76
<b>Figure 28.</b> Topographie de la Grotte-aux-Points .....	77
<b>Figure 29.</b> Représentation du panneau des deux-signes de la Grotte-aux-Points .....	78
<b>Figure 30.</b> Plan de la Grotte-aux-Ours de Châteaubourg .....	79
<b>Figure 31.</b> Topographie de la Grotte de Siréjol .....	82
<b>Figure 32.</b> Restes osseux retrouvés dans le grand éboulis de la Grotte de Siréjol .....	83



<b>Figure 33.</b> Topologie de la Grotte des Trois-Frères .....	84
<b>Figure 34.</b> Gels d'électrophorèse montrant les fragments d'ADN amplifiés par PCR d'un fragment du cytochrome b pour trois coprolithes d'hyène des cavernes .....	94
<b>Figure 35.</b> Gels d'électrophorèse montrant les fragments d'ADN amplifiés par PCR sur une gamme de volume pour plusieurs échantillons d'hyènes des cavernes .....	95
<b>Figure 36.</b> Gels d'électrophorèse montrant les banques de fragments construites avec deux volumes d'extraits différents pour deux coprolithes d'hyène des cavernes .....	97
<b>Figure 37.</b> Emplacement des coprolithes de canidés dans la Grotte Chauvet-Pont d'Arc ...	104
<b>Figure 38.</b> Gels d'électrophorèse montrant les fragments d'ADN amplifiés par PCR, à partir des extractions effectuées sur trois coprolithes de canidés .....	107
<b>Figure 39.</b> Gels d'électrophorèse montrant les banques d'ADN construites pour les trois coprolithes de canidés .....	108
<b>Figure 40.</b> Pourcentage d'ADN de canidé dans les banques pour les trois coprolithes .....	109
<b>Figure 41.</b> Fragment du coprolithe de canidé envoyé pour datation .....	110
<b>Figure 42.</b> Nombre de lectures pour chaque position du génome mitochondrial du canidé de la Grotte Chauvet-Pont d'Arc .....	112
<b>Figure 43.</b> Fréquence des mutations observées sur les lectures utilisées pour reconstituer le génome mitochondrial .....	113
<b>Figure 44.</b> Matrice de différences entre le génome mitochondrial du canidé de Chauvet et ceux des chiens et loups actuels .....	114
<b>Figure 45.</b> Position phylogénétique du canidé de la Grotte Chauvet-Pont d'Arc .....	122
<b>Figure 46.</b> Couverture individuelle des 38 autosomes et du chromosome sexuel X .....	123
<b>Figure 47.</b> Récapitulatif des catégories des 80 GO d'intérêt .....	125
<b>Figure 48.</b> Fréquences des haplotypes pour les gènes MGAM et SLC5A1.....	126
<b>Figure 49.</b> Récapitulatif des haplotypes pour les mutations du gène nucléaire MC1R .....	128
<b>Figure 50.</b> Nombre de séquences de la banque qui s'alignent en tolérant un mésappariement sur les génomes de proies potentielles .....	129
<b>Figure 51.</b> Gel d'électrophorèse montrant les fragments d'ADN d'ours des cavernes amplifiés par PCR à partir de l'extrait d'ADN de CCH3 .....	131
<b>Figure 52.</b> Statistiques des endommagements des fragments d'ADN d'ours des cavernes amplifiés à partir de l'extrait du coprolithe .....	132
<b>Figure 53.</b> Os rongés d'ours des cavernes présents dans la Grotte Chauvet-pont d'Arc .....	133
<b>Figure 54.</b> Localisation des sites archéologiques d'où proviennent les coprolithes d'hyène des cavernes étudiés .....	139
<b>Figure 55.</b> Position phylogénétique des génomes mitochondriaux d'hyène des cavernes reconstitués à partir des échantillons des différents sites étudiés .....	147
<b>Figure 56.</b> Coprolithe d'hyène des cavernes de la Grotte-aux-Ours .....	150
<b>Figure 57.</b> Position des échantillons prélevés dans la Grotte des Trois-Frères .....	151
<b>Figure 58.</b> Pourcentage des lectures de <i>Bos</i> et de <i>Bison</i> dans les séquences amplifiées à partir des extraits réalisés pour les 4 ossements de <i>Bison priscus</i> .....	152
<b>Figure 59.</b> Datation de l'ossement de <i>Bison priscus</i> .....	154
<b>Figure 60.</b> Caractéristiques des lectures utilisées pour reconstituer le génome mitochondrial de <i>Bison priscus</i> .....	157

<b>Figure 61.</b> Fréquence des mutations observées sur les séquences utilisées pour reconstituer le génome mitochondrial de <i>Bison priscus</i> .....	158
<b>Figure 62.</b> Position phylogénétique du génome mitochondrial de <i>Bison priscus</i> .....	159
<b>Figure 63.</b> Nombre de séquences des 4 pistes de la banque présentant un alignement parfait sur un échantillonnage de 11 génomes mitochondriaux de proies potentielles .....	160
<b>Figure 64.</b> Corrélation entre le taux de couverture et le pourcentage en GC pour la séquence mitochondriale du boviné de la Grotte-aux-Ours .....	163
<b>Figure 65.</b> Position phylogénétique des génomes mitochondriaux de <i>Bison priscus</i> et du boviné de la Grotte-aux-Ours .....	164
<b>Figure 66.</b> Nombre de séquences des 4 pistes de la banque qui présentent un alignement parfait sur les génomes mitochondriaux de trois bovidés .....	165
<b>Figure 67.</b> Localisation de la Grotte-aux-Ours et de la Grotte de Siréjol.....	166
<b>Figure 68.</b> Os canon de <i>Bison schoetensacki</i> sélectionné pour les analyses .....	167
<b>Figure 69.</b> Matrice de différences entre le génome mitochondrial reconstitué à partir de l'ossement de <i>Bison schoetensacki</i> et ceux de cinq bovidés.....	169
<b>Figure 70.</b> Position phylogénétique de <i>Bison priscus</i> et des deux génomes de <i>Bison schoetensacki</i> .....	170

## Liste des tableaux

<b>Tableau 1.</b> Échantillons collectés pour chaque site .....	89
<b>Tableau 2.</b> Liste des amorces utilisées pour amplifier l'ADN contenu dans les différents échantillons .....	94
<b>Tableau 3.</b> Liste des échantillons les plus riches en ADN pour chaque site .....	96
<b>Tableau 4.</b> Noms des extraits utilisés pour la construction des banques .....	96
<b>Tableau 5.</b> Génomes utilisés pour aligner les lectures des banques .....	98
<b>Tableau 6.</b> Nombre de lectures alignées sur le génome nucléaire du chien .....	105
<b>Tableau 7.</b> Liste des amorces utilisées pour réaliser les amplifications ciblant des gènes nucléaires .....	106
<b>Tableau 8.</b> Datation radiocarbone du coprolithe de la Grotte Chauvet-Pont d'Arc .....	110
<b>Tableau 9.</b> Nombre de lectures obtenues pour chaque banque construite à partir du coprolithe de canidé .....	111
<b>Tableau 10.</b> Génomes mitochondriaux utilisés pour la construction de la matrice de différences .....	114
<b>Tableau 11.</b> Nombre de lectures de la banque qui s'alignent en tolérant un mésappariement sur 50 génomes mitochondriaux de proies potentielles .....	130
<b>Tableau 12.</b> Nombre de lectures des banques construites pour chaque coprolithe d'hyène des cavernes .....	142
<b>Tableau 13.</b> Nombre de bases couvertes et taux de couverture pour les génomes mitochondriaux d'hyène des cavernes reconstitués à partir des coprolithes des différents sites étudiés .....	143
<b>Tableau 14.</b> Liste des amorces utilisées pour amplifier l'ADN de l'échantillon de <i>Bison priscus</i> .....	144
<b>Tableau 15.</b> Liste des amorces utilisées pour amplifier l'ADN du coprolithe de la Grotte-aux-Ours .....	145
<b>Tableau 16.</b> Liste des amorces utilisées pour amplifier l'ADN de l'ossement de <i>Bison schoetensacki</i> .....	146
<b>Tableau 17.</b> Datation radiocarbone des différents coprolithes d'hyène des cavernes .....	148
<b>Tableau 18.</b> Séquences animales présentes dans les coprolithes d'hyène des cavernes .....	149
<b>Tableau 19.</b> Nombre de lectures de la banque qui s'alignent sur des génomes mitochondriaux de référence de bovidés .....	155
<b>Tableau 20.</b> Nombre de lectures de la banque qui s'alignent parfaitement sur 49 génomes mitochondriaux de proies potentielles .....	161
<b>Tableau 21.</b> Datation de l'os canon de <i>Bison schoetensacki</i> .....	168

## *Index*

**ADN** : Acide désoxyribonucléique

**AMS** : Accelerator Mass Spectrometer

**BP** : Before Present (avant 1950)

**BSA** : Bovine Serum Albumine

**Caméra CCD** : Camera comprenant un dispositif à transfert de charge

**EDTA** : Acide Ethylène Diamine Tétracétique

**Guilde** : ensemble d'un groupe

**Ma** : Million d'années

**N-PTB** : N-Phénacylthiazolium

**Oligocène** : époque géologique allant de - 33,9 à - 23 Ma

**pb** : paire de bases

**PCR** : Polymerase Chain Reaction

**Pléistocène** : époque géologique allant de - 2,6 Ma à - 12 000 ans

**Pliocène** : époque géologique allant de - 5,3 à - 2,6 Ma

**Villafranchien** : subdivision de l'échelle des temps géologiques allant de - 5,2 à - 1,2 Ma



## Introduction générale

Le bestiaire qui orne les parois des cavernes est le dépositaire de l'attention des Hommes du Pléistocène pour les animaux qui les entouraient. Ces animaux, pour la plupart éteints actuellement, ont longtemps été étudiés uniquement à travers leurs restes osseux. Ces recherches bien qu'apportant des réponses sur le plan anatomique ne permettaient pas d'obtenir des informations complètes sur l'éthologie et l'écologie de ces espèces disparues. De plus, sans information d'ordre génétique, certaines phylogénies demeuraient incomplètes. Les prémices de l'étude de l'ADN ancien dans ces restes osseux ont permis d'obtenir de nouvelles informations sur des espèces éteintes et montrer l'intérêt de s'intéresser d'un point de vue génétique aux échantillons étudiés. Aujourd'hui, la construction de banques de fragments et le séquençage haut-débit de celles-ci semblent de mise pour obtenir de nombreuses informations génétiques sur les espèces éteintes. Le régime alimentaire - si les échantillons étudiés sont des coprolithes - et la phylogénie de certaines familles peuvent être établis grâce à ces méthodes d'analyse.

L'objectif de cette thèse est de caractériser le génome, la position phylogénétique et l'alimentation de deux espèces dont on trouve des traces dans les grottes : le loup, *Canis lupus* et l'hyène des cavernes, *Crocota crocuta spelaea*.

Ces études ont été menées grâce à l'analyse de l'ADN contenu dans des échantillons archéologiques - ossements et coprolithes - appartenant à ces deux espèces.

Dans la première partie, nous allons nous intéresser à la famille des canidés grâce aux excréments fossilisés (coprolithes) prélevés dans la partie profonde de la Grotte Chauvet-Pont d'Arc (Ardèche, France). Ils seront analysés au moyen du séquençage haut-débit afin d'obtenir des informations génétiques sur les génomes nucléaire et mitochondrial ainsi que sur le régime alimentaire du spécimen. L'animal, qui les a produits, était-il venu en compagnie des Hommes qui ont orné la caverne ou représentait-il pour eux un prédateur ?

Dans la deuxième partie, nous allons nous intéresser à de nombreux coprolithes d'hyène des cavernes provenant de plusieurs sites de trois régions de France, dans le but d'obtenir un panorama représentatif de l'alimentation de cette espèce au Pléistocène. Nous allons plus particulièrement porter notre attention sur l'un d'eux qui contient des séquences d'un spécimen inconnu de bovidé. D'autres échantillons vont être analysés afin de recueillir les

informations nécessaires à l'identification de l'espèce présente dans le coprolithe mais également pour apporter des éclaircissements à la phylogénie des bovidés.

*Partie 1 :*

*Introduction*



# 1. L'ÉTUDE DE L'ADN ANCIEN

Discipline dérivée de la paléogénétique, la paléogénomique est l'analyse des génomes anciens présents dans des échantillons pré-, proto-, ou simplement historiques grâce aux techniques de séquençage haut-débit.

## 1.1. Les premières molécules d'ADN ancien

Les premières molécules d'ADN ancien obtenues l'ont été grâce à des techniques de clonage dans des bactériophages ou des plasmides.

En 1984, l'équipe d'A. Wilson parvient à obtenir la toute première séquence d'ADN ancien à partir d'un muscle d'un spécimen naturalisé de quagga (sous-espèce de zèbre aujourd'hui disparue). Ils réussissent à obtenir des séquences d'environ 100 paires de bases après clonage dans un bactériophage (Higuchi *et al.*, 1984).

Quelques mois plus tard en 1985, S. Pääbo annonce avoir obtenu la première séquence d'ADN ancien pour un spécimen humain (Pääbo, 1985). En effet il est parvenu, à partir d'une momie d'enfant égyptien, à cloner un fragment d'ADN de 3 400 paires de bases dont 919 ont été séquencées. Ce séquençage va permettre d'identifier deux répétitions Alu qui appartiennent au système HLA humain. Pourtant cette séquence sera rapidement contestée et considérée comme une contamination par de l'ADN moderne ; en effet le fragment cloné est très proche d'une lignée cellulaire utilisée dans le laboratoire (Delpozso & Guardiola, 1989).

La première véritable molécule d'ADN ancien humain va être obtenue grâce à l'invention de la réaction en chaîne par polymérase (PCR) en 1986 (Mullis *et al.*, 1986). Cette nouvelle technique permet d'amplifier de courts fragments d'ADN présents en faible quantité dans des échantillons. De plus, grâce à la spécificité des amorces, les régions du génome analysées peuvent être choisies. Ainsi grâce à la PCR, une séquence d'ADN mitochondrial a pu être obtenue à partir d'un cerveau amérindien retrouvé dans une tourbière de Little Salt Spring en Floride et daté par la méthode du radiocarbone à environ 7 000 ans (Pääbo *et al.*, 1988). Après séquençage, il s'est avéré que le fragment obtenu présentait des mutations et ne se rattachait à aucun haplotype connu. Celui-ci a, depuis lors, été retrouvé chez de nombreux amérindiens.

La réussite dans l'étude menée sur le quagga va encourager les scientifiques à s'intéresser à d'autres espèces disparues telle que celle du loup de Tasmanie, *Thylacinus cynocephalus*. L'analyse de 219 paires de bases d'ADN mitochondrial de spécimens de musée va permettre de montrer que cette espèce est plus proche génétiquement des marsupiaux australiens que de ceux d'Amérique latine (Thomas *et al.*, 1989) ; ce que les analyses morphologiques n'avaient pas pu démontrer. Cette recherche a permis de montrer l'avantage de combiner analyse de l'ADN ancien et paléontologie.

La première étude menée sur les végétaux a porté sur des feuilles fossilisées de magnolia retrouvées dans des dépôts d'argile datés du Miocène (-20 à -17 Ma). Dans ces expériences, publiées en 1990, 820 paires de base de l'ADN des chloroplastes ont été analysées. L'étude suggérait que les magnolias du Miocène étaient proches des actuels d'un point de vue phylogénétique (Golenberg *et al.*, 1990). Mais, à l'image de l'analyse menée sur les momies égyptiennes, les séquences d'ADN issues de ces expériences ont été mises en doute, notamment à cause de l'absence de contrôle négatif pour les PCR et l'impossibilité pour les autres équipes de reproduire les résultats (Wayne *et al.*, 1999).

Cependant, la possibilité que l'ADN puisse se conserver plus d'un million d'années a poussé les chercheurs à tenter d'amplifier de l'ADN contenu dans des échantillons très anciens. Des études ont été réalisées sur des fossiles de dinosaures, des végétaux fossilisés dans l'ambre (Poinar *et al.*, 1993), des insectes-eux aussi emprisonnés dans ce matériau- (DeSalle *et al.*, 1992), ainsi que sur les bactéries contenues dans le tube digestif de ces derniers (Cano *et al.*, 1994). Mais ces expériences ont rapidement été discréditées. Les autres équipes les ont tout d'abord mises en doute à cause de leur non reproductibilité (Austin *et al.*, 1997), puis une étude réalisée en 1998 a montré que la chitine (présente chez les insectes) ainsi que la lignine et la cellulose (pour les plantes) n'étaient pas conservées dans les échantillons étudiés. Il paraissait très étrange que l'ADN se soit conservé dans des échantillons où les autres macromolécules avaient, quant à elles, disparues (Stankiewicz *et al.*, 1998). Devant toutes ces incertitudes, les résultats observés ont donc été attribués à des contaminations. Ces diverses erreurs ont jeté un grand discrédit sur l'étude de l'ADN ancien (MacHugh *et al.*, 2000). Afin de réhabiliter cette discipline, il a fallu établir des protocoles très stricts pour s'assurer que les séquences d'ADN amplifiées correspondaient bien à l'ADN endogène de l'échantillon. La crédibilité de la paléogénétique nécessitait cette remise en question.

## 1.2. La paléogénétique

C'est la mise en place de protocoles rigoureux tels que des espaces dédiés dans les laboratoires, la reproductibilité des résultats et la présence de témoins négatifs qui ont permis de rendre sa crédibilité à l'étude de l'ADN ancien (Cooper & Poinar, 2000). De plus, le séquençage de multiples clones pour chaque échantillon a permis de montrer les altérations caractéristiques des molécules d'ADN ancien (qui seront explicitées dans un paragraphe ultérieur). Un autre critère important, qui devrait être rendue systématique, est la datation de l'échantillon par la méthode radiocarbone qui permet d'apporter la certitude de l'authenticité des restes étudiés.

De nombreuses informations ont pu être obtenues grâce à l'étude paléogénétique d'échantillons anciens. Ainsi, à partir du milieu des années 90, l'analyse de l'ADN contenu dans des restes appartenant à des espèces disparues, a permis d'obtenir les premières séquences génétiques considérées comme authentiques pour le mammouth (Hagelberg *et al.*, 1994 ; Hoss *et al.*, 1994), l'hyène des cavernes (Rohland *et al.*, 2005), le mégacéros (Hughes *et al.*, 2006), l'ours des cavernes (Hänni *et al.*, 1994), le lion des cavernes (Barnett *et al.*, 2009), le bison des steppes (Shapiro *et al.*, 2004), le rhinocéros laineux (Willerslev *et al.*, 2009), ainsi que l'Homme de Néandertal (Kriings *et al.*, 1997). Toutes ces études ont permis de construire une phylogénie moléculaire pour les différentes espèces citées.

La paléogénétique apporte aussi des informations dans l'étude génétique des populations. Ainsi, lorsqu'il y a suffisamment d'échantillons pour une espèce donnée, il est possible d'observer la diversité génétique de celle-ci, comme cela a été fait pour des espèces animales du Pléistocène ou de l'Holocène (Hofreiter *et al.*, 2004).

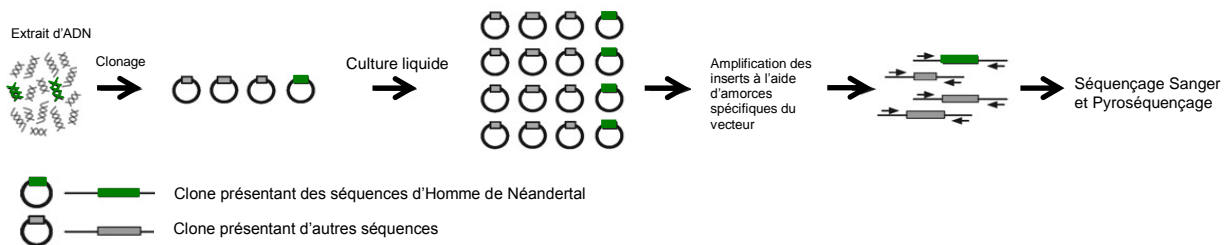
Cette discipline permet de répondre aux questions soulevées par l'extinction de certaines espèces, telles que l'ours des cavernes (Stiller *et al.*, 2010 ; Bon *et al.*, 2011) ou le mammouth laineux (Debruyne *et al.*, 2008 ; Gilbert *et al.*, 2008a) ; et de connaître le moment de la séparation des différentes lignées pendant les maxima glaciaires comme, pour l'ours brun au Pléistocène (Valdiosera *et al.*, 2007 ; Valdiosera *et al.*, 2008).

### 1.3. La naissance de la paléogénomique

Ce sont deux études, publiées à très peu de temps d'intervalle par la même équipe de recherche et reposant sur la construction de banques de fragments, qui ont marqué les débuts de la paléogénomique.

Le but de la première étude était de caractériser un grand nombre de nucléotides du génome nucléaire de l'ours des cavernes. Pour cela, à partir d'un os et d'une dent d'*Ursus spelaeus*, datés d'environ 40 000 ans, deux banques métagénomiques (une pour chaque échantillon) ont été construites par clonage des fragments dans *Escherichia coli* (Noonan *et al.*, 2005). Environ 26 800 paires de bases ont été obtenues grâce à 15 000 clones séquencés par la méthode de Sanger. Cette étude a montré que la quantité d'ADN endogène aux échantillons était très faible par rapport aux ADN contaminants. En effet, l'ADN d'ours des cavernes ne représentait respectivement que 1,1 % et 5,5 % des lectures contenues dans les banques.

La deuxième étude portait sur l'ADN extrait d'un ossement d'Homme de Néandertal (Noonan *et al.*, 2006) ; la construction d'une banque de fragments puis l'analyse par séquençage Sanger et pyroséquençage de ceux-ci a permis de caractériser près de 65 500 paires de bases du génome nucléaire de cette espèce (Figure 1).



**Figure 1.** Construction classique d'une banque de fragments d'ADN ancien par clonage (d'après Noonan *et al.*, 2006).

Ces deux études ont montré qu'analyser sans sélection les fragments d'ADN contenus dans des échantillons archéologiques permettait d'obtenir des informations génétiques pour des espèces éteintes. Des études de paléogénomique ont été initiées en même temps, mais en utilisant les technique de séquençage haut-débit.

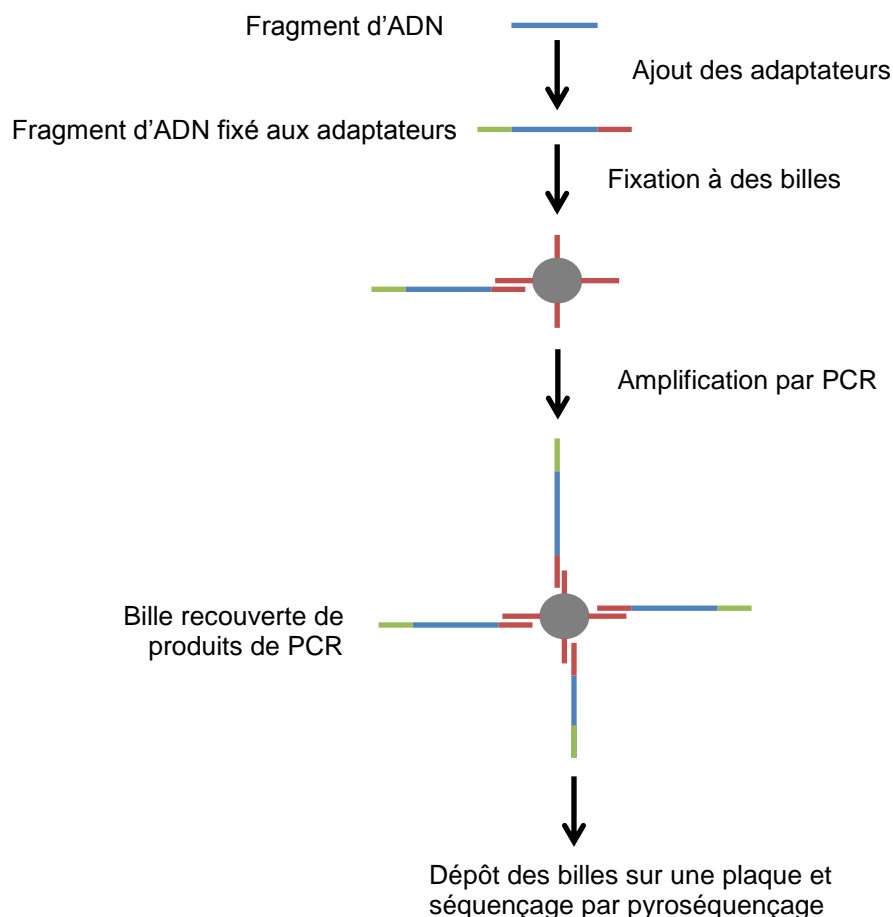
## 1.4. Les techniques de séquençage haut débit

Actuellement, les études paléogénomiques se basent sur deux approches : le séquençage métagénomique de bibliothèques natives et le séquençage de bibliothèque enrichies pour l'ADN de l'organisme d'intérêt. De ce fait, le développement de ces techniques par Roche Applied Science (454 Genome Sequence FLX Instrument), Illumina (Solexa Genome Analyzer), Applied Biosystems (ABI SOLiD system), Life Technologies (Ion Torrent) et Helicos Bioscience (Heliscope) a grandement facilité les analyses paléogénomiques. Les méthodes 454 et Illumina sont les plus couramment utilisées.

### 1.4.1. La méthode de séquençage haut débit 454

Cette première méthode permet de séquencer en parallèle un million de fragments d'ADN d'une longueur de 400 paires de bases en moyenne ; on peut ainsi obtenir un milliard de nucléotides en 24 heures. Cette technique n'a plus qu'un intérêt historique car sa capacité n'atteint pas celle de la méthode Illumina.

La première étape de cette technique consiste à ligaturer de fragments d'ADN de petite taille sur des adaptateurs spécifiques (Margulies *et al.*, 2005). Puis par l'intermédiaire d'amorces complémentaires de ces adaptateurs, les fragments sont fixés sur des billes de streptavidine. Les molécules d'ADN sont ensuite amplifiées par une PCR à émulsion qui va permettre de recouvrir la bille d'une grande quantité de produits de PCR. Après amplification, les billes sont déposées sur une puce dans laquelle sont creusés des puits prévus pour ne recevoir chacun qu'une seule bille. Le séquençage des fragments d'ADN se fait par pyroséquençage : chaque type de nucléotide est ajouté un par un, si c'est celui attendu, il est incorporé ce qui libère un pyrophosphate. Une ATP sulfurylase transforme ce pyrophosphate en ATP, qui va être utilisé à son tour par une luciférase. Cette réaction induit la production d'oxyluciférine et d'un signal lumineux qui va être enregistré par une caméra CCD grâce aux fibres optiques qui sont reliées à chacun des puits (Figure 2).

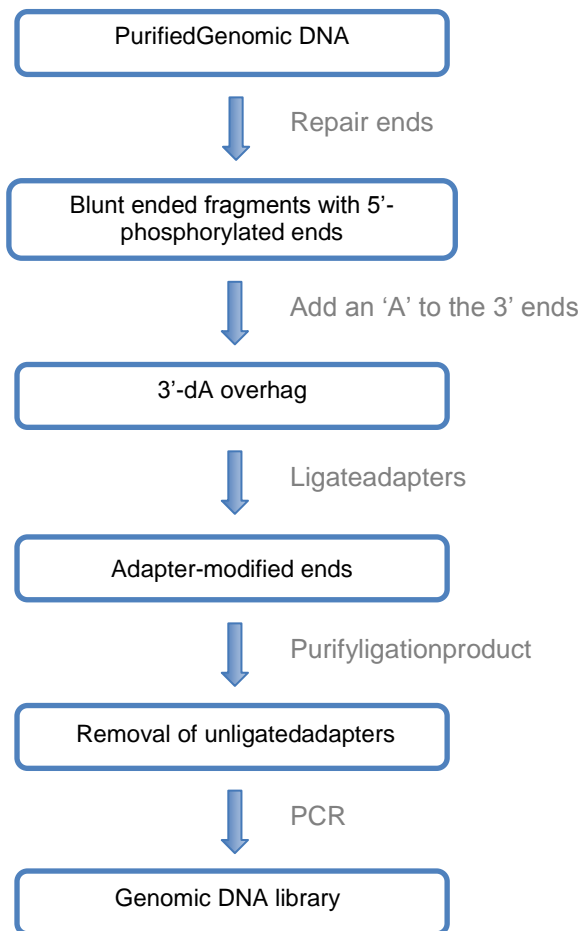


**Figure 2.** Principe de la méthode de séquençage à haut-débit 454 pour un fragment d'ADN.

#### 1.4.2. La méthode de séquençage à haut débit Illumina

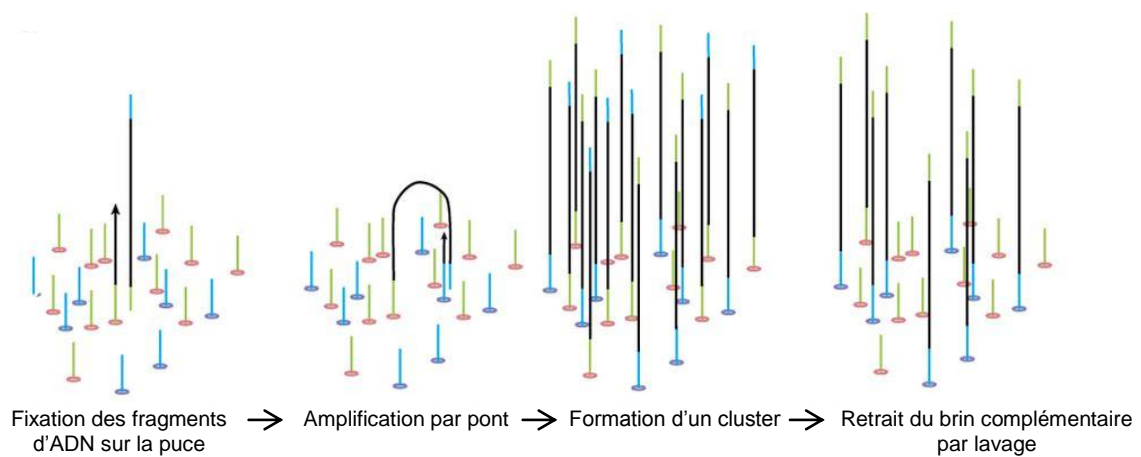
La deuxième méthode a été mise au point par la société Solexa, rachetée depuis par Illumina (Bentley *et al.*, 2008). Elle permet de séquencer un grand nombre de fragments d'ADN, plus d'un milliard par puce. La profondeur des lectures est toutefois relativement faible, avec en routine 150 paires de bases lues.

C'est cette méthode qui a été utilisée pour séquencer les banques construites au cours de cette thèse car malgré une profondeur de lectures relativement courte, elle permet d'obtenir un nombre très important de lectures. Comme pour la méthode 454, cette technique repose sur le séquençage de banques de fragments. Les grandes lignes de la construction des banques sont décrites dans la figure 3.



**Figure 3.** Schéma de la méthode de construction d'une banque métagénomique (d'après Illumina protocols).

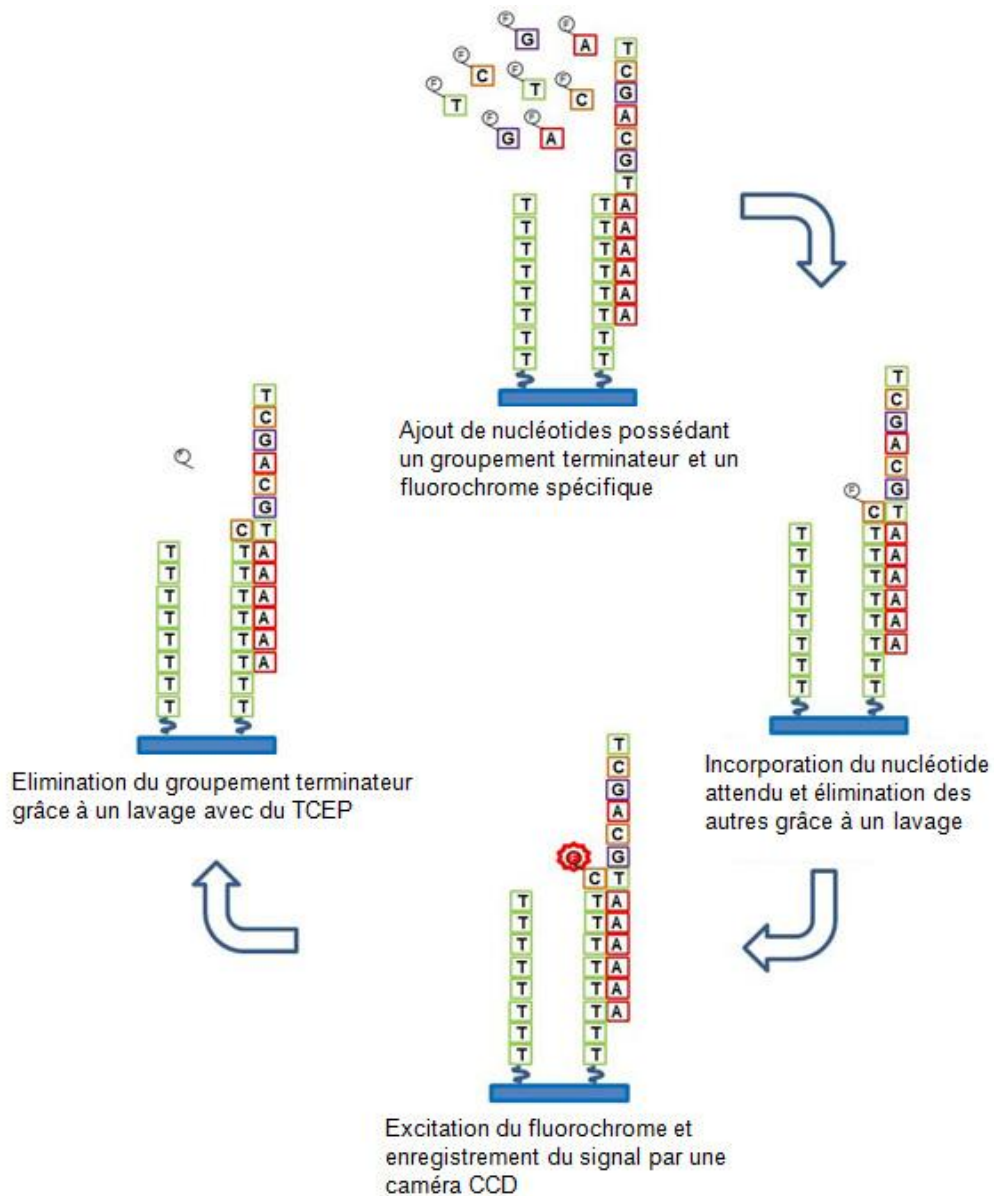
Après amplification des fragments d'ADN, les séquences de la banque vont être fixées sur une puce qui comporte une grande quantité d'amorces complémentaires à ces adaptateurs. Une amplification PCR par pont permet d'obtenir environ un millier de copies « simple brin » (aussi appelé cluster) de ces fragments d'ADN (Figure 4).



**Figure 4.** Schéma montrant l'amplification par pont sur la puce Illumina avant l'étape de séquençage en cours de synthèse.

Une fois formés, ces clusters de fragments identiques sont analysés par séquençage en cours de synthèse. Celui-ci fonctionne par la répétition de cycles identiques. Tout d'abord, quatre types de nucléotides terminateurs réversibles - les déoxynucléotides tri-phosphate possédant un groupement O-azidométhyle à leur extrémité 3' - sont incorporés ; ce groupement terminateur empêche que plusieurs nucléotides soient incorporés sur un même fragment au cours d'un même cycle. Un fluorochrome spécifique est également fixé à chacune des bases. Une fois le nucléotide incorporé au brin d'ADN, un signal laser est émis ; la fluorescence provoquée par cette excitation est enregistrée via une caméra CCD et permet d'identifier la position et la nature de ce nucléotide. Puis, le groupement chimique responsable de la fonction terminatrice est éliminé grâce à un lavage avec du tris(2-carboxyethyl)phosphine (TCEP), pour qu'un nouveau cycle puisse commencer. Les cycles s'enchaînent ensuite jusqu'à la profondeur de lecture sélectionnée pour le projet, généralement 75 à 101 paires de bases pour l'ADN ancien (Figure 5).





**Figure 5.** Principe du séquençage en cours de synthèse lors du séquençage Illumina pour un cycle.

### 1.4.3. Les autres techniques de séquençage haut débit

La méthode ABI SOLiD débute par une étape identique à celle utilisée par la méthode 454. Les fragments d'ADN de la banque sont ligaturés à des billes, amplifiés par PCR à émulsion, puis déposés sur un support. Par contre, le séquençage de ces fragments est très différent. En effet, après fixation de l'amorce à l'adaptateur, des oligonucléotides de 8 paires de bases sont hybridés au fragment d'ADN puis ligaturés. Ces octamères sont caractérisés à leur extrémité par un marqueur fluorescent, dont la couleur dépend de la nature d'un doublet de nucléotides. Après détection de la fluorescence, l'octamère est coupé après le dernier nucléotide du doublet ; le cycle d'hybridation et de ligature recommence jusqu'à

recouvrement complet du fragment d'ADN. Puis un nouveau cycle de séquençage démarre alors en utilisant une amorce plus courte d'un nucléotide, ce qui permet de se décaler dans les doublets incorporés. Le processus est répété autant de fois que nécessaire pour déterminer la séquence complète du fragment d'ADN.

La méthode HeliScope, elle, s'affranchit des étapes d'amplification en travaillant sur une molécule unique pour chaque fragment d'ADN de la banque métagénomique (Thompson & Steinmann, 2010). Pour cela, il faut que la caméra soit suffisamment sensible pour enregistrer un signal émis par une seule molécule. Les fragments d'ADN sont fixés à une plaque de verre, sur laquelle sont ajoutés les réactifs de séquençage. Le séquençage est effectué en cours de synthèse comme pour la méthode Illumina, avec cependant un seul type de nucléotide ajouté à chaque cycle.

La méthode Ion torrent, développée par Ion Torrent Systems Inc. en 2010, se base sur le changement du pH entrainé par la libération d'un ion hydrogène et d'un pyrophosphate lors de l'incorporation d'un nucléotide dans un brin d'ADN. Les fragments d'ADN à séquencer sont placés en plusieurs copies simple brin dans les micropuits d'une puce semi-conductrice. Chaque type de dNTP est ajouté tour à tour dans les puits avec de l'ADN polymérase. S'il s'agit du nucléotide attendu, son incorporation va changer le pH dû à la libération de l'ion hydrogène, ce qui va être observé par le système IFSET. Une fois le signal enregistré, les nucléotides non incorporés sont enlevés grâce à un lavage et un nouveau cycle recommence jusqu'à séquençage complet des fragments d'ADN.

#### **1.4.4. Apports et limites des méthodes de séquençage haut débit pour l'analyse des génomes anciens**

Les techniques de séquençage à haut débit sont bien adaptées à l'étude de l'ADN ancien. En effet, elles permettent le séquençage de nombreux fragments de petite taille (entre 50 et 200 paires de bases) caractéristiques de l'ADN ancien. De plus, toutes les séquences présentes dans l'échantillon vont être séquencées sans sélection ni a priori. Cela permet également d'éviter qu'une contamination ne fausse totalement un résultat comme lors des PCR ciblées. Le séquençage par la méthode 454 a permis des analyses poussées sur le génome nucléaire de l'Homme de Néandertal, avec l'obtention d'un génome mitochondrial complet (Green *et al.*, 2008) et plus de 4 millions de nucléotides pour son génome nucléaire

(Green *et al.*, 2010). Le séquençage par la méthode 454 des fragments d'ADN contenus dans des poils de mammoths (Gilbert *et al.*, 2007) et des cheveux paléo-esquimaux (Gilbert *et al.*, 2008b) a permis d'obtenir des informations sur leurs génomes respectifs. Grâce à la technique de séquençage Illumina, des génomes mitochondriaux d'animaux éteints ont pu être reconstitués tel que celui de l'hyène des cavernes (Bon *et al.*, 2012), celui de l'aurochs (Edwards *et al.*, 2010) ou encore celui du rhinocéros laineux (Willerslev *et al.*, 2009). Des informations sur les génomes nucléaires de paléo-esquimaux (Rasmussen *et al.*, 2010) et de l'Homme de Denisova (Reich *et al.*, 2010) ont également pu être obtenues grâce au séquençage Illumina de banques de fragments. De plus, 115,9 Mb du génome nucléaire d'un cheval ont pu être séquencées grâce à la méthode HeliScope à partir d'un ossement issu du permafrost (Orlando *et al.*, 2013).

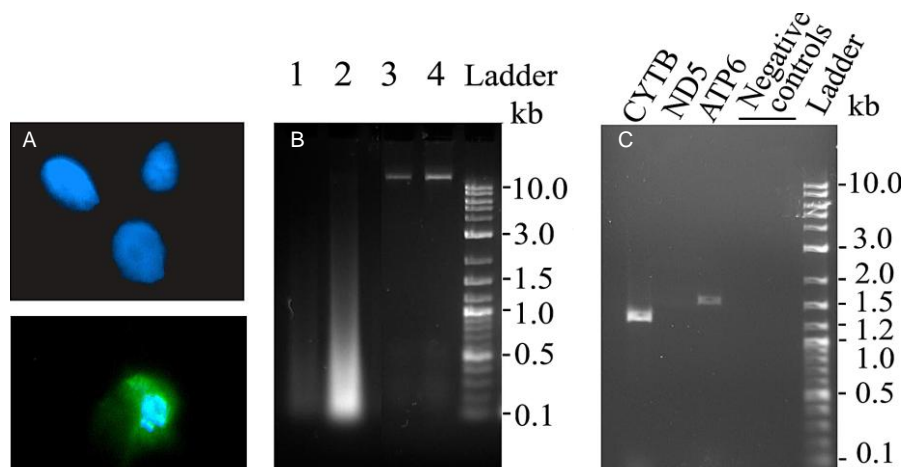
L'utilisation de ces techniques de séquençage à haut débit pose toutefois quelques problèmes. En effet, des études ont montré que la quantité d'ADN endogène dans les banques varie selon les échantillons. Dans les poils et tissus congelés de mammoths retrouvés dans le permafrost, elle est comprise entre 45,4 % et 90 % (Poinar *et al.*, 2006 ; Miller *et al.*, 2008) ; en revanche, elle est plus faible dans les os où elle varie de 0,27 % dans un os de chèvre des Baléares (Ramirez *et al.*, 2009) à 5 % dans un os d'Homme de Néandertal (Green *et al.*, 2010). La faible quantité d'ADN dans certains échantillons rend ces méthodes sensibles aux contaminations, notamment si la construction des banques n'est pas réalisée dans des conditions rigoureuses. Ainsi, la première banque métagénomique construite à partir d'un os d'Homme de Néandertal était contaminée par de l'ADN humain moderne et entre 40 et 78 % des lectures attribuées à *Homo neanderthalis* appartenaient en fait à *Homo sapiens sapiens* (Wall & Kim, 2007 ; Green *et al.*, 2008).

## **1.5. Les problèmes liés à l'ADN ancien**

Les limites de la paléogénomique sont les mêmes que celles qui étaient connues pour la paléogénétique. En effet, malgré une amélioration des techniques de séquençage, les facteurs limitant sont toujours les mêmes : ceux liés à l'ADN ancien. Dans ce chapitre, nous faisons une présentation des altérations de l'ADN ancien qu'elles aient été détectées par séquençage haut débit ou par PCR.

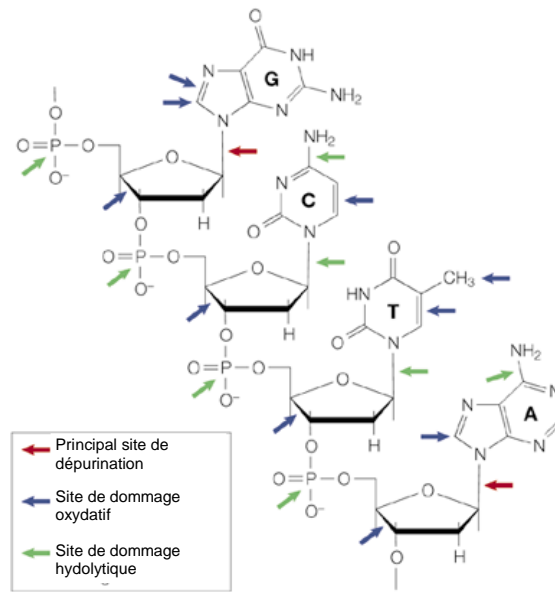
### 1.5.1. La dégradation de l'ADN ancien

L'ADN ancien, même lorsque les échantillons sont dans un état de préservation exceptionnelle - comme le sont ceux issus du Permafrost - est retrouvé dégradé en fragments de petite taille. Ainsi, dans un muscle de mammouth, si bien conservé que les noyaux des cellules étaient encore visibles (Figure 6), les plus longs fragments d'ADN retrouvés n'excédaient pas 1 600 paires de bases (Rogaev *et al.*, 2006).



**Figure 6.** Préservation de l'ADN dans des tissus de mammouths conservés dans le permafrost. A : Coloration montrant l'ADN des cellules au DAPI (en bleu). B : Gel d'acrylamide montrant la taille des molécules d'ADN conservés dans les cellules (1 et 2) et dans un extrait de sang humain (3 et 4). Les pistes 1 et 3 correspondent à une dilution au dixième de l'extrait déposé dans les pistes 2 et 4. C : Gel d'acrylamide montrant l'amplification de longues séquences d'ADN pour des gènes mitochondriaux entiers de mammouth.

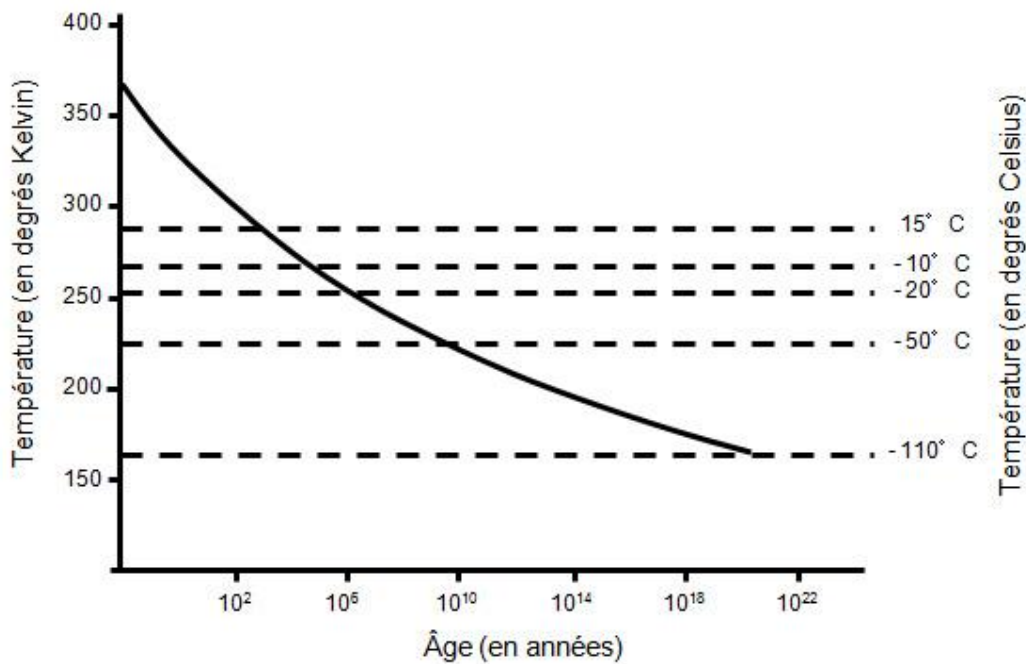
Une des causes de la dégradation de l'ADN est la coupure des molécules sous l'action de plusieurs facteurs. Ces coupures peuvent être induites par la libération des nucléases endogènes peu de temps après le décès de l'organisme ; en effet, les parois des organites où sont localisées les nucléases, se rompent lors de la mort, les laissant libres d'aller dégrader les molécules d'ADN (Lindhal, 1993). Les coupures de l'ADN peuvent également être provoquées par des réactions d'hydrolyse et d'oxydation. Celles-ci entraînent des coupures « simple brin » en rompant les liaisons phosphodiester entre les nucléotides (Pääbo *et al.*, 2004 ; Willerslev & Cooper, 2005). Ces réactions peuvent aussi induire la rupture de la liaison N-glycosidique entre la base azotée et son sucre, créant de ce fait des sites apyrimidiques et apuriques (Figure 7). L'instabilité de ces sites abasiques va provoquer des coupures « double brin ». Il est à noter que les purines sont beaucoup plus touchées que les pyrimidines (Lindhal, 1993).



**Figure 7.** Sites des principales lésions conduisant à la coupure des molécules d'ADN ancien, d'après (Hofreiter et al., 2001).

Les molécules d'ADN peuvent également être dégradées sous l'action de microorganismes (tels que des champignons et des bactéries) ou encore d'invertébrés du sol, qui se nourrissent de macromolécules (Pääbo *et al.*, 2004).

La température de l'environnement dans lequel est conservé l'échantillon joue un grand rôle dans la préservation des molécules d'ADN (Figure 8). En effet, la dégradation induite par les nucléases et les microorganismes va être limitée par des températures basses, qui inhibent l'action des premières et ralentissant l'activité des seconds. Ainsi, les milieux où les températures sont inférieures à 0°C, comme le permafrost, vont permettre une conservation de l'ADN pendant plusieurs centaines de milliers d'années contrairement aux climats chauds (Willerslev *et al.*, 2004 ; Orlando *et al.*, 2013).



**Figure 8.** Courbe de survie d’une molécule d’ADN de taille supérieure à 100 pb en fonction de la température. Le calcul est effectué sur la base d’un génome complet comprenant 3 milliards de nucléotides, subissant des dommages uniquement sur les purines et de façon identiques sur tout le génome. D’après (Willerslev *et al.*, 2004).

Les amplitudes de température ont également un effet dans la dégradation des molécules d’ADN. Par exemple, on pense qu’il y a peu d’ADN dans les échantillons de musées à cause des variations de températures dans les réserves ; mais les lavages et vernissages subis par ces échantillons peuvent également être responsables de cette perte d’ADN.

En plus de la température, la conservation de l’ADN va également dépendre des propriétés physico-chimiques de l’environnement : un pH neutre ou basique, comme dans les grottes, permettra une meilleure préservation. Les conditions optimales de conservation sont des températures toujours basses et constantes, une haute salinité du milieu ou une dessiccation post-mortem rapide (Shapiro, 2008).

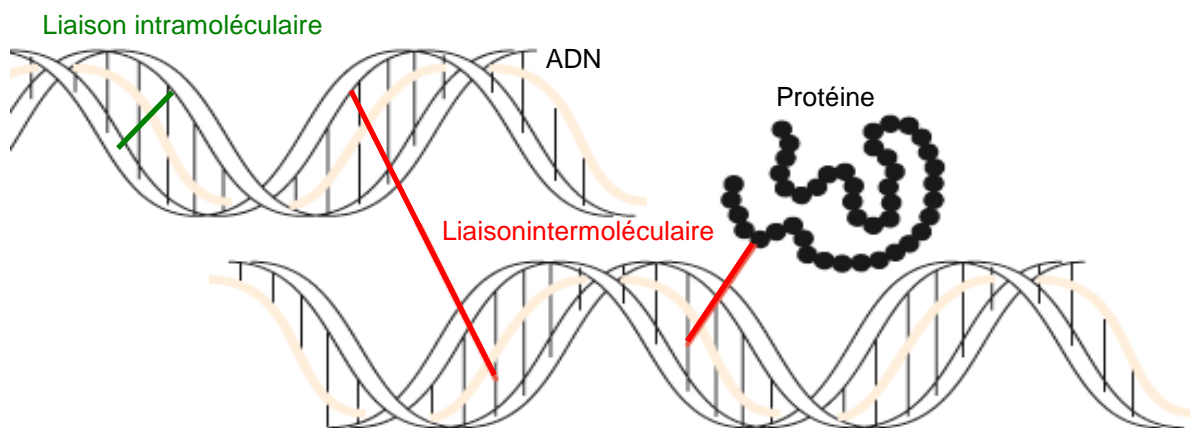
### 1.5.2. Les lésions bloquantes

Les dégradations ne sont pas les seuls dommages subis par les molécules d’ADN ancien. En effet, certains échantillons, dont le contenu en ADN était visualisé par hybridation, pouvaient ne donner aucun résultat lors de PCR (d’Abbadie *et al.*, 2007). L’utilisation de polymérase modifiée dans le but d’amplifier l’ADN malgré ces dommages, a permis d’obtenir

de l'ADN dans des échantillons qui en étaient apparemment dépourvus. La présence de lésions bloquantes sur les molécules d'ADN ancien a ainsi pu être mise en évidence (Shapiro, 2008).

Afin d'étudier ces lésions bloquantes, la technique de la Polymerase Extension Profiling (PEP) a été mise au point (Heyn *et al.*, 2010). Cette dernière permet de détecter directement, sur chaque fragment, les lésions qui empêchent l'avancée de la polymérase. Cette étude a montré que les lésions bloquantes touchaient une quantité variable de séquences d'ADN en fonction des conditions de préservation de celui-ci. Ainsi, les séquences issues d'échantillons du permafrost - un os de mammouth et deux de chevaux - présentent respectivement 2, 10 et 39 % de lésions bloquantes (Heyn *et al.*, 2010).

Ces lésions peuvent être dues à la présence de liaisons créées entre des macromolécules lors de la réaction de Maillard ; celle-ci faisant apparaître des liaisons entre les sucres et les groupements amines des molécules organiques (Figure 9). L'analyse par spectrométrie de masse de coprolithes de paresseux a révélé la présence d'alkylpyrazines, de furanones et de furaldéhydes. Ces produits sont spécifiques de la réaction de Maillard et pourraient résulter de la création d'un grand nombre de liaisons croisées (Poinar *et al.*, 1998).



**Figure 9.** Les différents types de liaisons covalentes empêchant la progression de la Taq Polymérase.

L'utilisation de N-phenacylthiazolium bromide (PTB) facilite l'amplification d'ADN en coupant ces liaisons entre les sucres réducteurs et les amines primaires (Vasan *et al.*, 1996).

Ces lésions peuvent également être induites par des radicaux libres. En effet, les réactions d'oxydation peuvent conduire à la rupture du cycle carboné du sucre entraînant la formation de bases azotées oxydées, telle que la 5-hydroxy-thymine, qui bloque l'élongation de l'ADN-polymérase. Les attaques oxydantes provoquées par les rayons ultraviolets sont à l'origine de la formation d'une structure en anneau entre deux pyrimidines successives.

Des études menées sur des échantillons provenant du permafrost ont montré que les lésions bloquantes sont les premières causes de dommage dans les molécules d'ADN ancien avec une fréquence de  $4.10^{-4}$ . Elles sont plus nombreuses sur l'ADN nucléaire que sur le mitochondrial. Cela serait dû à la proximité du premier avec les protéines des histones, ce qui favoriserait la réaction de Maillard (Binladen *et al.*, 2006). Cependant, ces données ont été obtenues sur des échantillons du permafrost où l'eau liquide, à l'origine des hydrolyses et des oxydations, est rare. Il est possible que des échantillons provenant de milieux tempérés présentent un autre type de profil de dommages.

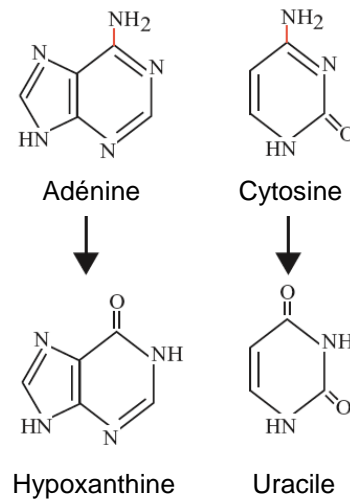
### **1.5.3. Les dommages subis par l'ADN ancien**

Des études ont montré que, lors d'une réaction d'amplification, le taux d'erreur était de  $8.10^{-3}$  avec de l'ADN ancien alors qu'il n'est que de  $2.10^{-3}$  avec de l'ADN moderne (Olivieri *et al.*, 2010). Cette différence est occasionnée par des dommages causés au niveau des bases azotées de l'ADN par des réactions d'oxydation. En effet, celles-ci en modifiant les interactions électrostatiques entre les bases conduisent à l'incorporation de nucléotides erronés lors des réactions de PCR.

La plupart des mutations observées sont des transitions. Elles sont de deux types : A vers G et T vers C pour le type 1 (TS1), ou G vers A et C vers T pour le type 2 (TS2). Les TS2 sont largement majoritaires par rapport aux TS1 dans l'ADN ancien avec un ratio de 8/1 (Binladen *et al.*, 2006 ; Olivieri *et al.*, 2010) ; elles représentent 94% des transitions (Briggs *et al.*, 2007) et 88% des mutations totales subies par l'ADN (Gilbert *et al.*, 2007). Ces observations s'expliquent par le fait que les TS1 sont de simples artefacts de PCR, conduisant aléatoirement à l'incorporation d'un nucléotide erroné, tandis que les TS2 sont le reflet des dommages subis par les molécules d'ADN ancien (Figure 10). Ces transitions de type 2 sont causées par la désamination de la cytosine ou de son homologue, la 5-méthylcytosine, en

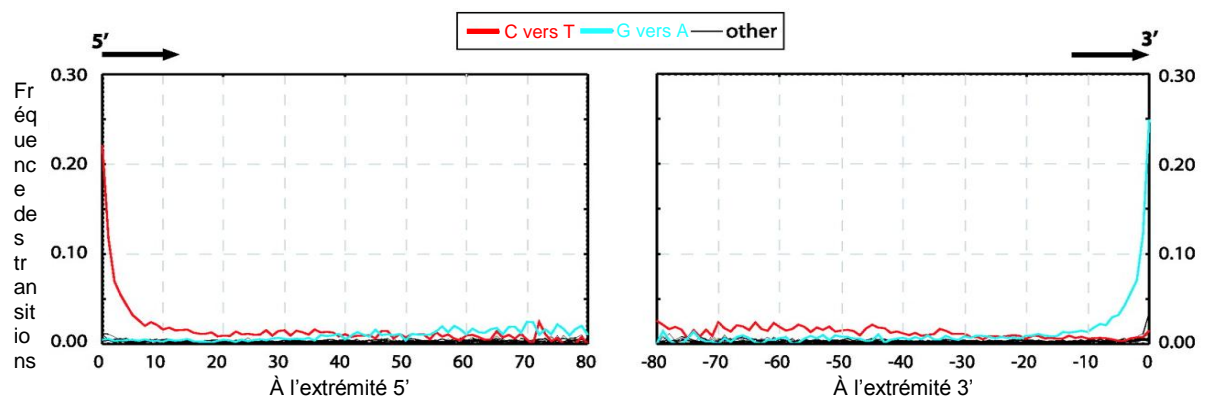


uracile, et celle de l'adénine en hypoxanthine (Willerslev & Cooper, 2005) ; les plus fréquentes étant les désaminations de cytosine.



**Figure 10.** Les transitions de type 2 dues aux désaminations. D'après (Willerslev & Cooper, 2005).

Ces TS2 ne sont pas réparties équitablement le long de la séquence d'ADN amplifiée. En effet, la proportion des transitions C vers T est 50 fois plus importante à l'extrémité 5'. Il a également été montré que les transitions G vers A sont beaucoup plus nombreuses près de l'extrémité 3', et correspondent souvent aux transitions C vers T touchant le brin complémentaire (Figure 11). Les molécules « simple brin » étant très sensibles aux désaminations, ces mutations observées aux extrémités sont dues aux molécules d'ADN « simple brin » dépassant de part et d'autre du « double brin » et accumulant les désaminations (Briggs *et al.*, 2007).



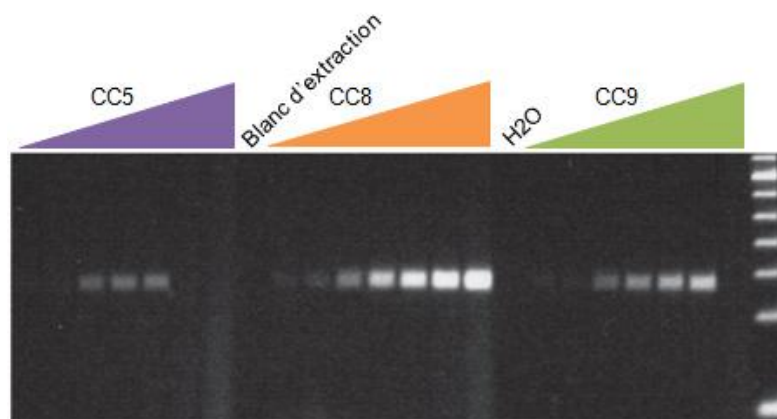
**Figure 11.** Mise en évidence de l'augmentation des transitions de type 2 aux extrémités 5' et 3' des molécules d'ADN ancien. D'après (Briggs *et al.*, 2007).

Pour obtenir une séquence fiable à partir d'un échantillon, il faut donc avoir plusieurs lectures pour une même position du génome. Grâce aux techniques de séquençage à haut débit, il n'y a plus besoin d'effectuer de nombreuses PCR car plusieurs lectures pour une même position du génome peuvent être obtenues avec une seule banque. Une autre technique mise au point est d'utiliser une uracile-DNA-glycosylase qui va couper les bases azotées du brin d'ADN afin de ne pas amplifier les brins abîmés lors de la PCR (Hofreiter *et al.*, 2001).

#### 1.5.4. Les molécules inhibitrices

Des molécules présentes dans le sol environnant ou dans l'échantillon lui-même peuvent être extraites en même temps que l'ADN ancien. Certaines, comme des acides humiques et fulviques principalement trouvés dans l'humus, vont avoir une action inhibitrice qui va empêcher l'amplification de l'ADN lors de la PCR. Ces molécules seraient plus présentes dans certains échantillons que dans d'autres ; ainsi les cheveux, les coprolithes et le sol en contiendraient plus que les restes osseux (King *et al.*, 2009).

La présence de ces molécules inhibitrices nécessite de mettre une petite quantité d'extrait d'ADN ; en effet, lorsque celles-ci sont trop nombreuses l'amplification échoue (Figure 12). Des études ont montré que l'addition de certaines substances comme la protéine T4 gp32, la BSA, la bétaïne ou le sperme de saumon permettait de diminuer l'action de ces molécules inhibitrices.



**Figure 12.** Amplification d'un fragment par PCR de 33 cycles à partir de 0,04 à 2,5 % de l'extrait d'ADN. D'après (Bon *et al.*, 2012).

### 1.5.5. Les contaminations

En raison des difficultés liées à son amplification et de sa faible quantité dans les restes archéologiques, l'ADN ancien est très sensible aux contaminations par de l'ADN moderne. Les échantillons peuvent ainsi être contaminés par différents types ADN à plusieurs moments du processus.

Tout d'abord, contaminé par des séquences microbiennes et environnementales provenant du sol dans lequel se trouvaient les échantillons (Green *et al.*, 2006). Avec les PCR à amorces spécifiques, ces séquences n'étaient pas un réel problème ; mais avec la construction de banques de fragment il faut être attentif à ce que ces séquences d'ADN exogène n'éclipsent pas le signal de l'ADN endogène étudié.

Une autre source d'ADN parasite est la contamination par de l'ADN humain, qui peut se produire à partir du prélèvement jusqu'aux analyses en laboratoire.

Si ces contaminations sont rapidement détectées quand les analyses portent sur des animaux ou des végétaux, elles sont beaucoup plus difficiles à gérer quand les études portent sur des échantillons de la lignée humaine (Delpozzo & Guardiola, 1989).

De plus, avant le séquençage complet du génome humain, les espèces animales éteintes pouvaient également être touchées par des contaminations humaines. En effet, l'ADN nucléaire contient des « numts », qui sont des insertions de séquences provenant du génome mitochondrial. Avant que ces « numts » ne soient répertoriés, ils pouvaient être décrits à tort comme des séquences mitochondriales des animaux étudiés (Willerslev & Cooper, 2005). Ainsi, la séquence retrouvée dans un échantillon de dinosaure est en fait un « numt » humain (Zischler *et al.*, 1995).

Dans les laboratoires, en plus des contaminations par de l'ADN humain moderne, il peut y avoir deux autres types de contaminations. Tout d'abord, celles qui sont dues aux produits et réactifs utilisés pour effectuer les extractions, les amplifications et le séquençage de l'ADN ancien. En effet, certains produits, telle que la BSA, sont fabriqués à partir de molécules animales qui peuvent laisser des traces d'ADN. Les contaminations peuvent également être le fait de molécules d'ADN amplifiées lors de précédentes expériences et qui se sont répandues dans le laboratoire. Ces molécules qui appartiennent souvent à une espèce

proche de celles étudiées sont très difficiles à détecter à posteriori (Willerslev & Cooper, 2005).

Afin de limiter au maximum les contaminations, des stratégies ont été mises en place sur le terrain et dans les laboratoires. Dans ces derniers, les échantillons et les extraits d'ADN ancien sont manipulés dans des emplacements dédiés où il n'y a pas eu d'amplification d'ADN, qu'il soit moderne ou ancien. Ces lieux peuvent être en surpression afin d'éviter l'entrée de molécules venues de l'extérieur, et posséder un système de radiation aux ultraviolets afin d'éliminer l'ADN entre chaque manipulation. Sur le terrain, les échantillons doivent être manipulés avec précaution (port de gants) et mis en sachet stérile avant même d'être sorti du site de prélèvement afin d'éviter toute contamination par un ADN extérieur.

Un autre contrôle possible, afin de s'assurer qu'il n'y a pas eu de contamination, est de réaliser des blancs d'extraction et d'amplification. La reproductibilité dans un autre laboratoire est également un bon test afin de s'assurer qu'il n'y a pas eu de contamination (Willerslev & Cooper, 2005).

## **1.6. Les échantillons étudiés**

Tous les échantillons d'origine animale ou végétale retrouvés dans les sites de fouille sont susceptibles d'être des sources d'ADN amplifiable. Malheureusement, l'étude de l'ADN ancien étant destructive et les paléontologues et les anthropologues ne voulant pas - à juste titre - sacrifier leurs plus belles pièces, les échantillons fournis ne sont pas forcément les mieux conservés et donc ceux contenant le plus d'ADN ou aisément datables.

### **1.6.1. Les tissus mous**

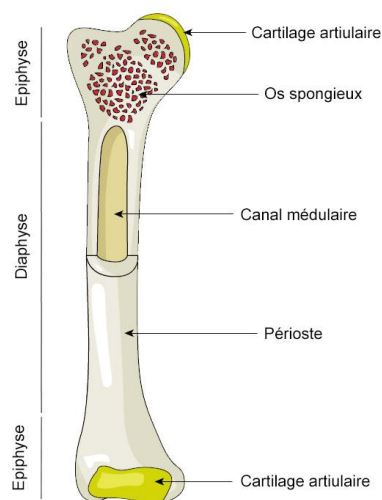
Les tissus mous, constitués par les muscles, les vaisseaux sanguins ou encore les tissus adipeux, sont les échantillons les plus à même de fournir de l'ADN amplifiable car ce sont les parties du corps qui contiennent le plus de cellules. C'est pourquoi, les recherches sur les animaux de musée ou sur les momies sont été menées à partir des tissus mous (Higuchi *et al.*, 1984 ; Pääbo *et al.*, 1988). Malheureusement ces tissus, qu'ils soient momifiés ou naturalisés, ne contiennent pas toujours de l'ADN amplifiable, à l'image des momies égyptiennes étudiées lors des premières expérimentations sur l'ADN ancien (Pääbo, 1985). Cette absence serait expliquée par les températures élevées régnant en Egypte. De plus, les produits utilisés

par les embaumeurs lors de la momification pourraient, eux aussi, être responsables de la non-conservation de l'ADN.

Les milieux froids permettent une congélation des tissus mous et donc une bonne conservation de ceux-ci ; c'est donc dans des échantillons issus du permafrost que la plus grande quantité d'ADN amplifiable a été retrouvée. Ainsi des études ont été menées sur des restes de muscles de mammouths issus du permafrost (Noro *et al.*, 1998 ; Rogaev *et al.*, 2006) et sur un homme, daté à 4 546 ans BP, retrouvé dans un glacier du Tyrol (Ermini *et al.*, 2008) permettant l'obtention de données génétiques sur leurs génomes mitochondriaux et nucléaires.

### 1.6.2. Les os

Les restes osseux contiennent environ 300 fois moins d'ADN que les tissus mous ; en effet, pour un prélèvement frais, il y a entre 3 et 16  $\mu\text{g}$  d'ADN par gramme d'os compact alors qu'il y en a entre 1 000 et 5 000  $\mu\text{g}$  dans les tissus mous (Hofstetter *et al.*, 1997 ; Hochmeister *et al.*, 1991). Cependant, comme les os sont présents en très grand nombre sur les sites archéologiques, c'est vers eux que les paléogénéticiens vont se tourner.



**Figure 13.** Structure d'un os long.

La matrice de l'os compact est très solide et va permettre de protéger l'ADN des contaminations et des dégradations (MacHugh *et al.*, 2000). De plus, il a été montré que l'hydroxyapatite de calcium présente naturellement dans les os adsorbe l'ADN et le protège

des dommages qu'il pourrait subir. En effet, l'adsorption de l'ADN diminuerait par deux le risque de dépurination de celui-ci. Cette théorie est soutenue par l'observation de la corrélation négative entre les altérations minérales subies par l'os et la conservation de l'ADN (Götherström *et al.*, 2002).

La partie de l'os préférentiellement sélectionnée pour les amplifications d'ADN est le cortex solide. Bien que la partie spongieuse de l'os soit plus riche en cellules, la matrice permet une meilleure conservation de l'ADN contenu dans le cortex, ainsi qu'une relative protection contre les contaminations extérieures par les bactéries ou les champignons.

### 1.6.3. Les dents

Tout comme les os, les dents sont des restes archéologiques très fréquemment retrouvés. Elles sont considérées comme une meilleure source d'ADN amplifiable que les restes osseux car leurs tissus vivants contiennent plus de cellules. De plus, la barrière formée par l'émail et le cément est moins poreuse que celle des os et protège donc plus efficacement l'ADN contre les dégradations et les contaminations. Lors d'une étude comparative sur la conservation de l'ADN dans des dents humaines et de bisons, il a été montré que le cément contenait plus d'ADN amplifiable que la dentine ou l'émail (Adler *et al.*, 2011).

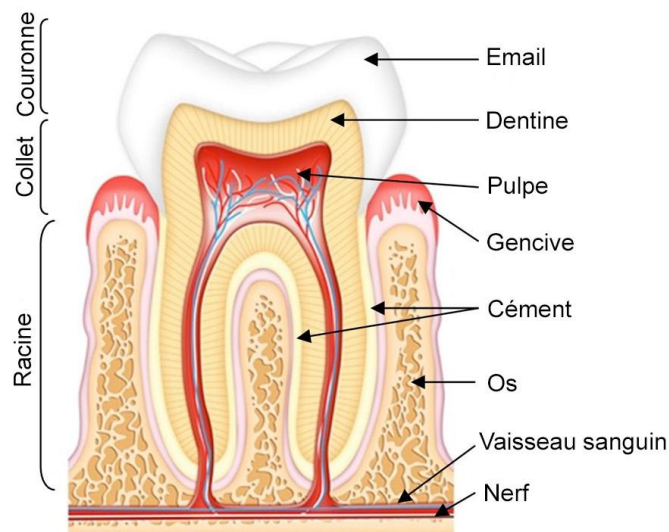
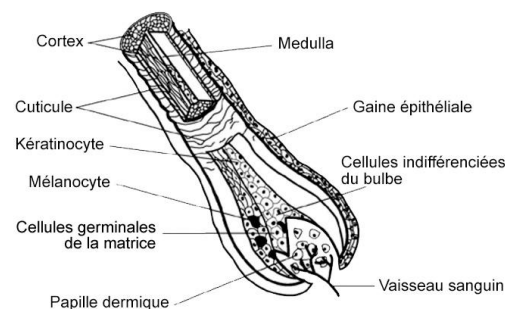


Figure 14. Structure d'une dent.

Plusieurs études paléogénomiques ont été réalisées à partir de dents très anciennes telles que des molaires de mammoths nains retrouvées sur une île de l'océan Arctique (Noonan *et al.*, 2005 ; Reich *et al.*, 2010).

#### 1.6.4. Les cheveux et les poils

Les poils et les cheveux sont moins riches en ADN que les autres échantillons archéologiques étudiés. Cependant, bien qu'en moindre quantité, les molécules d'ADN sont bien conservées grâce aux propriétés hydrophobes de la kératine qui permettent d'éviter les réactions d'hydrolyse et d'oxydation. De plus, le très faible taux de contamination extérieure observé suggère que ces échantillons y sont imperméables grâce à la protection apportée par le cortex et la medulla (Gilbert *et al.*, 2008b). La conservation de l'ADN est meilleure à la racine du poil ; la dégradation de l'ADN dans le reste du cheveu serait due à l'étape de kératinisation subie par celui-ci lors de la formation du cheveu (Bengtsson *et al.*, 2011).



**Figure 15.** Structure d'un follicule pileux. D'après (Bengtsson *et al.*, 2011).

Des études ont montré que l'ADN mitochondrial est particulièrement abondant dans les poils et les cheveux. Ce fait étonnant n'a pas encore été totalement expliqué. Une des hypothèses avancées est que la membrane des mitochondries protégerait mieux l'ADN que celle du noyau lors des étapes de kératinisation. Une autre hypothèse est que la haute activité métabolique engendrée par la formation du follicule pileux nécessiterait une grande quantité de mitochondries (Bengtsson *et al.*, 2011).

Les cheveux et les poils ont été utilisés comme source d'ADN amplifiable lors de l'analyse d'échantillons historiques (Gilbert *et al.*, 2004 ; Amory *et al.*, 2007 ; Melchior *et al.*, 2010) et de restes beaucoup plus anciens d'Hommes (Gilbert *et al.*, 2008b ; Rasmussen *et al.*, 2010) ou d'animaux tels que des mammoths ou des bisons issus du Permafrost (Gilbert *et*

*al.*, 2004 ; Gilbert *et al.*, 2008a ; Miller *et al.*, 2008). Un des avantages des cheveux et des poils est qu'ils permettent d'éviter la destruction de restes archéologiques d'une valeur élevée. Cependant, il faut rester vigilant car l'eumélanine et la phéomélanine, deux constituants de la mélanine naturelle, peuvent agir comme des inhibiteurs lors des PCR (Bengtsson *et al.*, 2011).

Des ongles, des sabots ainsi que des plumes peuvent également être utilisés comme source d'ADN amplifiable pour des échantillons anciens (Bengtsson *et al.*, 2011 ; Olsen *et al.*, 2011).

### 1.6.5. Les coprolithes

Le terme coprolithe a été créé en 1829 par W. Buckland pour désigner les excréments minéralisés de dinosaures (Buckland, 1829). Aujourd'hui, cette terminologie sert pour tous les excréments fossilisés, aussi bien dans un contexte paléontologique qu'archéologique. Les coprolithes sont le plus souvent produits par des carnivores broyeurs d'os (canidés, hyénidés) ; le calcium contenu dans les os des proies permettant leur fossilisation. Les coprolithes ont longtemps été utilisés en palynologie pour la reconstitution des paléo-environnements grâce aux pollens qu'ils contiennent ; ceux-ci provenant de l'eau absorbée ou des viscères des herbivores ingérés et permettant de connaître la flore et le climat du milieu dans lequel évoluait le spécimen qui a produit le coprolithe (Martin & Sharrock, 1964 ; Leroi-Gourhan, 1965).



**Figure 16.** Coprolithes présents à l'entrée de la Galerie des Croisillons de la Grotte Chauvet-Pont d'Arc, attribués à un canidé.



Des expériences menées sur des coprolithes de paresseux (Poinar *et al.*, 1998 ; Poinar *et al.*, 2003) et d'Hommes de la période Pré-Clovis (Gilbert *et al.*, 2008c) ont montré que de l'ADN amplifiable par PCR était présent dans les coprolithes. Ces derniers contiennent l'ADN du producteur, des centaines de milliers de cellules épithéliales de l'intestin desquamant chaque jour. Cet ADN permet de connaître l'espèce productrice de manière certaine, sans se baser sur une comparaison de forme des excréments avec les espèces actuelles. Un autre avantage des coprolithes est qu'ils permettent d'avoir accès à l'alimentation du spécimen ; ainsi de l'ADN de renne a pu être mis en évidence dans des coprolithes d'hyène des cavernes (Bon *et al.*, 2012). Il est également possible de caractériser les parasites intestinaux contenus dans ces échantillons (Loreille *et al.*, 2001). Les coprolithes sont donc des échantillons de choix pour des études métagénomiques.

#### **1.6.6. Le sol**

Le sol est lui aussi un échantillon très intéressant à étudier en paléométagénomique, car il peut contenir des fragments d'ADN de nombreuses espèces. Des carottes extraites des sols constamment gelés du permafrost ont permis de mettre en évidence des séquences de vertébrés, de végétaux et de bactéries (Willerslev *et al.*, 2003 ; Willerslev *et al.*, 2004). Les températures très basses atteintes dans les régions présentant ce type de sol (entre - 9°C et - 50°C) permettent une bonne conservation de l'ADN. De plus les conditions anaérobies, dues à la grande quantité de méthane présent (plus de 40 ml kg<sup>-1</sup>), diminuent fortement les réactions d'oxydation subies par l'ADN (Willerslev *et al.*, 2004). Ces deux conditions, combinées au pH neutre qui règne dans le permafrost, font que l'ADN peut se conserver jusqu'à plusieurs centaines de milliers d'années dans ces régions (Willerslev *et al.*, 2004).

Des molécules d'ADN peuvent également être retrouvées dans les sols des régions tempérées. Des séquences de deux espèces de moas (oiseaux aujourd'hui disparus) et de plantes présentes avant l'arrivée des ancêtres des Maori ont été amplifiées à partir de carottes de sédiments, âgés de 600 ans, issus d'une grotte de Nouvelle-Zélande (Willerslev *et al.*, 2003). De plus, l'analyse de ces carottes a permis de remarquer que la circulation verticale de l'ADN dans le sol est très limitée. En effet, les séquences d'ADN de moas retrouvées dans les couches les plus vieilles n'étaient pas mélangées avec les séquences d'ADN de moutons modernes qui étaient localisées dans les couches supérieures des carottages (Willerslev *et al.*, 2003).

### **1.6.7. Les coquilles d'œufs**

L'amplification d'ADN nucléaire et mitochondrial à partir de l'extérieur et de l'intérieur d'œufs de moas retrouvés en Nouvelle-Zélande a montré que les coquilles de ces derniers étaient des sources d'ADN amplifiable. Protégé dans la matrice minérale de la coquille, l'ADN peut se conserver très longtemps. De plus la présence de molécules antibiotiques dans la coquille permet de limiter les contaminations extérieures par les bactéries (Oskam *et al.*, 2010). Des analyses complémentaires de l'ADN présent sur les couches externes des œufs ont également permis de montrer que c'était les mâles qui couvaient la portée chez certaines espèces de moas (Huynen *et al.*, 2010).

## 2. LES CARNIVORES DU PLÉISTOCÈNE

### 2.1. Les loups du Pléistocène

L'un des deux carnivores étudiés au cours de cette thèse est le loup, *Canis lupus*. Présente dans notre environnement depuis des dizaines de milliers d'années, ancêtre du chien domestique et pourtant excessivement crainte et chassée, cette espèce a toujours eu un rapport très particulier avec les Hommes. Le loup appartient à la sous-famille des *Caninae* et plus particulièrement au genre *Canis*.

#### 2.1.1. La famille des *Canidae*

La famille des *Canidae* est composée d'un large groupe de mammifères prédateurs de l'ordre des carnivores. Ces membres possèdent une paire de dents carnassières (la quatrième prémolaire supérieure et la première molaire inférieure) qui se sont modifiées au fil de l'évolution et donnent une grande efficacité dans le déchirement de la peau, des tendons et des muscles des proies. Les *Canidae* sont également caractérisés par un bulbe tympanique enflé composé d'une chambre unique divisée par un pseudo-septum (contrairement aux *Félidae* qui possèdent un véritable septum), la perte d'une artère stapédienne et par la position médiane de leur artère carotide (Sillero-Zubiri *et al.*, 2004). La famille des *Canidae* est composée de deux sous-familles fossiles (les *Hesperocyoninae* et les *Borophaginae*) et d'une sous-famille actuelle (les *Caninae*).

Les *Hesperocyoninae* forment le groupe le plus ancien et l'hespérocyon, qui en est le membre basal, est également l'ancêtre des *Borophaginae* et des *Caninae* (Wang, 1994). Une des différences majeures entre les *Hesperocyoninae* et les deux autres sous-familles est la transformation de la carnassière inférieure qui va passer de tranchante (en forme de lame) à creusée. C'est cette caractéristique commune qui justifie l'hypothèse d'un ancêtre commun aux *Canidae* et aux *Borophaginae*. Cette modification dentaire permet une alimentation s'étalant sur la gamme d'un régime très à peu carné.

La deuxième sous-famille fossile, celles des *Borophaginae*, était le premier groupe de *Canidae* à présenter une importante diversité morphologique et taxonomique au sein du

continent Nord-Américain (Wang *et al.*, 1999). Dans le large spectre morphologique de cette sous-famille, se côtoient des espèces omnivores (avec des habitudes alimentaires semblables aux *Procyonidae* actuels) et des espèces au régime alimentaire très carné avec une dentition très développée capable de broyer les os, comme le font les hyènes actuelles.

Les premiers *Caninae* font leur apparition à l'Oligocène inférieur avec l'émergence du genre *Leptocyon* qui va persister jusqu'au Miocène inférieur. Les animaux de ce groupe sont de la taille d'un petit renard et possèdent des prémolaires étroites et minces dues à une mâchoire inférieure allongée. Sans doute du fait de la forte concurrence des deux autres sous-familles de *Canidae*, le genre *Leptocyon* présentait une faible diversité, avec seulement deux ou trois espèces qui coexistaient et étaient toujours de petite taille. Le premier représentant du genre *Canini* à faire son apparition est l'*Eucyon* qui fait la taille d'un petit chacal et se distingue du genre *Vulpini* par un processus paroccipital étendu et la présence d'un sinus frontal. Ce dernier caractère initie des transformations qui vont entraîner le développement de sinus complexes et d'un crâne en forme de dôme chez le loup gris. Des espèces du genre *Eucyon* apparaissent au Miocène supérieur en Asie et au Pliocène inférieur en Europe. Les fossiles nord-américains sont antérieurs et suggèrent donc une dispersion vers l'ouest de cette forme. Les *Caninae* sont caractérisés par de nombreuses molaires, des griffes non-rétractiles et par la longueur de leurs membres par rapport à la taille de leur corps. La famille des *Caninae* regroupe 21 genres (fossiles ou actuels) et c'est au genre *Canis*, auquel appartient le loup gris, que nous allons nous intéresser plus particulièrement.

### **2.1.1.2. Le genre *Canis***

Le genre *Canis* apparaît à la limite entre le Miocène et le Pliocène (entre 5 et 6 Ma) en Amérique du Nord. Les espèces du genre *Canis* sont déjà présents sous leur forme ancestrale et sont de taille plus réduite qu'actuellement (telles que *C. ferox* ou *C. legophagus*). Au même moment, les premiers *Canidae* apparaissent en Europe (*Eucyon monticinensis* en Italie au Pléistocène inférieur et *Nyctereutes donzani* en France au Pliocène inférieur). Le premier fossile attribuable au genre *Canis* en Eurasie est un spécimen vieux de 3,4 Ma, retrouvé en Chine. L'histoire des *Canis* est complexe à retracer à cause de la rapide et massive radiation du genre qui s'est produite lors de son arrivée en Eurasie. Cette expansion a donné naissance à de nombreuses espèces proches au Pléistocène et au Pliocène. Des études ont montré que cette radiation massive pourrait être due à un changement dans la guildes des carnivores et à un

développement des ruminants de grande taille au détriment de ceux de petite taille (Brugal & Croitor, 2007). Les taxons du genre *Canis* possèdent des canines moins longues indiquant une préhension plus réduite, une morphologie crânienne différente avec entre autre une région frontale élevée, un processus post-orbitaire convexe et une stature plus imposante (Sillero-Zubiri *et al.*, 2004). Actuellement, les *Canis* représentent six des trente-quatre espèces de la famille des *Canidae* mais au Pléistocène ce nombre était beaucoup plus important.

## **2.1.2. Histoire évolutive des *Canis* durant le Pléistocène en Europe**

### ***2.1.2.1. Les espèces décrites au Pléistocène inférieur***

La radiation rapide des canidés en Eurasie a laissé de nombreux fossiles dans un très grand nombre de gisements. Cette richesse archéologique rend ardu le travail des paléontologues pour définir ce qui correspond à une nouvelle espèce et ce qui est dû à la variabilité naturelle. De ce fait, de très nombreuses espèces ont été décrites au Pléistocène inférieur, dont 5 majeures.

La première, *Canis arnensis*, a été décrite à partir des ossements retrouvés sur le site de Tasso (Corse du Sud, France) ; des restes sont également présents au Valdarno (Toscane, Italie). Les représentants de cette espèce sont de petite taille avec des dents inférieures alignées et une morphologie crânienne différente de celle des loups modernes. En effet, la boîte crânienne est large et bombée avec une crête sagittale nette comme chez le chacal actuel mais ne se prolonge pas à l'avant du crâne. Ce canidé peut être rapproché soit des chacals comme le propose D. Torre sur la base du rapport LM1/LM2 (Torre, 1967) soit des coyotes en utilisant la longueur des dents inférieures LP4/LM1 selon B. Kurtén (Kurtén, 1974).

Les premières traces de la deuxième espèce, *Canis etruscus*, sont visibles en Chineau Villafranchien basal (Sotnikova & Rook, 2010), mais c'est à partir des restes retrouvés au Valdarno que la diagnose a été établie. *Canis etruscus* était de la taille d'un chien de berger, ressemblant écologiquement au loup indien actuel : il était favorisée dans les milieux boisés mais absent des steppes (Kurtén, 1968). *Canis etruscus* possède un museau long et étroit. Il est caractérisé par une dentition primitive avec des M1 et des M2 relativement développées et il se différencie de *Canis arnensis* par un rapport LM1/LM2 plus grand. Il se distingue de

*Canis lupus* par la possession d'une carnassière inférieure moins développée et de *Canis latrans* par la conformation de sa cavité céphalique et le développement de sa crête sagittale.



**Figure 17.** Héli-mandibule d'un spécimen de *Canis etruscus* provenant du Valdarno supérieur.

*Canis senezensis* a été décrit grâce à deux fragments d'hémi-maxillaires du gisement de Senèze (Haute-Loire, France). L'association faunique a permis de définir cette espèce comme appartenant au Villafranchien supérieur. Les spécimens de cette espèce présentent un museau long avec des prémolaires très espacées à couronne assez basse et sans denticule postérieur. Ces caractéristiques qui diffèrent de *Canis etruscus* et *Canis arnensis* ont conduit à la diagnose de cette nouvelle espèce (Martin, 1973).

En comparant les cinq fragments de crânes, de maxillaires et de mandibules (découverts dans le gisement grec d'Apollonia 1, situé en Albanie) avec les restes déjà connus de *Canis etruscus* et *Canis arnensis*, Koufos et Kostopoulos ont défini une nouvelle espèce : *Canis apolloniensis* (Koufos & Kostopoulos, 1997). Cette espèce diffère de *C. arnensis* par une plus grande stature, des denticules postérieurs nettement développés et une carnassière plus longue. En revanche, il est plus petit que *C. etruscus* et possède un palais plus long et plus large, des PM2 et PM3 moins développées et une carnassière inférieure plus courte. Il se caractérise également par une forte constriction médiale sur la première molaire supérieure.

Enfin, la dernière espèce, *Canis accitadinus*, a été décrite grâce aux restes retrouvés dans le gisement de Fonelas P-1 (Grenade, Espagne). La diagnose a été établie à partir d'un

crâne comportant sept dents et d'un radius. *Canis accitanus* est un petit spécimen avec un crâne long et étroit. La partie caudale des os nasaux se développe au delà de l'extrémité caudale du maxillaire. Le frontal est proéminent avec une crête sagittale peu développée. Les prémolaires supérieures sont espacées et seule la UP3 présente un denticule postérieur (Garrido & Arribas, 2008).

Beaucoup d'autres espèces ont été décrites à la fin du 19<sup>ème</sup> siècle telles que *Canis falconeri major*, *Canis davisi*, *Canis chiliensis* ou encore *Canis adoxus*. Aujourd'hui, *Canis falconeri major* est affilié à la lignée des *Xenocyon* tandis que les trois autres espèces le sont au genre *Eucyon*. Ce débat sur le statut d'espèce porte aussi sur les spécimens décrits ci-dessus. En effet, la distinction entre *Canis etruscus* et *Canis arnensis* est sujette à débat. Pour M.-F. Bonifay, ces deux espèces devraient être regroupées sous le nom unique de *C. etruscus* car les différences observées seraient dues aux contraintes auxquelles ont été soumis les différents spécimens (Bonifay, 1971). Pour L. Rook, au contraire, les deux espèces sont bien distinctes avec une évolution régionale parallèle avec *C. arnensis* autour de la Méditerranée et *C. etruscus* en Europe centro-septentrionale (Rook, 1993). Le statut de *Canis senezensis* a lui aussi été remis en question et il semblerait qu'il soit à rapprocher de *C. arnensis*. Pour *Canis apolloniensis*, le statut d'espèce n'était pas clair car il apparaissait comme une sorte d'intermédiaire entre *C. etruscus* et *C. arnensis* d'un point de vue biométrique et morphologique. Des études menées récemment sur des données morphologiques et biométriques (Boudadi-Maligne, 2012) proposent de ne retenir que deux des cinq espèces évoquées : *C. arnensis* et *C. etruscus*. En effet, d'un point de vue statistique, *C. senezensis* et *C. accitanus* seraient indissociables de *C. arnensis* pour les données biométriques. Les différences observées pour *C. accitanus* ne seraient dues qu'à de la variabilité individuelle. *C. apolloniensis* serait lui aussi à classer avec *C. arnensis* au vue de l'absence de différence significative dans la forme de son système masticateur. *C. etruscus* garde son statut d'espèce à part dû fait de sa stature plus importante que *C. arnensis*, de ses prémolaires supérieures plus robustes et de ces LM2 et LM3 plus réduites (Boudadi-Maligne, 2010).

### **2.1.2.2. *Canis mosbachensis***

À la fin du Pléistocène inférieur apparaît en Europe occidentale, la forme *mosbachensis*. Il est morphologiquement proche de *Canis arnensis* bien que légèrement plus grand que celui-ci. Il est néanmoins de petite stature avec un crâne court et possède une

mandibule puissante par rapport à sa dentition. Ses dents sont peu espacées et le denticule postérieur peu développé sur les P2 inférieures. W. Soergel qui a effectué la diagnose y voit un descendant de *C. arnensis*. Il propose que *Canis mosbachensis* puisse être l'ancêtre des loups du Pléistocène moyen et supérieur. Mais ici aussi il y a plusieurs hypothèses sur le statut de cette espèce. Il est proposé que *C. mosbachensis* soit une espèce à part entière au vu de ses particularités morphologiques qui le distinguent de *C. etruscus* mais aussi des premiers vrais loups. Une autre hypothèse voudrait que ce spécimen soit une forme de *C. etruscus* ou une sous-espèce de loup en raison de sa morphologie déjà très proche des premiers *C. lupus* par exemple celle de son trigonide et ce malgré une différence de stature. Des restes de *C. mosbachensis* retrouvés dans la grotte de l'Escaie présentent des caractères beaucoup plus proches des *Canis lupus*, ce qui soutient l'hypothèse de la connexion phylogénétique entre *C. mosbachensis* et *C. lupus*. Certains auteurs lui donnent même l'appellation de *C. lupus mosbachensis* (Torre, 1974).

La fin du Villafranchien a donc vu la coexistence de trois espèces de canidés, *C. etruscus*, *C. arnensis* et *C. mosbachensis* avant que ceux-ci ne disparaissent au profit des premières occurrences de la lignée des *lupus*.

### **2.1.2.3. Les premiers *Canis lupus***

La lignée des *Canis lupus* stricto sensu fait son apparition vers le milieu du Pléistocène moyen, il y a 450 000 ans environ. À partir de ce moment et jusqu'à la fin du Pléistocène supérieur, la lignée des *Canis lupus* va voir la taille de ses représentants augmenter de façon continue. Les restes de loups retrouvés dans les gisements du Pléistocène moyen et supérieur ont permis la diagnose de trois nouvelles sous-espèces chronologiques qui soulignent bien cette évolution.

La première, *Canis lupus lunellensis*, décrite par M.-F. Bonifay à partir des ossements du site de Lunel-Viel (Languedoc-Roussillon, France), est composée d'individus plus petits que les loups actuels mais qui présentent la même morphologie (Bonifay, 1971). La première molaire supérieure de *Canis lunellensis* est de faible longueur. Cependant les segments sécants et broyeur sont comparables à ceux des loups actuels.

Ce sont les restes présents sur le site du Santenay (Bourgogne, France) qui ont permis à A. Argant de décrire la deuxième sous-espèce : *Canis lupus santenaisiensis* (Argant, 1989). Ses



représentants ont les dents plus longues et plus épaisses que *C. l. lunellensis*, et leur stature est plus imposante. Cette sous-espèce marque un stade intermédiaire avec des individus dont la taille est comprise entre celle des *C. lupus lunellensis* et celles des *C. lupus* actuels. Des restes retrouvés sur le site du Lazaret avaient été décrits comme appartenant à une nouvelle sous-espèce, *Canis lupus mediterraneus* (Patou, 1984), avant d'être mis en synonymie avec *C. lupus santenaisiensis* (Valensi, 1994).

La diagnose de la troisième sous-espèce a été réalisée à partir des restes exhumés sur les sites de Jaurens et de Mالدیدier (Aquitaine, France). Ses représentants possèdent une stature beaucoup plus imposante que celles des *C. lupus santenaisiensis* ou que celles des loups modernes (Boudadi-Maligne, 2012). L'accroissement de la stature chez cette sous-espèce s'accompagne d'une augmentation globale de la taille des dents (carnassières supérieures et inférieures ainsi que la UM1) et celle des segments sécants et broyeurs inférieurs. Le calcaneus subit lui aussi une augmentation de taille. Les datations de certains ossements par la méthode radiocarbone montrent que cette sous-espèce était présente à la fin du Pléistocène supérieur.

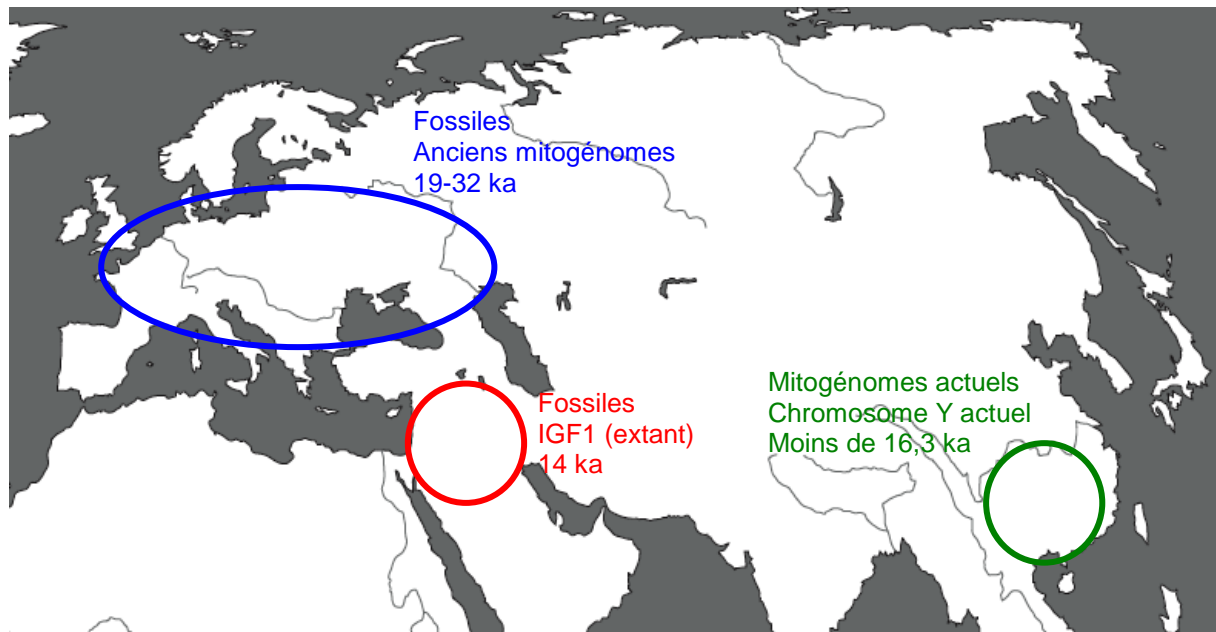
L'augmentation constante de la taille chez les loups du Pléistocène pourrait être due à des changements climatiques. Une corrélation est possible entre l'émergence des trois sous-espèces avec les trois épisodes de refroidissement climatiques qui ont eu lieu au Pléistocène moyen et inférieur. Ainsi, l'apparition de *C. l. lunellensis* serait liée à la glaciation qui a eu lieu pendant l'Elstérien (Pléistocène moyen) ; l'épisode climatique rigoureux du Saalien (toujours au Pléistocène moyen) aurait vu l'émergence de *C. l. santenaisiensis* ; et enfin le refroidissement qui s'est produit au Weichselien (Pléistocène supérieur) aurait entraîné l'avènement de *C. l. maximus*.

Ces différences de stature pourraient également être liées à des changements dans la faune, comme par exemple le remplacement de la faune du Galérien par celle de l'Aurélien pour *C. l. lunellensis*. En association avec les ossements de *C. l. maximus*, sont retrouvés de très nombreux fossiles d'animaux spécifiques des climats froids (comme le renne). La forte disponibilité en proie aurait pu avoir un rôle dans l'augmentation de la stature de ces individus en parallèle de celui joué par le climat rigoureux (Boudadi-Maligne, 2010). À l'issue du Pléistocène supérieur, une diminution de la taille des loups est observable jusqu'à ce qu'elle atteigne celle qu'on leur connaît aujourd'hui.

### 2.1.2.3. *Canis lupus familiaris*

En parallèle de ses cousins sauvages et sous l'action de l'homme, est apparue durant le Pléistocène supérieur une sous-espèce bien particulière : *Canis lupus familiaris*. Les premières hypothèses qui tentaient d'expliquer la domestication proposaient que les chacals aient été utilisés lors de la domestication mais ces théories n'ont plus cours. Tout comme celles qui envisageaient que plusieurs espèces sauvages avaient été croisées afin d'aboutir aux chiens actuels. Il est maintenant communément admis que les chiens domestiques dérivent des loups gris. Malgré cette certitude, la localisation ainsi que la date de la domestication restent fort incertaines. De nombreuses études ont été réalisées afin de répondre à cette interrogation mais trois hypothèses sont toujours en débat (Figure 18).

La première de ces hypothèses suggère que les chiens dérivent des loups du Moyen-Orient, et situe la domestication vers - 14 000 ans. Cette théorie est basée sur de nombreux fossiles ainsi que sur une étude du gène nucléaire IGF-1 montrant que les chiens de petites tailles possèdent les mêmes mutations que les petits loups du Moyen-Orient (Gray *et al.*, 2010). La seconde hypothèse est que la domestication aurait eu lieu en Asie du Sud-Est et plus exactement dans la région située au sud de la rivière Yangtze, il y a au moins 16 300 ans. Cette théorie est basée sur l'analyse de la diversité mitochondriale qui est la plus variée dans cette région, et sur les haplogroupes du chromosome Y dont deux sont uniques à cette région (Pang *et al.*, 2009 ; Ding *et al.*, 2012). La troisième hypothèse plaide pour une domestication européenne du loup entre 19 000 et 32 000 ans. Elle est basée sur l'analyse d'anciens génomes mitochondriaux partiels ou complets (datés entre 36 000 ans et l'actuel). Les analyses phylogénétiques montrent que les chiens actuels sont plus proches de ces anciens canidés que des loups modernes (Thalmann *et al.*, 2013).



**Figure 18.** Schéma récapitulatif des différents lieux et intervalles de temps où auraient pu avoir lieu la domestication du loup.

Ces trois hypothèses et toutes les études qui en découlent montrent clairement qu'il est nécessaire de mieux connaître les loups du Pléistocène à partir desquels sont issus les chiens modernes si l'on veut comprendre le processus de domestication.

### 2.1.3. Mode de vie des *Canis lupus*

Les loups gris sont des animaux ubiquistes qui ont su s'adapter à presque tous les climats (ils sont présents sur les continents nord-américain et eurasien) et à des biotopes variés (toundras, steppes, prairies ou encore forêts). Le nombre d'individus dans une région est lié au climat, au nombre de proies disponibles et également au degré d'anthropisation du milieu.

#### 2.1.3.1. Espace de vie des loups

Les loups vivent en meute qui se répartissent sur un espace vital, chaque meute occupant une portion exclusive de celui-ci appelée territoire. Chaque meute est séparée des autres par une distance de plusieurs dizaines de kilomètres où évoluent les loups dits « solitaires » (Landry, 2006). La superficie du territoire d'une meute varie en fonction des contraintes du milieu, du nombre de proies disponibles, de la taille de la meute et parfois de la saison (Prucca, 2003). Ainsi, la taille d'un territoire peut varier entre cinquante et deux milles

kilomètres carré. Lorsque qu'une meute choisit un territoire, elle fait attention à ce que plusieurs éléments indispensables soient présents. Un espace couvert est requis afin de se protéger des prédateurs et trouver des tanières où les femelles en gestation et les jeunes louveteaux pourront résider. Un espace ouvert où les proies pourront être chassées est lui aussi nécessaire, ainsi qu'un point d'eau.

### ***2.1.3.2. Organisation des meutes de loups***

Une meute de loups est soumise à une hiérarchie bien précise. Elle comporte par ordre hiérarchique : un couple d'alphas qui sont les dominants et uniques reproducteurs, un loup bêta qui pourra à l'occasion se reproduire avec la femelle dominante, des subordonnés (mâles et femelles) et enfin un mâle oméga qui est le souffre-douleur de la meute. Les louveteaux et jeunes de moins d'un an ne sont pas soumis à la hiérarchie jusqu'à ce qu'ils aient atteint la maturité sexuelle (Landry, 2006). La taille de la meute peut varier en fonction de plusieurs facteurs : la taille des proies disponibles, la région ou la saison. Un autre facteur est la pression exercée par les dominants, si celle-ci est trop forte, des subordonnés peuvent quitter la meute afin d'en rejoindre une autre. En Europe, les meutes comprennent entre deux et sept loups alors que sur le continent nord-américain la moyenne est plutôt d'une dizaine d'individus.

Sur le plan de l'alimentation, les loups sont des animaux communautaires qui chassent en groupe. Cette technique leur permet d'attraper des proies de taille plus importante. La taille des animaux chassés dépend également de celle de la meute. Ce sont principalement des ongulés de taille moyenne qui sont consommés mais également des petits gibiers (lagomorphes, oiseaux, petits carnivores). Ils peuvent aussi consommer des végétaux ou des carcasses quand le nombre de proies devient insuffisant.

### ***2.1.3.3. Des données actuelles à prendre avec précaution***

Comme toutes ces observations ont été réalisées sur des spécimens actuels, il est très difficile d'affirmer que les loups du Pléistocène suivaient exactement le même mode de vie. Les tanières occupées par les loups actuels sont toujours vides d'ossements et de fèces car les louveteaux ingèrent de la nourriture prédigérée et leurs mères avalent les excréments de ces derniers. De plus, les louveteaux morts dans les tanières sont mangés par un individu de la meute et ne laissent donc pas de traces. C'est pour cette raison sans doute qu'aucune tanière

du Pléistocène n'a pu être découverte. Il faut également garder à l'esprit que l'éthologie a pu évoluer ; en effet, le degré d'anthropisation était bien moindre au Pléistocène (chasse et occupation du territoire) et le contexte de compétition entre les espèces, très différent.

## 2.2. L'hyène des cavernes

Le deuxième animal dont les coprolithes ont été étudiés dans cette thèse est l'hyène des cavernes. Il s'agit de l'un des carnivores les mieux connus de l'ère glaciaire grâce à des fossiles retrouvés dans de nombreuses cavités. Cet animal a également été d'une grande aide pour les Paléontologues en ramenant les os de ses proies dans ces mêmes grottes. C'était un grand prédateur des herbivores du Pléistocène et un compétiteur de l'Homme pour l'accès aux ressources carnées.

Mis à part le protèle (*Proteles cristata*), un hyénidé d'Afrique qui se nourrit principalement de termites, toutes les hyènes actuelles (*Crocuta crocuta*, *Hyena hyena* et *Parahyaena brunnea*) étaient présentes en France au Pléistocène. Dans les assemblages osseux des cavernes, l'hyène tachetée est l'espèce la plus représentée ; ce qui a conduit à la définition d'une espèce : l'hyène des cavernes, *Crocuta crocuta spelaea*. L'hyène des cavernes est très peu voire pas représentée dans l'art mobilier et pariétal. Deux œuvres sont discutées : un objet en ivoire de la grotte de la Madeleine et l'animal tacheté représenté en profil droit dans la Grotte Chauvet-Pont d'Arc sur le panneau de la panthère. Elles pourraient correspondre respectivement à un bovidé et à un ours des cavernes.



**Figure 19.** Sculpture probable d'une hyène des cavernes sur un propulseur en ivoire

### 2.2.1. Origine évolutive de l'hyène des cavernes

Les quatre espèces d'hyènes actuelles (représentant 4 genres différents) sont les reliquats d'une famille autrefois très diversifiée. En effet, des fossiles d'environ cent espèces différentes ont été retrouvés. La famille des hyénidés est la plus jeune de l'ordre des carnivores. Elle a émergé de la famille des viverridés, il y a environ 25 millions d'années à la fin de l'Oligocène (Kurtén, 1968). C'est à la fin du Miocène que l'on retrouve la plus grande diversité avec vingt espèces qui cohabitent. A part le protèle qui est très spécifique, les trois espèces restantes se ressemblent beaucoup, ce qui suggère une divergence des lignées très récente.

L'origine de l'hyène des cavernes est soumise à débat. Selon une étude basée sur des séquences d'ADN ancien extraites de fossiles, les trois haplogroupes retrouvés chez les spécimens européens et asiatiques seraient dus à trois événements de migrations indépendants d'une population ancestrale depuis l'Afrique (Rohland *et al.*, 2005). Cependant, les fossiles retrouvés dans les régions concernées sont trop jeunes pour soutenir cette théorie. En effet, les fossiles les plus anciens sont retrouvés en Chine. Cette hypothèse est en corrélation avec celle soutenue par B. Kurtén. Selon lui, l'hyène des cavernes descend d'une hyène dont l'aire de répartition se situait en Inde au Villafranchien : *Crocota sivalensis*. Celle-ci possédait une dentition primitive avec la présence de molaires supérieures qui vont devenir vestigiales et disparaître dans les espèces modernes. Pendant le Pléistocène moyen, *Crocota sivalensis* quitte son aire d'origine et se répand en Afrique, en Chine et en Europe (Kurtén, 1968).

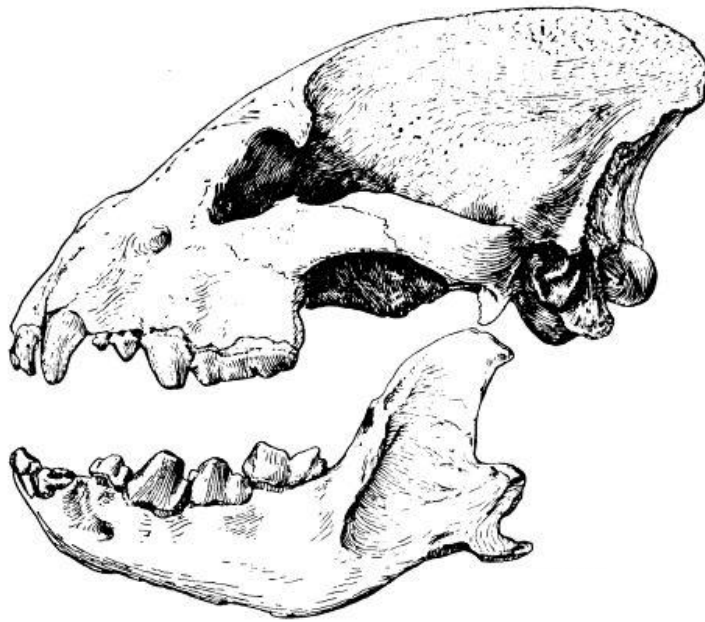
Les premières hyènes du genre *Crocota* apparaissent en Europe pendant la glaciation du Günz. Le plus vieux fossile d'hyène des cavernes est observé à la Sierra d'Atapuerca et est daté à 0,78 million d'années. L'hyène a eu deux chemins possibles pour arriver en Europe : à partir de la péninsule arabique (où elle est détectée il y a 1 million d'années), elle a pu soit traverser le Bosphore quand les glaciations le permettaient, soit contourner la Mer Noire par le Caucase (où des restes de *Crocota sinensis* ont été retrouvés au Pléistocène inférieur). À la fin de l'ère glaciaire, avec le développement de l'agriculture, le nombre d'hyènes des cavernes s'est peu à peu réduit jusqu'à sa disparition totale d'Europe et d'Asie (Kurtén, 1968).

La forte ressemblance entre l'hyène tachetée actuelle d'Afrique et l'hyène des cavernes divise les paléontologues qui cherchent à établir quelle est leur relation taxonomique. Pour certains, l'hyène des cavernes est une sous-espèce de l'hyène tachetée (*Crocuta crocuta spelaea*) (Kurtén, 1968 ; Turner, 1984 ; Werdelin et Solounias, 1991) alors que pour d'autres, il s'agit d'une espèce distincte (*Crocuta spelaea*) (Markova *et al.*, 1995 ; Baryshnikov, 1999) ; une troisième hypothèse la présente comme la forme ancestrale de l'hyène tachetée actuelle d'Afrique (*Crocuta crocuta*). Cette dernière hypothèse est soutenue par une étude récente (Sheng *et al.*, 2013) basée sur l'analyse de portions de génomes mitochondriaux anciens d'hyènes des cavernes et ceux d'*Hyenidae* actuels. Cette étude montre que la population ancestrale eurasiatique d'hyène des cavernes se serait séparée en plusieurs lignées ancestrales dont deux auraient migré en Afrique et donné les hyènes tachetées actuelles.

Un gradient de taille continu est observé entre les fossiles qui sont retrouvés en Europe, au Sud de la Russie, au Proche-Orient et les spécimens africains. Des restes d'animaux de même taille que les hyènes tachetées actuelles sont retrouvés sur les sites de Süssenborn et Gombaseck (Kurtén, 1968). Quelques centaines de milliers d'années plus tard, pendant le Cromérien, l'hyène des cavernes atteint sa taille distinctive ; elle est 25% plus grande que l'hyène tachetée actuelle d'Afrique.

### **2.2.2. Morphologie de l'hyène des cavernes**

L'hyène des cavernes est le plus spécialisé des hyénidés dans le broyage des os. En effet, elle a développé de puissantes molaires et prémolaires de structure conique afin d'extraire la moelle et les tissus spongieux des os de ses proies. Leurs dents carnassières sont allongées et forment des lames aiguisées ; les canines en revanche sont peu imposantes (Kurtén, 1968).



**Figure 20.** Crâne d'hyène des cavernes. D'après (Kurtén, 1968).

Malgré leur ressemblance, l'hyène des cavernes diffère de l'hyène tachetée actuelle par sa taille ainsi que par l'allongement de son fémur et de son humérus dans les stades tardifs, doublé par un épaissement des pattes antérieures et postérieures. La morphologie du crâne ainsi que la structure des dents sont également un peu différentes entre les deux spécimens. Comme chez sa cousine africaine, le dimorphisme sexuel est peu marqué. Les femelles peuvent développer des pseudo-pénis par allongement de leur clitoris (Fosse, 1997).

### **2.2.3. Mode de vie de l'hyène des cavernes**

Comme leur nom l'indique, les spécimens de l'espèce *Crocota crocuta spelaea* vivaient dans des grottes appelés repaires. Le premier décrit est la grotte de Kirkdale dans le Yorkshire par William Buckland en 1823. Dean Buckland a ensuite montré que l'importante accumulation d'ossements d'hyène (des juvéniles aux séniles) ainsi que ceux de leurs proies était dû à une utilisation très prolongée de la cavité. Un autre repaire très connu est la grotte de Tornewton dans le South Devon où 200 000 dents d'hyènes ont été retrouvées (Kurtén, 1968).

Ces repaires présentent globalement tous la même topologie. Ce sont des cavités avec plusieurs d'entrées (au minimum 2), qui conduisent à une vaste salle basse de plafond



(souvent moins d'un mètre) où les animaux étaient traînés pour être mangés ; c'est là que sont retrouvés la plupart des restes osseux de proies. De cette salle partent des tunnels étroits qui conduisent à des diverticules où se passent les mises-bas et où sont concentrés des ossements d'hyènes. Dans ces nurseries, on trouve des galeries creusées par les jeunes hyénons afin de se protéger de l'appétit des mâles adultes. Les vestiges osseux de proies sont riches aux abords du repaire et dans les zones proches de l'entrée. Les zones de défécation se situent souvent à l'extérieur. Mais des coprolithes sont parfois retrouvés dans les repaires. Ces excréments fossilisés permettent de reconstituer ponctuellement l'environnement végétal des hyènes, car ils concentrent les pollens issus de l'eau de boisson ou de l'ingestion des végétaux contenus dans les viscères des herbivores ingérés. Sutcliffe observe une différence de concentration des coprolithes dans les repaires en fonction du climat. Les grottes à faune tempérée contiennent beaucoup de coprolithes et d'ossements et sont de taille réduite (plus de 400 pièces à Lunel-Viel) alors que les cavités à faune froide comme celle du Würm ancien en contiennent beaucoup moins et sont de taille plus imposante (Fosse, 1997).

Les hyènes tachetées africaines actuelles utilisent également des repaires dont la taille varie en fonction de la durée d'occupation. Les repaires de petite taille vont être utilisés pendant quelques semaines ou quelques mois alors que les repaires de grande taille peuvent être occupés jusqu'à plusieurs années. Les portées comprennent un à deux petits qui sont élevés dans des repaires de petite taille uniquement par la cellule familiale. Lors de leur cinquième année, les jeunes sont réunis et élevés collectivement.

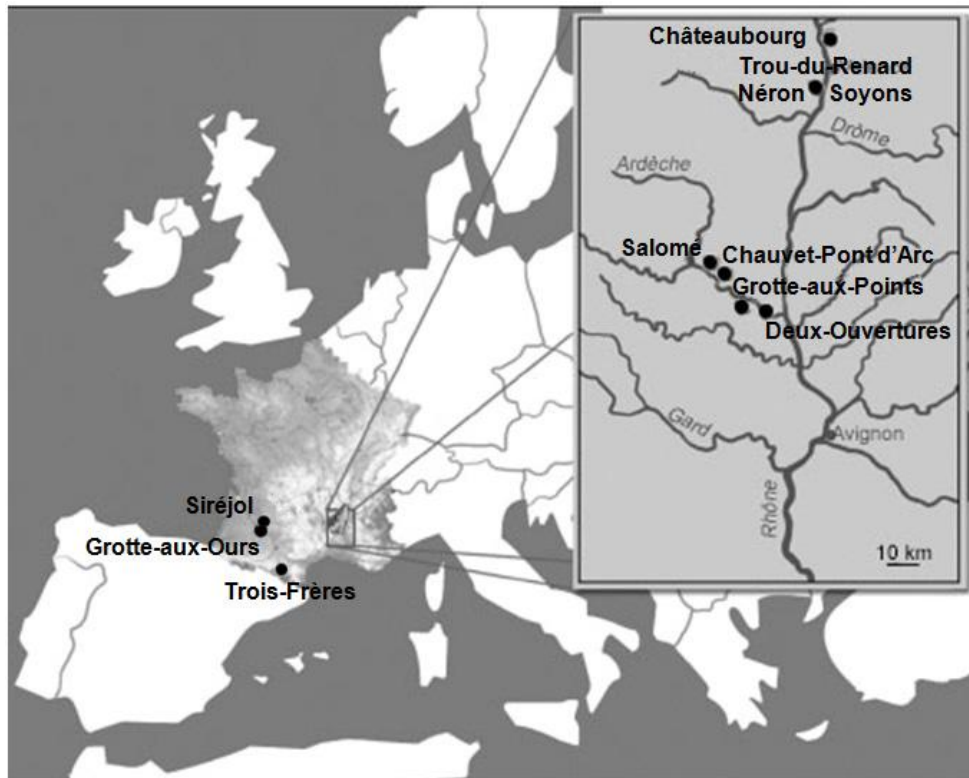
Dans les repaires d'hyène des cavernes, il est possible de distinguer deux types d'assemblage osseux : soit une prédominance d'artiodactyles (cerf élaphe, renne) et de grands bovidés (aurochs, bison), soit une prédominance de périssodactyles (équidés, rhinocéros) et proboscidiens (mammouth). Il est à noter que dans ces assemblages osseux, l'hyène est souvent l'animal le mieux représenté (Fosse, 1997).

Les hyènes du genre *Crocuta* se nourrissent principalement de proies de grande et de moyenne taille. Contrairement aux autres espèces d'hyènes qui vivent quasi exclusivement de charognage, l'hyène tachetée tue environ 70% des animaux qu'elle consomme (parfois jusqu'à 90%). Elle chasse en groupe afin d'avoir un meilleur taux de réussite, de plus grosses proies et une meilleure protection du butin de chasse contre les prédateurs. C'est le nombre et la disponibilité des proies qui va définir la taille de la meute. Le régime alimentaire de l'hyène

des cavernes varie en fonction du climat. Lors des périodes interglaciaires, les milieux boisés rendent la chasse difficile. Cela entraîne un plus fort taux de charognage et une plus grande mortalité chez les jeunes. En revanche, lors des périodes glaciaires les grands espaces ouverts, comme le paysage de la toundra, favorisent la chasse.

### 3. LES GROTTES D’OÙ SONT ISSUS LES ÉCHANTILLONS

Grâce à leur température fraîche et constante, correspondant à un milieu favorable pour la préservation de l’ADN. Dans cette étude, nous avons analysé des coprolithes et des ossements provenant de onze sites repartis dans trois régions de France : l’Ardèche, le Lot et l’Ariège (Figure 21).



**Figure 21.** Localisation des sites d’où proviennent les échantillons étudiés.

#### 3.1. Les grottes des gorges de l’Ardèche

Le passage de la rivière Ardèche a creusé un véritable canyon sur plus de trente kilomètres dans le plateau de calcaire urgonien du Bas-Vivarais. La rivière dans son chemin pour rejoindre la vallée du Rhône a également fait apparaître de nombreuses grottes et galeries à l’intérieur des falaises. Ces cavités ont servi de refuge aux animaux (hyènes et ours des cavernes) ; ainsi une quinzaine de grottes ont été référencées comme "grotte à ours" (Philippe, 2005). Elles ont également été occupées par les Hommes qui en ont orné certaines.

La présence humaine dans les gorges de l'Ardèche a commencé il y a environ 300 000 ans. Cette région est très riche en sites ornés, de peintures ou de gravures, car elle en compte une vingtaine répartis sur une période comprise entre - 37 000 et - 12 000 ans BP.

La Grotte Chauvet-Pont d'Arc est l'une des rares grottes ornées datées de l'Aurignacien (- 37 000 à - 29 000 ans BP). En effet, il y a très peu de vestiges de cette époque dans la région, à part une sagaie à base fendue retrouvée à l'Abri des pêcheurs et datée de - 26 760 ans BP (Clottes, 2001).

Les sites de l'époque gravettienne (- 27 000 à - 26 000 ans BP) sont plus nombreux avec six cavités ornées (dont la grotte Chauvet-Pont d'Arc) et un site en plein air (celui de Bouzil).

Il y a une dizaine de sites référencés pour le solutréen (- 23 000 à - 18 000 ans BP) qui est une époque caractérisée par un climat rigoureux. Les grottes de la Tête du Lion, de Chabot, des Deux-Ouvertures et la Grotte Sombre sont des grottes ornées de gravures ou de peintures qui lui sont attribuées.

C'est pour l'époque magdalénienne (- 17 000 à - 10 500 ans BP) qu'il y a le plus de traces d'occupation humaine, avec une trentaine de sites authentifiés. L'outillage en silex retrouvé est très diversifié, alors que l'industrie osseuse et l'art mobilier sont très peu développés.

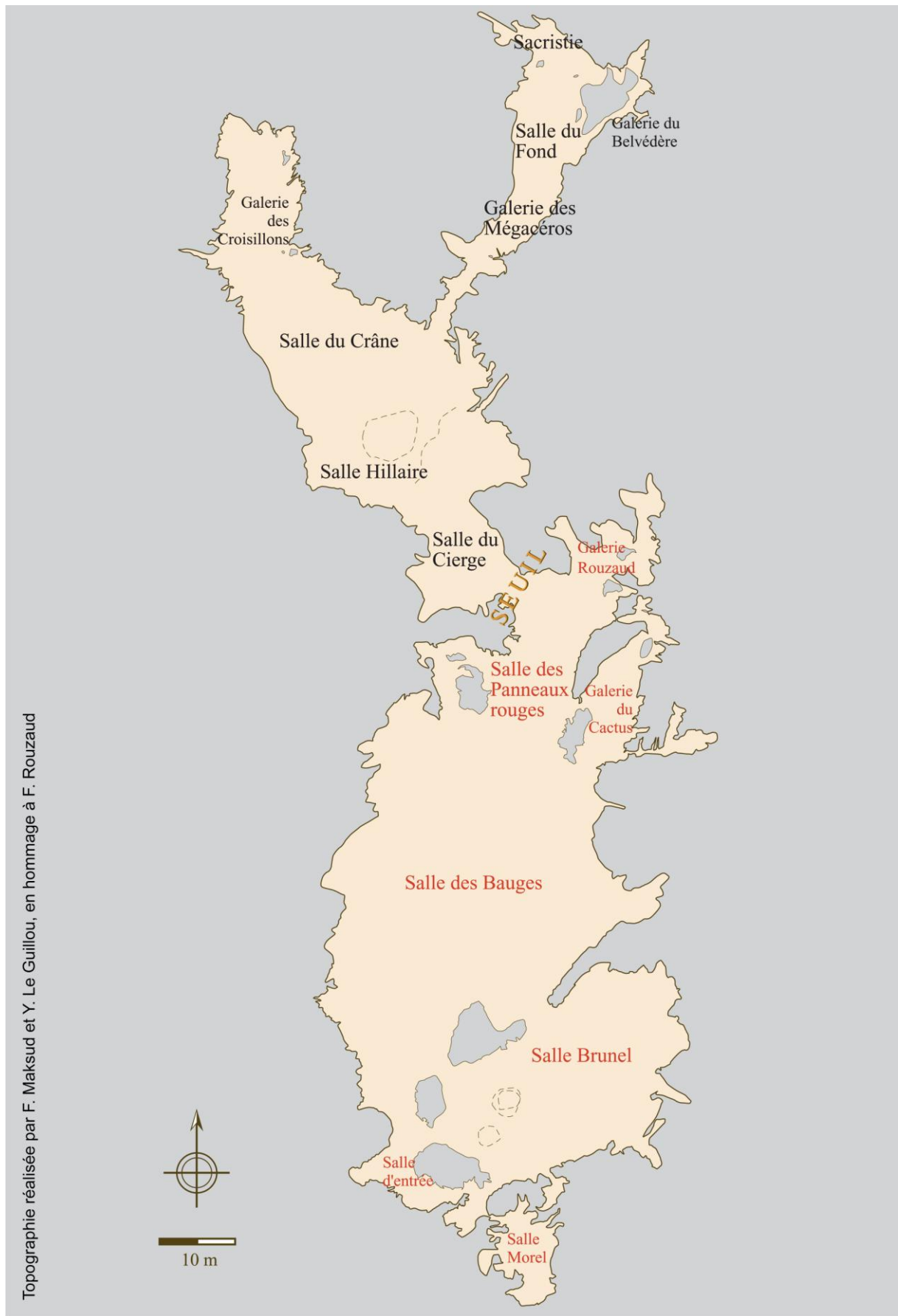
### **3.1.1. La grotte Chauvet-Pont d'Arc**

#### ***3.1.1.1. La découverte et la protection de la Grotte Chauvet-Pont d'Arc***

La grotte Chauvet a été découverte le 18 décembre 1994 au cours d'une prospection privée par E. Brunel, C. Hillaire et J.-M. Chauvet. Ce dernier avait repéré un souffle d'air émanant du fond d'une petite cavité déjà connue s'ouvrant dans le méandre abandonné de l'Ardèche face au Pont d'Arc. Avec l'aide de ses deux compagnons, il dégage le fond de cette cavité et ils descendent à pic dans le noir ; ils arrivent alors dans une salle aux concrétions abondantes. Au cours de leur visite, ils ne repèrent pas tout de suite les dessins paléolithiques ; c'est uniquement sur le chemin du retour qu'E. Brunel aperçoit le petit mammoth rouge dit « de la découverte ». Quelques jours plus tard les inventeurs, déclarent la grotte à la Direction Régionale des Affaires Culturelles qui envoie Jean Clottes afin

d'expertiser les peintures. L'existence de la grotte sera dévoilée au public le 18 janvier 1995, juste un mois après sa découverte. Elle a été classée au patrimoine historique en 1995 et plus récemment au patrimoine mondial de l'UNESCO (en juin 2014). Une réplique a été réalisée afin de permettre au monde entier d'admirer les représentations picturales de la grotte qui est fermée au public par souci de préservation.

Grâce à la prudence dont ont fait preuve les inventeurs, la grotte Chauvet-Pont d'Arc est en effet dans un état de préservation exceptionnel. Dès leur deuxième exploration, ils ont pris soin de recouvrir les traces de leur premier passage de lais en plastique afin de baliser un chemin et d'éviter que d'autres zones du paléosol ne soient abîmées. Le trou des inventeurs a été élargi et une porte blindée a été posée afin de réglementer l'accès à la grotte. Les lais de plastique ont été remplacés par des passerelles métalliques inoxydables (fixes ou mobiles). Toujours en vue de préserver l'état originel de la grotte, la climatologie de celle-ci est surveillée en permanence par des appareils contrôlant l'hygrométrie, la température de l'air, du sol et des parois, ainsi que le taux de dioxyde de carbone et le radon (Baffier, 2005). L'ensemble de ces mesures a permis de conserver la grotte dans l'état où elle se trouvait lors de sa fermeture définitive, il y a 22 000 ans (Sadier *et al*, 2012).



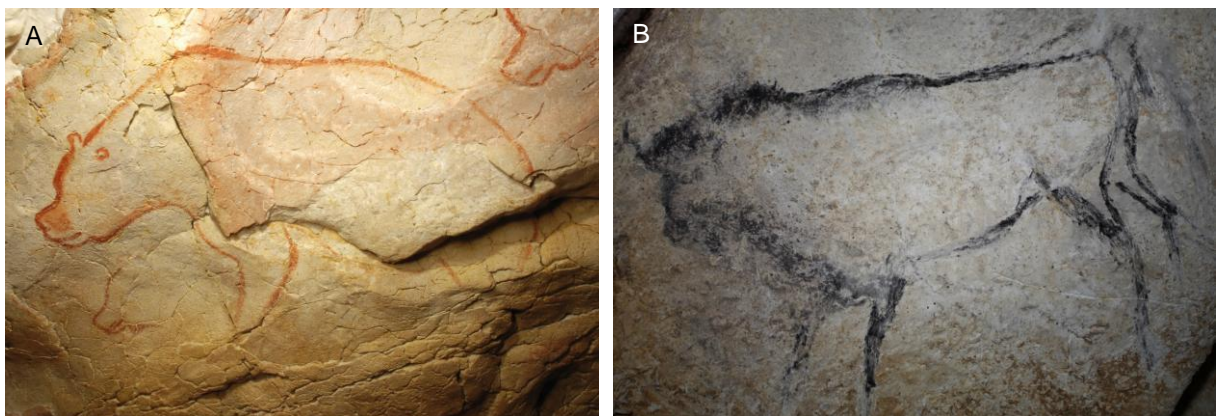
**Figure 22.** Topographie de la Grotte Chauvet-Pont d'Arc.

### 3.1.1.2. *L'Homme et la Grotte Chauvet-Pont d'Arc*

#### 3.1.1.2.1. *Les représentations picturales*

La renommée de la grotte Chauvet-Pont d'Arc est liée aux exceptionnelles peintures et gravures qui y sont présentes, et qui font partie des plus vieilles au monde. Les divers pigments inorganiques (ocres rouge et jaune) et organiques (charbon de bois) ainsi que les multiples techniques utilisées (raclage de paroi, estompe, gravure au doigt ou avec un outil pour accentuer un dessin) révèlent une technicité artistique déjà très développée (Tosello & Fritz, 2005).

La couleur des dessins divise la grotte en deux secteurs, séparés par le passage étroit du seuil. Les représentations effectuées avec des pigments rouges dans les zones proches de l'entrée, ont valu à celles-ci le nom de « Secteur rouge ». Les zones plus profondes de la cavité contiennent, quant à elles, des dessins réalisés au charbon de bois : c'est le « Secteur noir ». La grotte Chauvet-Pont d'Arc comporte près d'un millier de peintures et de gravures dont 420 sont des représentations animales.



**Figure 23.** Représentations picturales de la Grotte Chauvet-Pont d'Arc. A : Ours des cavernes à l'ocre rouge dans le Diverticule des Ours (© MCC/CNP). B : Bison au charbon de bois dans la Galerie des Mégacéros (© MCC/CNP).

Les animaux les plus présents sont des animaux dangereux pour l'Homme tels que des mammouths, des rhinocéros et des félidés (lion des cavernes ou panthère). Il y a également des représentations de cervidés, de chevaux, d'ours des cavernes, de bovidés (aurochs, bison ou bouquetin) ainsi qu'une unique gravure de hibou.

Un dessin très particulier, représentant la partie inférieure d'un corps féminin et une créature mi-homme mi-bison, a été réalisé au charbon sur un pendant de la salle du fond. Cette chimère peut être mise en parallèle avec une représentation de la Grotte du Gabillou. L'abondance de représentations de félins quant à elle semble être une caractéristique de l'art aurignacien, comme le montre des statuettes du Jura souabe de la même période. Des mains positives et négatives ainsi que de nombreux symboles et triangles pubiens sont également présents un peu partout dans la Grotte Chauvet-Pont d'Arc.

La chronologie entre les deux secteurs est difficile à établir en raison de l'impossibilité de dater les pigments organiques rouges. Cependant, la présence de marques rouges sous les dessins noirs dans la salle du fond indique que certains dessins rouges pourraient être très anciens. Les datations par la méthode du radiocarbone sur les dessins réalisés au charbon de bois (Cuzange *et al.*, 2007 ; Valladas *et al.*, 2001) ont permis de mettre en évidence deux périodes de fréquentation de la cavité par l'Homme : une aurignacienne (entre - 32 000 BP et - 30 000 BP) et une gravettienne (entre - 27 000 BP et - 25 000 BP). Jusqu'à présent, tous les dessins datés correspondent à la période aurignacienne. La fréquentation gravettienne est attestée par des charbons au sol ou des traces charbonneuses sur les parois.

#### ***3.1.1.2.2. Les autres traces du passage de l'Homme***

En plus des multiples dessins et gravures, la présence de l'Homme dans la grotte Chauvet-Pont d'Arc est attestée par de nombreuses traces et divers vestiges.

Sur certaines parois, les Hommes ont frotté leurs torches afin d'en raviver la flamme ; ces mouchages de torches, dont beaucoup remontent au Gravettien, ont laissé des traces noires un peu partout dans la cavité. De plus, par endroits, le sol de la grotte est composé de banquettes d'argile qui ont enregistré le passage des Hommes. Ainsi, entre la Salle du Crâne et la Galerie des Croisillons, l'ichnologie (l'étude des traces et des empreintes) a permis de repérer une piste de 70 mètres laissée par un adolescent pré-pubère (Garcia, 2005). Deux traces de main, pleine d'argile, visibles près des mammoths raclés, lui ont aussi été attribuées, ainsi que des mouchages de torche réguliers au-dessus de la piste. L'analyse des charbons de ces derniers a permis de proposer une date de son passage à - 25 550 BP, correspondant à la deuxième phase de fréquentation de la grotte.





**Figure 24.** Les traces du passage de l'Homme dans la Grotte. A : Mouchage de torche dans la Galerie des Mégacéros. B : Foyer dans la Salle des Bauges. C : Crâne d'ours des cavernes dans la Salle du Crâne. (© MCC/CNP)

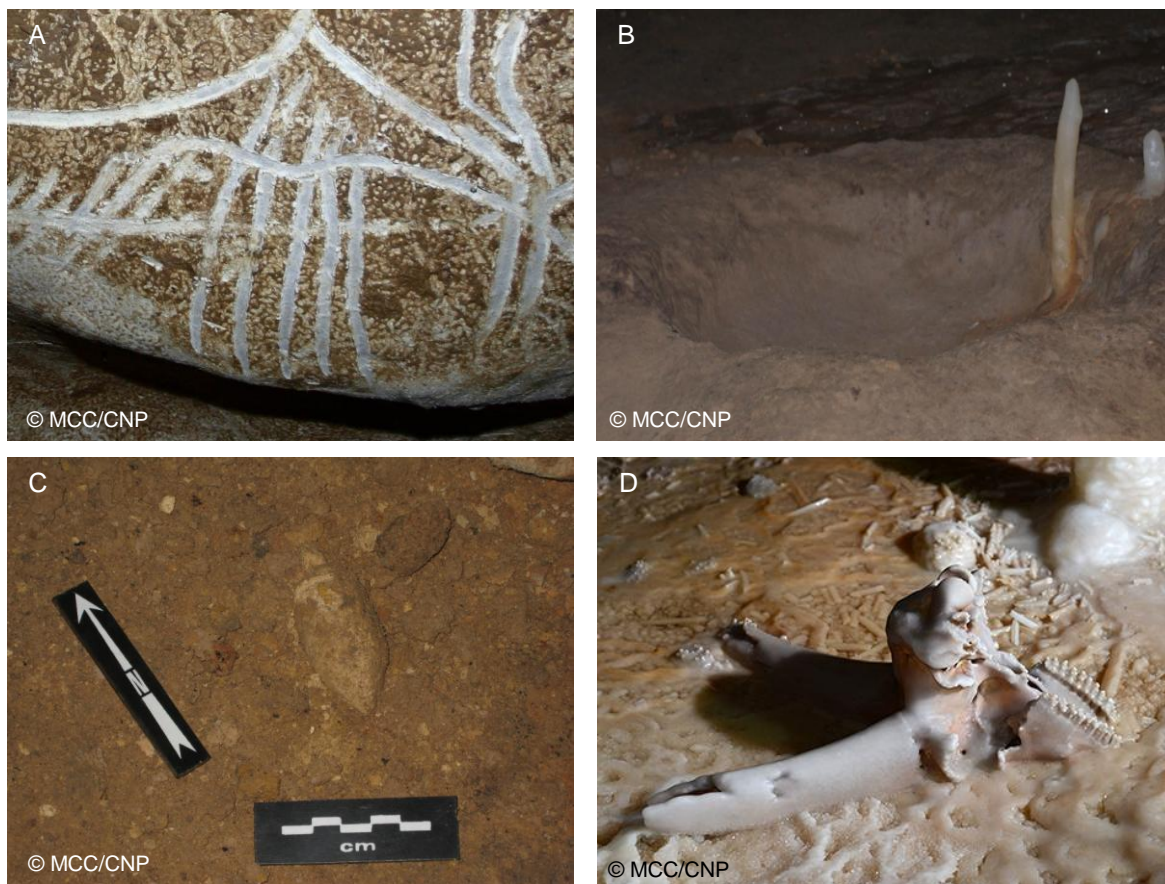
Un autre type de vestiges laissés par l'Homme est la présence de foyers dans les parties terminales de la cavité, principalement dans la Galerie des Mégacéros ; ils ont été datés grâce à leurs charbons de bois à - 32 000 ans BP. Ils auraient servi à la réalisation des dessins de la Salle Hillaire, de la Galerie des Mégacéros et de la Salle du Fond (Clottes, 2001). Fait remarquable, les charbons de bois prélevés dans les mouchages de torches, les foyers ou encore les dessins, proviennent tous d'une même espèce d'arbre : le pin sylvestre (*Pinus sylvestris*). Les caractéristiques de cet arbre par rapport aux autres espèces sont des flammes plus vives et plus durables, une production plus importante de charbon et la facilité d'acquisition due à un élagage naturel important. Ce sont les raisons de son utilisation unique dans la cavité (Théry-Parisot & Thiebault, 2005).

Enfin, la disposition de certains ossements d'ours suggère une manipulation par l'Homme, comme le crâne posé sur un bloc dans la salle dite « du Crâne ». Tout autour sont accumulés des dizaines d'autres crânes d'ours des cavernes qui ont éventuellement été amenés par l'Homme eux aussi. Il y a également deux humérus d'ours plantés à la verticale, dans la zone d'entrée dont la signification n'a pas encore été totalement comprise.

### **3.1.1.3. Les traces animales dans la Grotte Chauvet-Pont d'Arc**

Les Hommes du Paléolithique ne sont pas les seuls à avoir fréquenté la grotte Chauvet-Pont d'Arc. La cavité est d'une immense richesse pour les paléontologues, car elle renferme plus de 4 000 ossements éparpillés sur le sol, principalement concentrés dans la Salle du Crâne, la Salle Hillaire et l'éboulis d'entrée (Fosse & Philippe, 2005).

Ces restes appartiennent à 99% à l'ours des cavernes (*Ursus spelaeus*). En effet, en plus d'être une grotte ornée, la cavité est également définie comme grotte à ours en raison de sa forte occupation par ce dernier. Le passage d'*Ursus spelaeus* est attesté par des polissages d'ours sur le bas des parois, des griffades sur les murs, notamment sur certaines peintures, et la présence de bauges creusées lors des longues périodes d'hivernation. Grâce à la méthode du radiocarbone des os en surface, le passage des ours dans la cavité est estimé entre - 37 400 et - 28 800 BP (Bocherens *et al.*, 2006 ; Bon *et al.*, 2012).



**Figure 25.** Les traces du passage des animaux dans la Grotte Chauvet-Pont d'Arc. A : Griffade d'ours sur un panneau orné dans la Salle Hillaire. B : Bauge d'ours des cavernes dans la Salle des Bauges. C : coprolithe attribué à un canidé dans la Salle Hillaire. D : Crâne de Bouquetin.

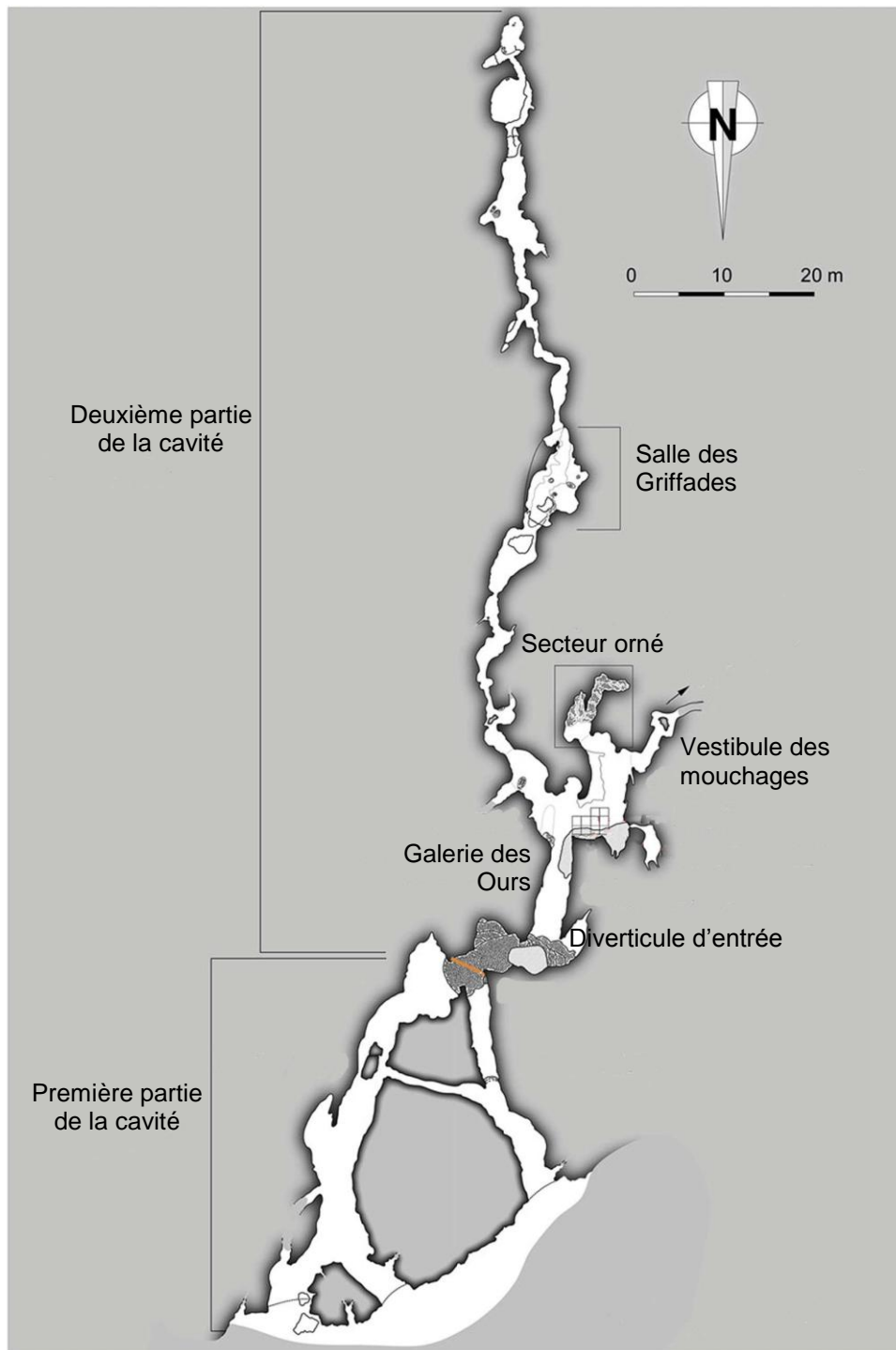
Comme pour les Hommes, les banquettes d'argile ont gardé la trace du passage des ours des cavernes, d'un bouquetin - dont une piste est visible dans la sacristie - et de canidés. Un grand canidé a laissé une piste à proximité de celle de l'adolescent (Garcia, 2005). Les empreintes de pattes laisseraient à penser qu'il s'agit plutôt d'un chien que d'un loup en raison de la morphologie des autopodes où les coussinets digitaux sont fortement engagés entre les latéraux. Toutefois, les deux pistes ne se recoupent jamais, il est donc impossible

d'affirmer que l'enfant et l'animal aient été présents au même moment dans la grotte. En revanche, les empreintes présentes dans le Secteur Rouge sont sans ambiguïté celles de loups. Elles ne forment pas de pistes mais des tapis dus au piétinement d'un grand nombre d'individus de tailles variées. En plus des empreintes, des ossements de loups sont présents dans la grotte, tels que des os longs, une hémi-mandibule, un métatarse, une phalange et deux vertèbres. Il y a également deux crânes, dont l'un est petit et pourrait appartenir à une femelle alors que le second est plus robuste. Le loup a également laissé des coprolithes comme trace de son passage dans les secteurs profonds de la cavité.

La présence mineure d'autres animaux est à noter, tels que le renard roux représenté par un squelette entier, la panthère, le chat sauvage et l'hyène des cavernes pour les carnivores ; le chevreuil, des cervidés, des bovidés et le cheval pour les herbivores.

### **3.1.2. La Grotte des Deux-Ouvertures**

Cette grotte, située à la sortie des gorges de l'Ardèche sur la rive gauche, est connue depuis très longtemps. Léopold Chiron, l'un des premiers préhistoriens français, en parlait déjà en 1896. Elle doit son nom à ses deux entrées l'une s'ouvrant à l'est et l'autre à l'ouest. Pourtant ce n'est qu'en décembre 1985 que C. Hillaire, F. Landraud, F. Monteil et M. Verdon font sauter le bouchon calcitique situé au fond de la partie connue, dégageant ainsi l'accès à la partie profonde du réseau. L'exploration de celle-ci révélera la présence de gravures paléolithiques. Des ossements d'ours des cavernes ainsi que des traces de son passage (bauges à ours, polis et griffades des parois) ont été remarqués par B. Gély et J.-L. Porte lors de leur examen de la grotte. Celle-ci sera classée au patrimoine historique en 1990 et fermée au public par la pose d'une porte blindée.



**Figure 26.** Topographie de la Grotte des Deux-Ouvertures, d'après (Philippe, 2009).

Les premières campagnes de fouilles ont commencé en septembre 1987 sous la direction de J.-L. Porte, avec pour objectif de dresser la topographie de la grotte, entreprendre un sondage stratigraphique, et procéder aux relevés des gravures. Si ces dernières ont été publiées rapidement, il faudra attendre une seconde campagne de fouilles en 2005 sous la direction de M. Philippe pour avoir des résultats sur les restes osseux de la grotte. La

topographie du réseau profond, entreprise en 1987 par J.-L. Porte, B. Gély et E. Dechamps, a été finalisée en 2007.

### ***3.1.2.1. Les occupations animales de la Grotte des Deux-Ouvertures***

A l'instar de la Grotte Chauvet-Pont d'Arc, la Grotte des Deux-Ouvertures est, elle aussi, une grotte à ours des cavernes.

Lors de la fouille d'une lentille fossilifère découverte au niveau du Croisement, près de 700 ossements ont été exhumés, dont environ 99% appartenaient à *Ursus spelaeus* ; les restes sont principalement ceux de jeunes oursons ou d'individus de moins de un an. Cette accumulation est due à l'occupation de la grotte par les ours des cavernes pendant les longues périodes hivernales.

La fréquentation de la Grotte des Deux-Ouvertures par l'hyène des cavernes est attestée par plusieurs coprolithes, un calcanéum, une carnassière supérieure droite et un métapode. Cependant, l'absence d'ossements digérés par des sucs permet de penser que la grotte ne faisait pas office de repaire pour l'hyène des cavernes.

Mises à part ces deux espèces, d'autres animaux ont fréquenté la grotte comme le loup, le renard, le cheval, le cerf ou le bouquetin. Il y a en revanche très peu de restes de microfaune.

### ***3.1.2.2. Les traces du passage de l'Homme dans la Grotte des Deux-Ouvertures***

Contrairement à la Grotte Chauvet, la Grotte des Deux-Ouvertures contient que des gravures mais pas de peintures. Elles sont quasiment toutes réparties dans une petite galerie et un diverticule situés au fond de la grotte, à quelques mètres du sondage stratigraphique (Gély & Porte, 1996). D'autres gravures ont néanmoins été observées en dehors de ce secteur (Philippe, 2009). Sur la soixantaine de représentations présentes dans le secteur orné, 52 ont été clairement identifiées. Ces gravures, attribuées au solutréen ancien, présentent un bestiaire varié avec des mammouths, des aurochs, deux bouquetins, un félin, un bison et un équidé. Il y a également des triangles pubiens, une figure féminine et un être composite qui n'a pas encore été clairement identifié.



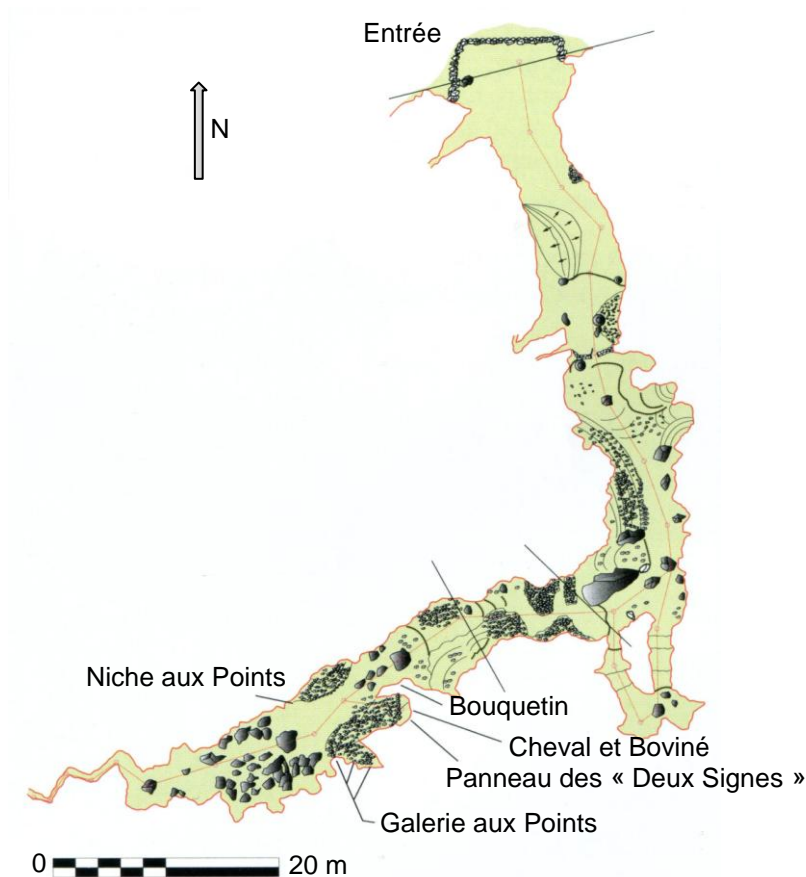
**Figure 27.** Gravure d'un bouquetin dans la Grotte des Deux-Ouvertures.

La présence de l'Homme est également observable par des mouchages de torches situés un peu partout sur les murs de la cavité. Deux ossements humains ont aussi été retrouvés dans le secteur orné (un radius et un humérus), où ils auraient pu être placés par les Hommes qui ont fréquenté la grotte au Paléolithique supérieur (Philippe, 2009)

### **3.1.3. La Grotte-aux-Points**

#### ***3.3.1.1. La découverte de la Grotte-aux-Points***

Située sur la rive droite de l'Ardèche près de la commune d'Aiguèze, la grotte s'ouvre à quatre-vingt-quinze mètres d'altitude sous un éperon de la falaise calcaire. Le porche d'entrée de la grotte a souvent servi d'abri-bergerie, comme en témoigne un muret en pierres sèches. Elle n'a toutefois été identifiée comme grotte ornée qu'en 1993. Le 7 novembre, lors d'une sortie spéléologique, J.-M. Chauvet, E. Brunnel et C. Hillaire repèrent des signes peints à l'ocre rouge dans les parties profondes de la cavité, anciennement appelée Grotte Yves. Une première expertise est réalisée par A. Colomer et P. Galant en 1993, puis J. Clottes sera sollicité en janvier 1995 pour une seconde authentification. La grotte est classée au patrimoine historique, depuis mai 1995 fermée au public et rebaptisée Grotte-aux-Points.



**Figure 28.** Topographie de la Grotte-aux-Points, dessin par M. Renda. D’après (Brunnel *et al.*, 2008).

La cavité s’ouvre par un porche relativement bas de trois mètres de haut pour treize de large. La grotte se présente sous la forme d’un couloir subhorizontal de cent-huit mètres de long orienté nord-sud puis est-ouest. Le paléosol de la grotte a été complètement bouleversé par ce qui semble être une recherche de terre et d’ossements à utiliser comme engrais. Des blocs ont été rejetés en amas un peu partout sur les bords du couloir délimitant clairement un cheminement central. Un remblai de paléosol est également présent dans le secteur d’entrée.

### ***3.1.3.2 Les représentations picturales de la Grotte-aux-Points***

La Grotte-aux-Points doit son appellation à la présence des nombreux points réalisés à l’ocre rouge qu’elle recèle. Ceux-ci, ainsi que les autres représentations artistiques se situent à environ soixante-quinze mètres de l’entrée. Les peintures de la grotte sont concentrées dans la partie la plus haute de plafond de la cavité, appelée la Rotonde.

La zone centrale du secteur orné est un panneau représentant deux grands signes dessinés par tamponnement et estompement à l’ocre rouge (Figure 29). Le dessin de gauche, qui est le mieux conservé, a la forme d’un champignon à chapeau convexe. Le deuxième signe est plus

grand et ressemble très fortement au symbole dit « du rapace » présent dans la Salle des Panneaux Rouges de la Grotte Chauvet-Pont d'Arc. Cette similarité permet l'hypothèse que les dessins des deux cavités sont contemporains.

À gauche du panneau des deux signes, sont représentés une tête de bouquetin ainsi qu'un arrière train de cheval et un bovidé tourné vers la gauche. De l'autre côté du panneau central de la grotte, se trouve la galerie des points. C'est un court diverticule de trois mètres de long où trois concentrations de points sont visibles. Sur le mur opposé, des points sont répartis de façon rayonnante autour d'un creux de la paroi : c'est la niche aux points. La répartition des dessins autour du panneau des deux signes donne à la grotte l'aspect d'un sanctuaire (Brunnel *et al.*, 2008).

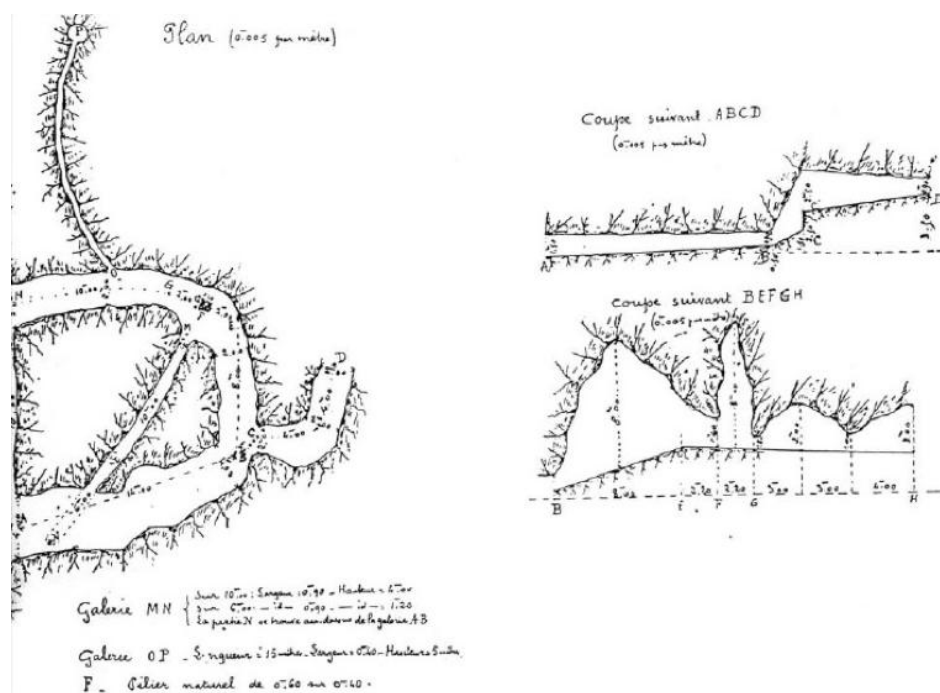


**Figure 29.** Panneau dit des deux-signes de la Grotte-aux-Points. (Brunnel *et al.*, 2008).

### **3.1.4 La Grotte-aux-Ours de Châteaubourg**

Située dans le canyon de la Goule en rive droite du Rhône, la grotte a été découverte en 1860 lors de l'ouverture d'une carrière dans les formations du Malm en vue de la construction de la digue du Rhône. Malgré quelques fouilles en 1871 et dans les années 1930, la quasi-totalité du remplissage karstique de la grotte a malheureusement été perdu (Figure 30). La mise à jour en 1995 d'une poche d'argile et de sable très fossilifère entraîne une fouille de sauvetage menée en urgence par B. Gély et G. Dal Pra en 1996. La grotte sera ensuite définitivement détruite par l'avancée de la carrière. La Grotte-aux-Ours était à l'origine constituée de hautes salles jouxtant de petites galeries qui s'inclinaient au gré des strates et s'ouvrait à cent cinquante mètres d'altitude.





**Figure 30.** Plan de la Grotte-aux-Ours de Châteaubourg. D'après (Bonnard et Goury, 1934).

De nombreux restes de faune du quaternaire et des silex attribuables au Néolithique ont été découverts en 1871 lors des fouilles menées par le Vicomte Lépici et J. de Lubac dans la deuxième salle de la grotte. Les autres chantiers de fouilles ont mis à jour des silex taillés, attribués au vu de leur faciès au Paléolithique moyen, principalement dispersés parmi les ossements d'ours, ainsi que des nucléus. La grotte doit d'ailleurs son appellation aux très nombreux fossiles d'ours des cavernes qu'elle a livré. L'ultime fouille effectuée en 1996 a été réalisée dans un diverticule terminal de la cavité rempli d'argile et de sables intercalés. Ce sauvetage, réalisé dans des conditions dangereuses à cause des menaces d'effondrement dues à l'avancement de la carrière, a permis de mettre à jour de nombreux restes d'ours et d'hyènes, dont des coprolithes. Quelques fossiles de loup et de cheval ont également été découverts. Les ossements de très jeunes hyènes, les os de proies mâchouillés ou régurgités ainsi que les coprolithes retrouvés dans la grotte permettent d'affirmer que celle-ci a servi de repaire à l'hyène des cavernes. L'étude des pollens contenus dans les coprolithes a montré que les ours et les hyènes de Châteaubourg évoluaient dans un climat rigoureux. La grotte n'est pas ornée ; les intrusions humaines seraient donc liées à la présence des ours en hibernation ou morts durant celle-ci, afin de récupérer leur peau ou leur viande.

### **3.1.5. Les Grottes du site de Soyons**

Situé sur la face est du massif du Guercy, le réseau complexe de grottes formant le site de Soyons a été découvert par le Vicomte Lepic et J. de Lubac en 1870. Il y a actuellement sept grottes connues : la Grotte de Néron, la Grotte des Enfants, l'Abri Moula, la Grotte de la Madeleine, le Trou du Renard, le Trou du Mouton et le Trou Roland. Les deux grottes d'où proviennent nos échantillons sont la Grotte de Néron et le Trou du Renard, qui sont également les seules à être ouvertes au public.

#### ***3.1.5.1. La Grotte de Néron***

S'ouvrant en rive droite du Rhône, la grotte de Néron est constituée d'une vaste salle de trente mètres de long sur dix de large. C'est la plus grande des cavernes du site archéologique de Soyons. Les fouilles menées par le Vicomte Lepic et J. de Lubac dans les années qui suivirent la découverte ont permis de mettre à jour un gisement moustérien très important. En effet, il compte parmi les plus riches de France. D'autres fouilles ont été conduites entre 1915 et 1919 par le Pr Goury mais les résultats ne seront malheureusement pas publiés. Longtemps après, en 1950, les fouilles menées par M. Veyrier, E. Beaux et J. Combièr révèlent trois phases moustériennes dont la plus ancienne est encadrée par deux charniers (Veyrier *et al.*, 1951). Ces derniers se sont formés lorsque les néanderthaliens ont abandonné la grotte et que celle-ci a servi de refuge à des carnassiers. Les deux charniers ne présentent pas la même composition, le plus récent est riche en coprolithes et ossements de carnivores (ours, loups et hyènes des cavernes) alors que le deuxième, plus ancien, révèle de nombreux restes de grand bovidé et de cheval associés à l'ours des cavernes. La couche moustérienne qu'ils encadrent est la moins riche des trois mises à jour. L'ensemble de l'industrie contenue dans celles-ci est typique du moustérien final, avec la présence de silex et une forte utilisation des matières d'appoint telles que le calcaire schisteux, le quartz ou des galets issus du Rhône. Des pointes polymorphes typiques du moustérien, des racloirs, des proto-grattoirs, des burins et quelques os appointés en véritables prototypes de sagaies ont également été retrouvés (Veyrier *et al.*, 1951). Puis, en 1991, lors de la vérification de l'état de conservation du gisement, deux dents ont été découvertes : une canine déciduale inférieure et une première molaire supérieure définitive qui appartenait à un individu relativement jeune (Defleur *et al.*, 1992). Les études ont montré qu'elles appartenaient toutes deux à des néanderthaliens. Ces dents constituent les premiers restes néanderthaliens découverts dans

cette région. La Grotte de Néron a été classée patrimoine historique en 1965 puis aménagée et ouverte au public en 1995.

#### **3.1.5.2. Le Trou du Renard**

Découverte en même temps que la Grotte de Néron, la grotte dite du Trou du Renard est une cavité présentant d'abondantes concrétions. Elle fait partie d'un réseau karstique qui la relie notamment à l'Abri Moula situé un peu plus en contrebas sur la colline du site de Soyons (Defleur, 1995). Si la grotte est connue depuis 1870, ce n'est qu'en 1989 que les parties profondes, appelées « Réseau Ursus », seront découvertes et topographiées par G. Crouzet, G. Dal Pra et F. Raynaud. Des ossements de grands carnivores (ours, lion et hyène des cavernes) ainsi que des coprolithes d'hyène des cavernes ont été mis à jour dans ce secteur. Cette partie de la grotte est également riche en restes d'Homme de Néandertal et en dépôts minéraux précipités appelés spéléothèmes (Gardien *et al.*, 2009). La grotte a été aménagée et ouverte au public en 1989.

#### **3.1.6. Le Réseau Salomé**

La grotte de Salomé est un réseau de galeries qui circule dans les falaises des gorges de l'Ibie. La fouille d'une lentille fossilifère au fond de la cavité a livré de nombreux ossements d'ours des cavernes ainsi que des coprolithes d'hyène. Aucune protection n'étant posée pour contrôler la fréquentation du réseau, il ne nous est pas possible de donner sa localisation précise.

### **3.2. Les grottes situées dans le Lot**

#### **3.2.1. La Grotte-aux-Ours**

La Grotte-aux-Ours est située sur le territoire de la commune de Pressignac dans le Lot. Elle a été découverte en 2008 par des spéléologues et se développe sur près de quatre cent mètres. Comme pour le réseau Salomé, la grotte, mal protégée, est susceptible d'attirer l'attention des rodeurs. Nous ne pouvons donc indiquer sa localisation exacte.

Comme son nom l'indique, la Grotte-aux-Ours a servi de refuge aux ours des cavernes lors de leurs longues périodes d'hivernation. En témoignent les griffades sur les parois, les ossements

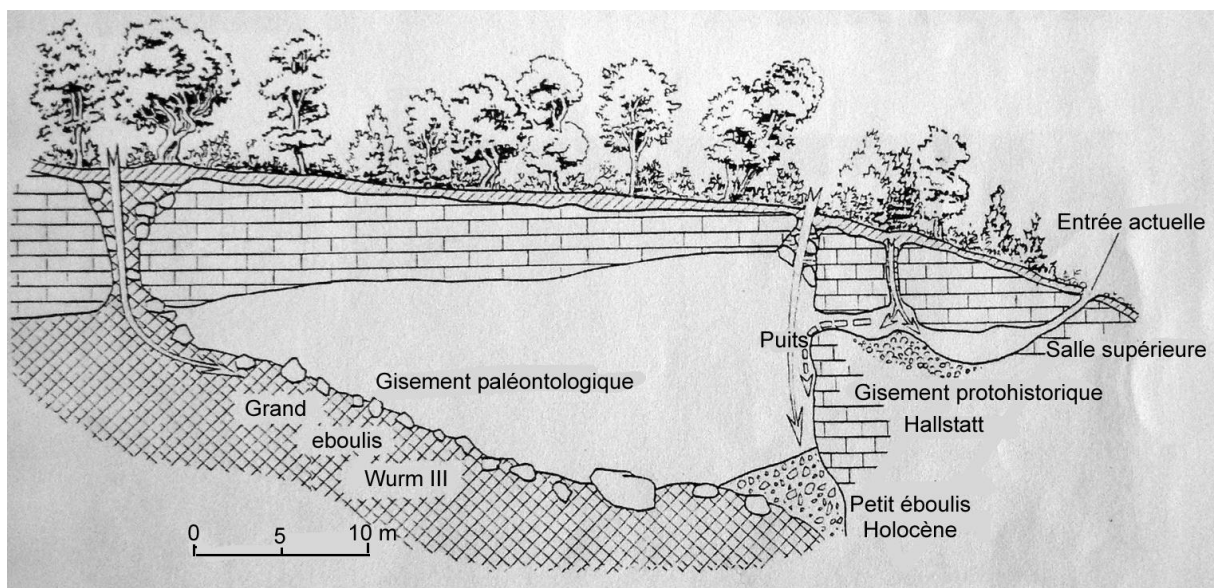
et les dents retrouvés ainsi que les bauges. De nombreuses et très nettes empreintes dans l'argile sont également visibles.

Les ours des cavernes ne sont pas les seuls à avoir fréquenté la Grotte-aux-Ours. En effet, la présence de l'hyène des cavernes est attestée par un crâne et par la présence de plusieurs coprolithes.

### 3.2.2. La Grotte de Siréjol

#### 3.2.2.1 La découverte de la Grotte de Siréjol

Située sur le Causse Martel, un plateau calcaire faisant partie des Causses du Quercy, la Grotte de Siréjol a été découverte en avril 1964 par A. Bélonie, M. Bélonie et E. Brousse. Une première description de la grotte et des gisements qu'elle renferme est faite en 1966.



**Figure 31.** Topographie de la Grotte de Siréjol réalisée par Michel Philippe. D'après (Philippe et Guérin, 1973).

La cavité se découpe en deux étages correspondant à deux gisements bien distincts (Figure 31). L'étage dit supérieur de la grotte est composé d'une petite salle renfermant un gisement protohistorique, qui a servi de grotte sépulcrale aux Hommes de cette période ; trois crânes ont pu être recueillis par les inventeurs lors de la découverte. L'étage dit inférieur, quant à lui, est constitué d'une large cavité contenant deux éboulis. Le premier, de grande taille, est constitué par des sédiments attribués au vu de la faune au Würm III. Ces sédiments ont pénétré dans la grotte par une large diaclase de deux à six mètres de large jusqu'à

obstruction complète de celle-ci ; le grand éboulis renferme de très nombreux fossiles de la faune du Pléistocène. Le second éboulis, de taille beaucoup plus modeste, est formé par des sédiments attribués à l'Holocène ; ceux-ci sont entrés par un puits de dix mètres de haut qui constitue la séparation entre les deux parties de la grotte. Lors du premier chantier de fouilles en 1972, l'entrée de la grotte est agrandie afin d'en faciliter l'accès puis fermée par une trappe pour éviter le renouvellement d'actes de vandalismes dont la grotte avait été victime.

### 3.2.2.2. *La faune très spéciale de la Grotte de Siréjol*

Les paléontologues se sont intéressés principalement au grand éboulis attribué au Würm III qui s'est avéré très fossilifère. Son âge a été confirmé par la datation au moyen de la méthode du radiocarbone des restes fauniques. Ils ont livrés des âges de 31 300 BP  $\pm$  1 700, 29 100 BP  $\pm$  1 450 et de 30 080 BP  $\pm$  1 600. Les datations obtenues pour le petit éboulis donnent un âge de 4 000 ans environ. Les deux éboulis ne sont donc pas contemporains.



**Figure 32.** Reste osseux retrouvés dans le grand éboulis de la Grotte de Siréjol. A gauche : Squelette presque complet d'un petit renne (*Rangifer tarandus*) . A droite : Crâne de bison des forêts (*Bison schoetensacki*). D'après (Philippe et Guérin, 1973).

Les fossiles de grands mammifères retrouvés dans les gisements de la Grotte de Siréjol appartiennent principalement à trois espèces. Il s'agit du cheval (*Equus caballus gallicus*), du renne (*Rangifer tarandus*) et d'une espèce rare de bison, le bison des forêts (*Bison schoetensacki*). La présence de ces trois espèces dans le même gisement peut surprendre. En effet, *Bison schoetensacki* comme son nom l'indique évoluait dans un biotope de type forestier. En revanche, *Equus caballus gallicus* et *Rangifer tarandus* évoluaient dans des espaces ouverts, steppe et prairie respectivement. L'environnement autour de la Grotte de Siréjol devait plutôt correspondre à un paysage de type steppe forestière à l'époque où ces animaux l'ont fréquentée.

### 3.3. La Grotte des Trois-Frères en Ariège

#### 3.3.1. La découverte de la Grotte des Trois-Frères

La Grotte des Trois-Frères située en Ariège est une grotte du Paléolithique supérieur qui doit son nom à ses inventeurs les trois fils du Comte H. Bégouen. Ils l'ont découverte en juillet 1914 alors qu'ils exploraient un passage étroit présent dans la grotte voisine du Tuc d'Audoubert. La Grotte des Trois-frères est reliée à la Grotte d'Enlène par une galerie de 60 mètres de long (Figure 33).



**Figure 33.** Topologie de la Grotte des Trois-Frères.

Ces trois cavités appartiennent au réseau karstique appelé les cavernes du Volp, du nom de la rivière qui les a creusées. Dans un souci de préservation, les grottes n'ont pas été ouvertes au public, les scientifiques peuvent néanmoins y pénétrer accompagnés d'un membre de la famille Bégouen.

#### 3.3.2. Les représentations pariétales de la Grotte des Trois-Frères

La Grotte des Trois-Frères comporte de nombreuses figurations pariétales attribuées au Magdalénien contrairement à la Grotte d'Enlève où elles sont quasiment inexistantes. La plupart des relevés des gravures et des peintures ont été effectués par l'Abbé Breuil.

La caverne ne compte pas moins de 350 figures animales répertoriées dont les plus présentes sont le bison et le cheval avec respectivement 170 et 84 représentations. Fait rare dans l'art pariétal, H. Bégouen fait mention d'un dessin de sauterelle très précis, qui permet d'affirmer la présence de cet animal dans les Pyrénées à cette période. La grotte est également célèbre pour ces deux chimères l'une, mi-homme mi-bison, appelée « le petit sorcier à l'arc musical » qui semble jouer d'un instrument et la seconde nommée « le dieu cornu » qui pourrait représenter un sorcier pratiquant un rite magique ou un chaman en transe. La grotte présente également des signes claviformes rouges dans la Galerie des Chouettes et des mains négatives dans la Galerie des Mains.







*Partie 2 :*

*Matériel et Méthodes*

# 1. PRÉLEVEMENTS DES ÉCHANTILLONS ÉTUDIÉS

Les échantillons étudiés, durant cette thèse, proviennent soit de collection de musée soit ils ont été prélevés directement dans leur environnement d'origine. Lorsqu'ils ont été collectés en grotte par l'équipe ou des collaborateurs, tous les permis nécessaires avaient été préalablement obtenus et toutes les précautions ont été prises afin de ne pas contaminer les échantillons. Au total, 39 échantillons - coprolithes ou ossements - provenant de 11 sites français répartis dans 3 régions, ont été prélevés en vue d'analyse ADN.

**Tableau 1.** Table des échantillons collectés pour chaque site, leur nature et leur origine.

Site/Département	Nature	Origine	Etat	Échantillons analysés
Chauvet-Pont d'Arc (07)	coprolithe	Prélèvement sur site	Natif	CCH1 ; CCH2 ; CCH3
Salomé (07)	coprolithe	Prélèvement sur site		Sal23 ; Sal533 ; Sal85 ; Sal192 ; Sal522
Trou du Renard (07)	coprolithe	Collection de musée	Vernis	TR1
Néron (07)	coprolithe	Collection de musée	Vernis	Ner1 ; Ner2 ; Ner3 ; Ner4
Deux-Ouvertures (07)	coprolithe	Prélèvement sur site	Natif	CDO451 ; CDO659 ; CDO676
Grotte-aux-Ours (07)	coprolithe	Prélèvement sur site		Chât1 ; Chât2 ; Chât3 ; Chât4
Soyons(07)	coprolithe	Collection de musée	Vernis	AX1
Grotte-aux-Ours (46)	coprolithe	Prélèvement sur site	Natif	Gao1 ; Gao2 ; Gao3 ; Gao4 ; Gao5
Grotte-aux-Points (07)	coprolithe	Prélèvement sur site	Natif	Cop01 ; Cop02
Siréjol (46)	ossement	Prélèvement sur site		SirB1 ; SirB2 ; SirB3 ; SirB4 ; SirB5
Trois-Frères (09)	ossement	Prélèvement sur site	Natif	SGE1 ; SGE2 ; SGE3 ; SGE5

## **2. DATATION RADIOCARBONE DES ÉCHANTILLONS**

Dans le but d'estimer l'âge des différents échantillons, des fragments de certains d'entre eux ont envoyées pour datation à Thomas Stafford de l'Université d'Aarhus (Danemark).

Les fragments d'ossements sont réduits en petits morceaux et sont baignés dans de l'eau déionisée (ED) avec une brève sonication pour retirer les sédiments qui sont collés sur ces derniers. Les fragments sont ensuite décalcifiés dans une solution d'HCl 0,2 N pendant 6 heures à 4°C et les protéines (le collagène pour les os), non dissous par l'acide, est rincé avec de l'ED. Les protéines sont traitées avec une solution composée de KOH 0,1 M pendant 2 heures à 4°C, puis il est rincé avec de l'ED. Les protéines décalcifiées sont réduites en gélatine dans une solution d'HCl 0,05 N à 90°C pendant une heure et le produit obtenu est filtré à travers une membrane Durapore de 0,45 µm. Les protéines solubilisées sont dialysées sur des filtres à 30-kDa, la fraction dont la taille est supérieure à 30-kDa est retenue puis lyophilisée. Les protéines purifiées sont ensuite chauffées à 850°C dans des tubes de quartz contenant de l'oxyde de cuivre II (CuO) et de l'argent. Le CO<sub>2</sub> est isolé par cryogénie, puis converti en graphite à l'aide de la méthode H<sub>2</sub>-Fe et enfin il est daté au W. M. Keck Carbon Cycle Accelerator Mass Spectrometry laboratory de l'Université de Californie-Irvine (UCIAMS). La date calibrée (calBP) et son intervalle de confiance sont obtenus en utilisant le logiciel OxCal 4.2 (Ramsey 2014) et la courbe de calibration IntCal13.

Le protocole de datation est le même pour les coprolithes, à la différence que ce sont les protéines totales qui sont datés à la place du collagène seul.

## **3. PRÉVENTIONS DES CONTAMINATIONS**

Pour éviter les contaminations, les phases initiales (extraction de l'ADN des échantillons archéologiques, construction des banques métagénomiques et préparations des réactions de PCR) sont réalisées dans un laboratoire dédié où aucune molécule d'ADN actuel et aucun ADN ancien amplifié n'a été manipulé. Ce laboratoire est situé dans un autre bâtiment que celui où les ADN amplifiés sont manipulés. Lors de ces phases initiales, des contrôles négatifs incluant des blancs d'extraction (« mock extract ») et des blancs de PCR sont inclus. Les personnes travaillent dans cette pièce ne sont pas entrées dans le laboratoire d'ADN moderne le même jour et portent des vêtements propres. Des chaussures dédiées sont

enfilées dans un sas, ainsi qu'une charlotte, un masque, une combinaison jetable et des gants qui sont fréquemment changés.

De plus, le matériel non jetable est nettoyé soigneusement après chaque utilisation. Les godets du Mikro-dismembrator ainsi que les joints et les billes sont immergés dans de l'eau MilliQ dans un bac à ultrason et soniquer dans deux bains successifs de 5 minutes. Puis, ils sont plongés dans une solution d'H<sub>2</sub>O DNA grade (Life technologies) et de SDS 0,1 % pendant au minimum 30 minutes. Ils sont ensuite rincés 3 fois à l'H<sub>2</sub>O DNA grade et séchés dans 10 à 15 ml d'éthanol absolu. Puis, ils sont laissés à sécher sur un kleenex propre et rangés dans un falcon propre jusqu'à leur prochaine utilisation.

#### **4. EXTRACTION D'ADN**

Pour les ossements, le cortex ou la moelle sont prélevés, selon que l'échantillon est dans un état natif ou vernis. Pour les coprolithes, lorsque l'échantillon est intact (non fragmenté), l'intérieur est gratté avec un scalpel après que l'échantillon ait été ouvert en deux, afin d'obtenir un maximum d'ADN endogène à l'échantillon. Pour les os comme pour les coprolithes, environ 0,5 g de matériel est récupéré. Puis les échantillons sont fragmentés au moyen d'un Mikro-dismembrator (Sartorius) en agitant 1 minute à 3 000 battements par minute. La poudre ainsi obtenue est progressivement solubilisée dans 5 ml tampon d'extraction (EDTA 0.5 M, Tris 10mM (pH 8.0), SDS 0,1 %, DTT 60mM et Protéinase K 0,5 mg/ml) en ajoutant des fractions de 0,5 ml par « up and down » à l'aide d'un cône coupé. Il n'est pas apparu nécessaire de rajouter du NPTB pour rompre les liaisons dues à la réaction de Maillard lors des extractions d'ADN de nos échantillons. La solution obtenue est incubée à 42°C pendant la nuit dans un falcon 15 ml sous agitation rotative. L'échantillon est ensuite centrifugé à 3 000 g pendant 5 min à température ambiante. Le surnageant est transféré dans un falcon 15 ml et 3 extractions successives sont réalisées (la première avec un volume (5ml) de phénol, la deuxième avec 0,5 volume de phénol et 0,5 volume de chloroforme et la dernière avec un volume de chloroforme). À chaque fois, c'est la phase aqueuse qui est récupérée. Elle correspond à la phase supérieure sauf pour la première extraction où la phase aqueuse se situe en dessous de la phase organique. L'extrait est dialysé et concentré par centrifugation sur une colonne Amicon. La membrane est ensuite rincée 4 fois avec 2 ml d'eau ultrapore sans nucléase (Life Technologies) et l'extrait d'ADN récupéré sous forme d'un volume final de 100 µl. Ce volume est ensuite purifié sur une colonne Qiagen 28004.

Cet extrait est utilisé par la suite pour la construction de banque ou la réalisation d'amplifications d'ADN ciblées.

## **5. RÉALISATION D'AMPLIFICATION D'ADN CIBLÉES**

Les PCRs sont réalisées dans un volume final de 50 µl contenant 300 nM d'amorce sens et antisens, 200 µl de dNTP, 2,5 mM de MgCl<sub>2</sub>, 5 µl de GeneAmp 10x PCR Buffer II et 2,5 µl d'AmpliTaq Gold DNA polymérase (Thermo Fisher Scientific, Waltham, MA, USA). Les amorces ont été sélectionnées grâce au logiciel Oligo 6.8 (Medprobe, Norvège). Des témoins négatifs, dans lesquels l'extrait d'ADN est remplacé par un mock ou de l'eau, sont ajoutés à chaque expérience de PCR. Le protocole de PCR consiste en une étape d'activation (95°C, 8'45''), suivie de 45 cycles (95°C, 15'' ; 50-58°C (en fonction du couple d'amorces), 20'' ; 70°C, 1') réalisés dans des thermocycleurs 7000 ou 7300 (Applied Biosystems), ou dans un Veriti thermal cycler (Thermo Fisher Scientific). Chaque fragment est analysé au moyen d'un minimum de deux réactions de PCR.

L'ADN amplifié est purifié par électrophorèse sur un gel vertical d'acrylamide à 8%. La totalité de la réaction de PCR est chargée sur le gel dans un tampon de dépôt contenant du bleu de bromophénol. Lorsque le front de migration est suffisamment avancé (bleu de bromophénol à 8-10 cm des puits de dépôts), le gel est récupéré et mis à incuber 20 minutes sous agitation dans une solution de TAE 1x, contenant du SyBr Green dilué (1/10000). Le gel est ensuite placé sur un transilluminateur UV (312 nm) pour visualiser l'ADN, et les fragments amplifiés de taille attendue sont découpés. Pour éluer l'ADN du gel, la bande d'acrylamide est transférée dans un tube de 0,5 ml. Ce tube, dont le fond a préalablement été percé avec une aiguille 21 g, est lui-même placé dans un tube eppendorf de 2 ml. L'acrylamide est pulvérisée dans le tube de 2 ml par centrifugation (5 minutes à 12 000 g), puis incubé une nuit à 37°C sous agitation dans 250 µl de tampon d'élution (Tris 2 mM (pH 8), EDTA 0,2 mM, acétate d'ammonium 1 M) contenant 20 µg de glycogène. L'ADN est ensuite purifié par filtration sur une cartouche de 0,22 µm de porosité (Costar Spin-X), concentré par précipitation à l'éthanol et le culot re-suspendu dans du TE 0,2 X de façon à obtenir une concentration d'ADN de 0,5 à 1 ng/µl.

Les produits de PCR sont ensuite clonés dans un vecteur pCR®4-TOPO®, puis le vecteur est inséré dans une souche d'*Escherichia coli* chimio-compétente. Etalés sur des boîtes de pétri contenant de l'ampicilline. Au minimum 8 colonies de chaque produit de PCR clonés sont repiquées dans des boîtes de culture contenant 2 ml de LB broth Lennox et de l'ampicilline (50 µg/ml). Les bactéries sont mises à pousser toute une nuit à 37°C sous agitation ; puis elles sont récupérées sous forme d'un culot après centrifugation du milieu de culture. Les parois des bactéries sont lysées au moyen d'une solution de TE 25 mM et d'une solution de NaOH 20 mM et SDS 1%. L'ADN génomique des bactéries est ensuite précipité car à une solution d'Acétate de Sodium 3 M (pH 5,2), l'ADN plasmidique d'intérêt resté en suspension est récupéré par retournement, ainsi que les ARNs. Une solution de TE 10 mM additionnée d'une ARNase (10 mg/ml) permet de se débarrasser de ces derniers. Les ADN plasmidiques sont récupérés dans 100 µl de TE 1,25 mM. Ils sont ensuite amplifiés par PCR dans un volume final de 20 µl (10 µl d'H<sub>2</sub>O, 3,5 µl de diluant 5X, 1 µl de BigDye Terminator 3.1 (Applied Biosystems) et 0,5 µl d'oligo Pu forward ou T3 à 5 pmol/µl). Le séquençage est réalisé en utilisant l'amorce universelle M13 des différents clones. Le protocole de PCR consiste en 30 cycles (95°C, 20'' ; 50°C, 30'' ; 60°C, 4'). Les produits de PCR sont ensuite précipités par centrifugation avec une solution froide d'éthanol et de NH<sub>4</sub>Ac ; puis ils sont lavés avec une solution glacée à 70 % d'éthanol. Les culots sont repris dans 10 µl de Hi-Di Formamide et séquencés par la méthode de Sanger sur un séquenceur 3130 XL.

Les séquences obtenues sont analysées grâce au logiciel Sequencher 5.3. Tout ce qui ne se trouve pas entre les deux promoteurs du vecteur (AATTCGCCCTT et AAGGGCGAATT) est éliminé, les promoteurs y compris. Les 8 fragments sont ensuite alignés et comparés afin d'établir un consensus pour chaque produit de PCR. Ce consensus est ensuite comparé à la banque de données en ligne GenBank afin de déterminer de quelle espèce il s'agit. Si les séquences sont de mauvaises qualités ou s'il y en a trop peu pour établir un consensus fiable, d'autres colonies sont mises en culture afin de compléter les informations avec d'autres séquences. De plus, si les consensus des deux PCRs ne sont pas identiques, une réaction supplémentaire de PCR est réalisée.

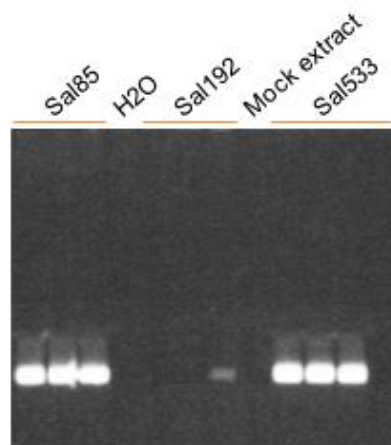
## 6. SÉLECTION DES ÉCHANTILLONS

La sélection des échantillons se fait en fonction de la préservation de l'ADN qui y est contenu. Pour les coprolithes comme pour les ossements, c'est la quantité d'ADN de l'espèce à laquelle correspond l'échantillon qui est informative. Ainsi, les coprolithes ont été testés avec des amorces spécifiques soit du loup, soit de l'hyène des cavernes et les ossements avec des amorces spécifiques du bison (Tableau 2).

**Tableau 2.** Liste des amorces utilisées pour amplifier l'ADN contenus dans les différents échantillons.

Espèce ciblée	Amorce sens	Amorce anti-sens	Taille
Hyène des cavernes	CAACCGCCTTCTCATCAGTGAC	AAGCTCCGTTGGCGTGTATGTAT	84
Loup	AACTATACCTGGCATCTGGTTCT	CCATCGAGATGTCCCATTT	89
Bison d'Europe	GATCAATCCCCAACAAAC	GTGTGTAGTAGGGGAATTAGA	104
Ours des cavernes	AACAACATATTCCTTCCATGAGC	AAGTAAGTGAGCATTTTGACTGGTAC	112

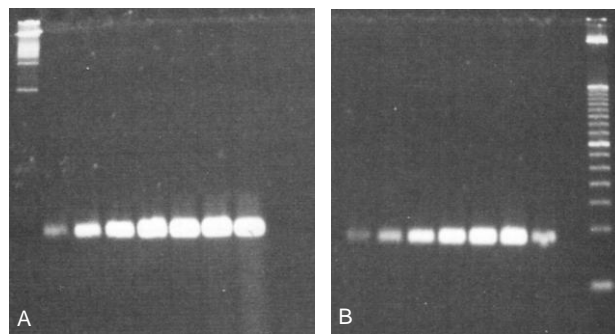
Afin de sélectionner le meilleur, chacun de 39 échantillons est testé à plusieurs reprises avec les amorces spécifiques à l'espèce d'intérêt. Le premier test consiste à effectuer un grand nombre de cycles de PCR (45 cycles) sur trois 'aliquots', correspondant à 0,1-0,3-0,9 µl de l'extrait (Figure 34).



**Figure 34.** Tentative d'amplification par 45 cycles de PCR d'un fragment - 84 pb - du gène mitochondrial du cytochrome B pour 3 coprolithes d'hyènes des cavernes collectés dans le Réseau Salomé avec 0,1-0,3-0,9 µl d'extrait.



Les amorces utilisées permettent une amplification qui cible un fragment du gène du cytochrome B du génome mitochondrial (84 nucléotides pour l'hyène des cavernes, 104 nucléotides pour les bovidés et 89 nucléotides pour le loup). Les amplicons sont clonés et séquencés. Les échantillons donnant des séquences d'ADN correspondant à l'espèce attendue et une amplification pour toutes les dilutions testées sont soumis à un second criblage. La figure 34 montre que les échantillons Sal85e2Q et Sal533e2Q passent le premier test contrairement à l'échantillon Sal192Q (Figure 35). Le second criblage consiste à analyser l'extrait des échantillons sélectionnés sur une large gamme de volume (de 0,01 à 6 µl) en effectuant moins de cycles de PCR que précédemment (33 cycles au lieu de 45).



**Figure 35.** Amplification de 33 cycles sur une gamme de volume (de 0,01 à 0,06 µl) pour les deux échantillons ayant passé le premier crible (A : Sal533e2Q et B : Sal85e2Q).

Cette gamme permet de se rendre compte de la plus petite quantité d'extrait à partir de laquelle une amplification est visible ; mais également à partir de quelle quantité d'ADN, les molécules inhibitrices commencent à avoir un impact sur la réaction de PCR. L'échantillon sélectionné est celui qui présente idéalement une amplification pour toutes les quantités de la gamme et peu d'inhibiteurs ; dans la figure XX, c'est l'échantillon Sal533e2Q. La quantité maximale testée révélant une amplification sans indication d'inhibition est utilisée pour la construction d'une banque de fragments.

**Tableau 3.** Liste des échantillons les plus riches en ADN et présentant le moins de molécules inhibitrices sélectionnés pour chaque site.

Site	Échantillon sélectionné
Grotte Chauvet-Pont d'Arc	CCH3
Réseau Salomé	Sal533
Trou du Renard	TR1
Grotte de Néron	Ner2
Grotte des Deux-Ouvertures	CDO676
Grotte de Châteaubourg	Chât1
Site inconnu	AX1
Grotte-aux-Ours	Gao1
Grotte-aux-Points	Cop01
Grotte des Trois-Frères	SGE2

## 7. CONSTRUCTION DE BANQUES DE FRAGMENTS

Pour chaque site étudié, les banques de fragments ont été construites à partir des échantillons les plus riches en ADN. Quand plusieurs extraits ont été réalisés pour un échantillon, nous avons sélectionné le meilleur pour la construction de la banque de fragments (Tableau 4).

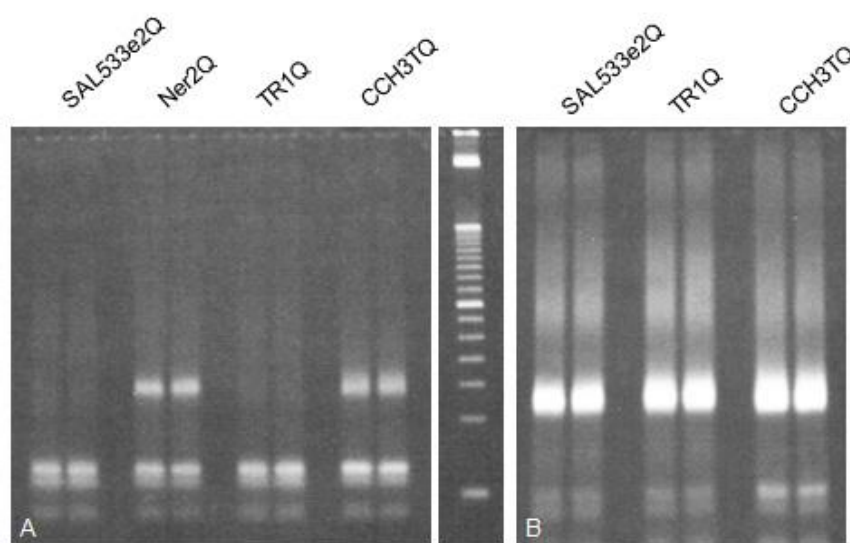
**Tableau 4.** Nom des extraits utilisés pour la construction des banques. La lettre Q correspond à la purification par Qiagen 20804 et e2 indique le numéro de l'extrait car il y en a parfois eu plusieurs de réalisés.

Site	Nom de l'extrait utilisé
Réseau Salomé	Sal533e2Q
Trou du Renard	TR1Q
Grotte de Néron	Ner2Q
Grotte Chauvet-Pont d'Arc	CCH3TQ CCH3Ce2Q
Grotte des Deux-Ouvertures	CDO676Q
Grotte de Châteaubourg	Chât1Q
Soyons	AX1Q
Grotte-aux-Ours	GAO1Q
Grotte-aux-Points	Cop01Q
Grotte des Trois-Frères	SGE2e2Q

Les banques pour le séquençage haut-débit au moyen de la méthode Illumina sont construites à partir d'une quantité variable d'extrait de 2.5  $\mu$ l à 10  $\mu$ l, selon la préservation du matériel génétique contenu dans les échantillons. Plusieurs kits (PE-102-1001 et FC-102-1001) de chez Illumina sont utilisés. Les recommandations du constructeur sont appliquées (San Diego, CA, USA).

Brièvement, les extrémités des fragments d'ADN sont réparées grâce à la T4 et la Klenow DNA polymérase et les extrémités 5' sont phosphorylées grâce à la T4 polynucléotide kinase. Une déoxyadénosine est ajoutée aux extrémités 3' des fragments en utilisant une Klenow *exo*<sup>-</sup> polymérase. Des adaptateurs Illumina sont ensuite ligaturés aux fragments d'ADN avec une DNA ligase. Le mix obtenu est purifié sur une colonne Qiagen 28104 et élué avec 30  $\mu$ l de tampon d'éluion. L'étape finale de la construction consiste en l'amplification des fragments d'ADN ligaturés aux adaptateurs. Un 'aliquot' de 20 $\mu$ l de fragments d'ADN ligaturés aux adaptateurs est ajouté à un mix de PCR contenant de la Phusion DNA polymerase, de l'H<sub>2</sub>O et les amorces universelles PE 1.0 et 2.0. Le protocole de PCR consiste en une étape d'activation (98°C, 30''), suivie de 12 cycles (98°C, 40'' ; 65°C, 30'' ; 72°C, 30'') et une étape finale (72°C, 5').

La quantité d'ADN ligaturé utilisé pour réaliser la réaction de PCR est très importante. En effet, à cause de la présence de molécules inhibitrices, il est nécessaire de bien doser la quantité d'ADN afin de ne pas inhiber totalement la réaction de PCR.



**Figure 36.** Banques de fragments construites avec 10  $\mu$ l (A) et 5  $\mu$ l (B) pour des extraits d'ADN de coprolithes d'hyènes des cavernes. CCH3 sert de témoin.

Lorsque les banques pour les extraits Sal533e2Q et TR1Q sont construites avec 20  $\mu$ l d'ADN ligaturé aux adaptateurs, il n'y a pas d'amplification de la banque ; pourtant les analyses précédentes avaient montré que les échantillons étaient très riches en ADN. Pour vérifier cela, une autre série de PCR où il n'y a que 5  $\mu$ l d'ADN ligaturé qui est rajouté au mix de PCR est réalisé ; cette fois-ci on observe une forte amplification d'ADN. Les inhibiteurs avaient empêché l'amplification lors de la première PCR.

## 8. RECONSTITUTION DE GÉNOMES MITOCHONDRIAUX

### 8.1. Assemblage des fragments en contigs

Les contigs sont construits avec le logiciel SOAP (Li *et al.*, 2009) avec comme critère un alignement parfait sur 23 nucléotides. Les contigs obtenus sont ensuite alignés sur la base de données en ligne GenBank afin de savoir à quelle espèce ils appartiennent.

### 8.2. Alignement des lectures de la banque

Afin de reconstruire un génome mitochondrial complet fiable, une stratégie en plusieurs temps est mise en place. Dans un premier temps, les lectures de la banque sont alignées sur les contigs ou sur un génome de référence grâce au logiciel BWA (Tableau 5).

**Tableau 5.** Génomes utilisés pour aligner les lectures des différentes banques.

Espèce	N° GenBank
Loup	DQ480505
Hyène des cavernes	JF894378
Bison d'Europe	HM045017
Bison d'Amérique	EU177871

Le crible choisi tolère 0,1 mésappariement (MM) ce qui correspond à 2 MM pour les lectures entre 20 et 27 nucléotides, 3 MM pour celles entre 28 et 56 nucléotides, 4 MM pour celles entre 57 et 88 nucléotides et 5 MM pour celles entre 89 et 101 nucléotides. Cette première analyse permet d'établir un consensus temporaire dont les trous sont comblés avec le génome d'une espèce proche. Puis un nouveau crible est effectué en alignant les lectures de

la banque sur le consensus temporaire avec le même critère d'alignement tolérant 0,1 MM, afin d'obtenir un consensus de deuxième génération. L'opération est répétée autant de fois que nécessaire afin d'obtenir une séquence mitochondriale complète et fiable.

### **8.3. Réalisation de PCR ciblées**

Si la couverture est trop faible ou si le génome présente trop de différences par rapport à la référence, il est possible que des trous subsistent dans la séquence du génome mitochondriale. Pour combler ce manque d'information et obtenir une séquence complète, des amorces sont sélectionnées en utilisant le consensus construit avec l'alignement des lectures afin de réaliser des PCR ciblées.

## **9. RECHERCHE DE L'ADN MITOCHONDRIAL D'AUTRES ESPÈCES ANIMALES DANS LES COPROLITHES**

Lorsque les échantillons étudiés sont des coprolithes, il est possible d'étudier le bol alimentaire de l'espèce productrice. Pour cela, deux approches ont été mises en place : l'analyse informatique des données de la banque et la réalisation de réactions de PCR ciblées.

Les analyses informatiques se déroulent en deux étapes. En premier lieu, un grand nombre d'espèces est criblé en parallèle grâce à un script qui aligne les lectures de la banque sur tous les génomes en même temps avec un crible tolérant jusqu'à un mésappariement. Les séquences sont comptabilisées pour l'espèce pour laquelle elles s'alignent avec le meilleur score. Cette première analyse permet de faire ressortir l'espèce majoritairement présente. Cependant, comme les lectures s'alignent sur un seul génome à la fois, il est possible de perdre des informations. Ainsi, une seconde approche plus ciblée vient compléter la première en alignant les lectures de la banque uniquement sur le ou les génomes qui sont ressortis à l'étape préliminaire. Dans les deux cas, l'alignement des lectures sur le ou les génomes est réalisé grâce au logiciel BWA. De plus, pour éviter tout biais par l'ADN de l'espèce productrice, les séquences qui ont le meilleur score d'alignement ou un score équivalent sur la séquence du producteur, sont éliminées. Sont également retirées les lectures dont le meilleur alignement dans BLAST n'est pas celui de l'espèce analysée.

Les analyses par réaction de PCR ciblées sont une méthode de confirmer les résultats observés par l'approche bio-informatique. A chaque fois, des amorces spécifiques de l'espèce ciblée sont dessinées et sélectionnées pour présenter des mésappariements sur les autres espèces et particulièrement l'espèce productrice du coprolithe, afin de ne pas amplifier son ADN majoritaire dans l'échantillon.

## 10. ANALYSES PHYLOGÉNÉTIQUES

Dans le but d'établir la position phylogénétique des spécimens auxquels appartiennent les échantillons ou parfois de ceux qui composent leur alimentation, des arbres phylogénétiques sont réalisés avec les génomes mitochondriaux - complets ou partiels - obtenus.

Nos séquences mitochondriales sont alignées avec les génomes mitochondriaux disponibles dans la banque en ligne GenBank à l'aide de l'outil ClustalW du logiciel Méga6 (Tamura *et al.*, 2013) avec les paramètres par défaut. Lorsque les analyses sont réalisées avec des génomes complets, la région de contrôle qui comporte un motif répété est retirée (de 16 256 à 16 831 pour l'hyène, de 16 039 à 1 549 pour les canidés).

Les arbres phylogénétiques ont été réalisés à chaque fois avec différentes méthodes afin de comparer les topologies. La suite MEGA6 permet de construire les arbres avec les méthodes du Neighbor-Joining (NJ), du Maximum Parsimony (MP), celle de Minimum Evolution (ME) et celle du Maximum Likelihood (ML). Les analyses bayésiennes sont réalisées au moyen de MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003).

La méthode du MrBayes a été utilisée avec le modèle général du temps réversible (GTR) + 4 $\Gamma$  + I, et la méthode de NJ a été réalisée avec le paramètre Tamura-3. L'analyse NJ a été réalisée en utilisant un modèle de taux de substitution gamma et une matrice de substitution Jones-Taylor-Tornton (JTT). Les analyses MP, ME et ML des séquences protéiques sont effectuées avec les paramètres par défaut de MEGA 6.

Pour évaluer la solidité de l'arbre, la méthode des répliques a été utilisée avec un minimum de 1000 répliques pour les méthodes NJ, ME et ML. L'analyse bayésienne a été réalisée sur un minimum de 1 000 000 générations. Les longueurs de branches utilisées sont celles des arbres construits avec MrBayes ou avec la méthode du Maximum Likelihood.



*Partie 3 :*

*Analyses de coprolithes de canidés de la Grotte  
Chauvet-Pont d'arc*



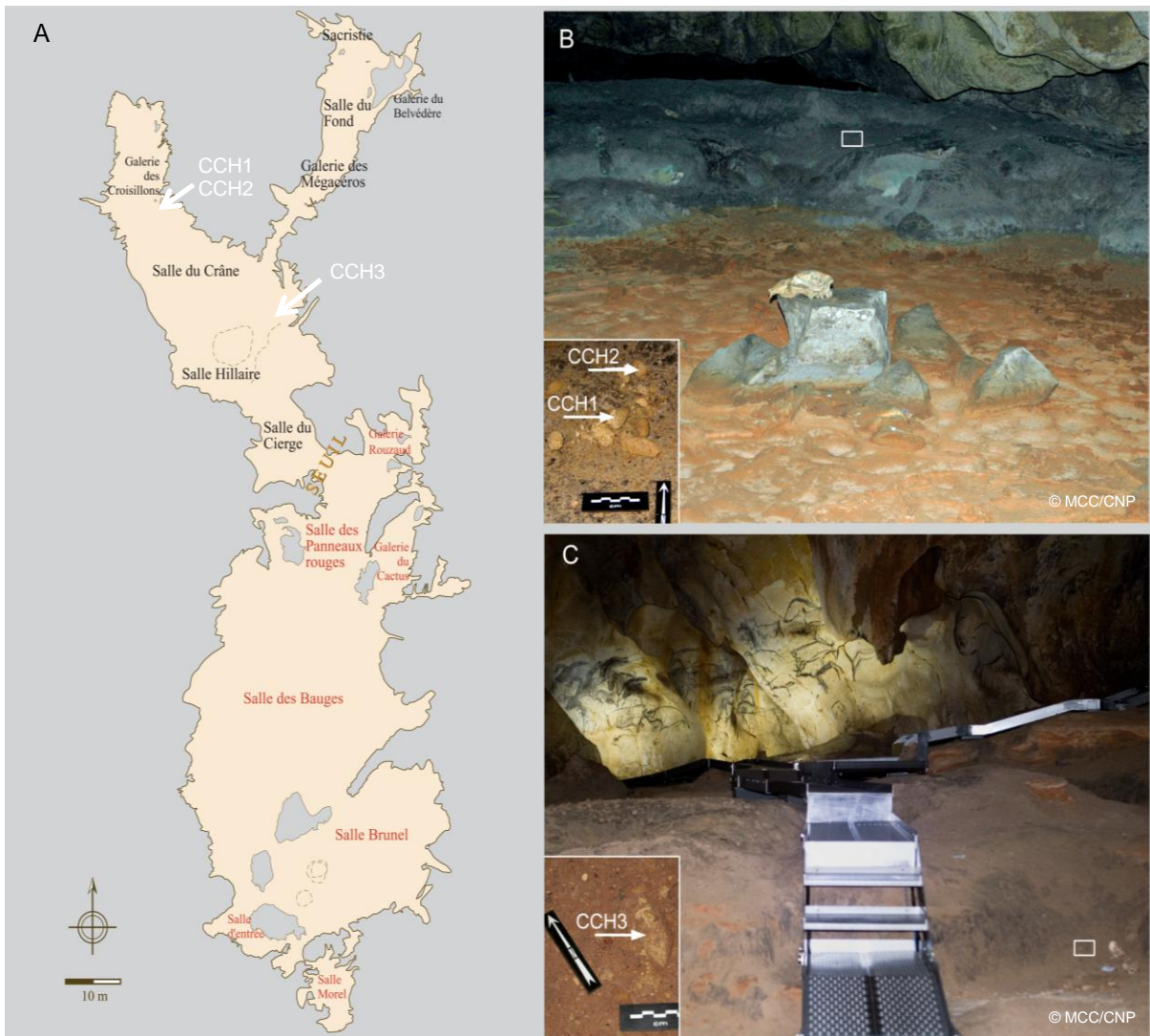
## 1. INTRODUCTION

La grotte Chauvet-Pont d'Arc, qui contient certaines des plus anciennes peintures connues, comporte également de nombreuses traces d'occupation animale : restes osseux, empreintes, poliss d'ours mais également coprolithes. Ce sont ces derniers échantillons, attribués à des canidés de par leur forme, qui ont été étudiés pour obtenir des informations sur les spécimens qui les ont produits.

Malgré de nombreuses études portant sur la domestication du chien, le lieu et le moment de celle-ci ne sont pas encore bien connus et restent fortement débattus. La domestication pourrait avoir eu lieu en Asie du Sud-Est il y a moins de 16 000 ans si on se base sur la diversité des génomes mitochondriaux et celle des haplogroupes du chromosome sexuel Y (Ding *et al.*, 2012). Elle pourrait également s'être déroulée au Moyen-Orient, il y a environ 14 000 ans si l'on juge d'après les fossiles et la mutation du gène nucléaire IGF1 (Gray *et al.*, 2010). La troisième hypothèse, basée sur l'analyse de génomes mitochondriaux anciens, propose que la domestication ait eu lieu en Europe occidentale entre 19 000 et 32 000 ans (Thalmann *et al.*, 2013). L'étude sur les anciens génomes mitochondriaux européens tendait à montrer que les chiens domestiques étaient issus de la diversité génétique des loups européens du Pléistocène (Thalmann *et al.*, 2013) ; cependant une partie de ces haplogroupes a aujourd'hui disparu.

Les canidés qui ont produit les coprolithes retrouvés dans la Grotte Chauvet-Pont d'Arc correspondent à une population ancienne, car, selon une étude récente, la fermeture de la caverne se serait produite il y a environ 21 000 ans (Sadleir *et al.*, 2012). L'analyse de restes aussi anciens est une grande opportunité pour essayer d'appréhender une partie de la diversité génétique des canidés du Pléistocène et comparer celle-ci à celle des loups et chiens actuels. Il n'est pas à exclure que ces canidés fassent partie des premiers chiens domestiques. Cette hypothèse permettrait d'expliquer la présence de canidés dans les secteurs profonds de la grotte.

Les coprolithes, prélevés par l'équipe, se situaient tous dans les secteurs profonds de la Grotte Chauvet-Pont d'Arc (Figure 37A). Les échantillons CCH1 et CCH2 faisaient partie d'un ensemble de coprolithes - brisés ou intacts - présents dans la Galerie des Croisillons (Figure 37B), tandis que l'échantillon CCH3 reposait seul et entier dans la Salle Hillaire (Figure 37C).



**Figure 37.** A : Topographie de la Grotte Chauvet-Pont d'Arc ; les flèches blanches indiquent les emplacements des coprolithes. B : La Salle du Crâne et l'entrée de la Galerie des Croisillons où les coprolithes CCH1 et CCH2 ont été trouvés. C : La Salle Hillaire où CCH3 a été trouvé à 10 mètres du Panneau des Chevaux.

L'ADN de ces trois coprolithes a été extrait et analysé par PCR et séquençage massif. Après sélection de l'échantillon le plus riche en ADN, nos recherches se sont concentrées en premier lieu sur la reconstruction du génome mitochondrial d'un canidé et la position phylogénétique de celui-ci. Puis, compte tenu de la position très intéressante au sein du genre *Canis* de notre spécimen, d'autres banques de fragments ont été construites, afin d'étoffer le corpus de données et d'explorer le génome nucléaire. En parallèle, l'alimentation de ce canidé a été explorée par deux approches différentes et un fragment de coprolithe a été envoyé en datation à un collaborateur.

## 2. MATÉRIEL ET MÉTHODES

Les analyses menées sur le génome nucléaire étant spécifiques au coprolithe provenant de la Grotte Chauvet-Pont d'Arc, des protocoles particuliers ont été mis en place.

### 2.1. Alignement des lectures sur le génome nucléaire du chien

Les lectures des 3 banques sont alignées sur le génome nucléaire du chien (CanFam3.1) par la plate-forme eBio de l'Université Paris XI. Avant cet alignement un contrôle de qualité a été effectué, il a permis de retirer les restes d'adaptateurs qui polluaient encore certaines lectures des banques 1 et 2. Après ce nettoyage, une base est retirée à chaque extrémité des lectures - l'ADN ancien y présentant un plus fort taux de dommages - puis, les lectures d'une taille inférieures à 25 nucléotides sont éliminées. Les lectures restantes sont divisées en deux sets de données, lectures longues (> 33 nucléotides) et lectures courtes (25 à 33 nucléotides, afin de les aligner en tolérant 2 mésappariements pour les lectures longues et 1 mésappariement pour les courtes. Le nombre d'indels autorisé est de 3 pour les lectures longues et de 1 pour les courtes. Le nombre de lectures alignées pour chaque set de données et chaque banque est compilé dans le tableau ci-dessous (Tableau 6).

**Tableau 6.** Nombre de lectures, longues et courtes, s'alignant sur le génome nucléaire du chien pour chaque banque. La taille des lectures longues est comprise entre 34 et 99 nucléotides, celle des lectures courtes entre 25 et 33 nucléotides.

Banque	Nombre de lectures longues	Nombre de lectures courtes	Total
1	14 805 086	5 131 160	19 936 246
2	13 261 273	5 846 668	19 107 941
3	13 687 253	4 521 617	18 208 870

Ensuite, les doublons (12 079 951) et lectures à alignements multiples (6 940 029) – jugées non pertinentes – sont retirés du set de données final ; au total 38 233 077 lectures sont alignées sur le génome nucléaire du chien.

### 2.2. Analyse des catégories de gènes

Les différences observées entre le génome partiel du canidé de Chauvet et le génome nucléaire du chien ont été répertoriées dans un fichier VCF (Virtual Card File). Elles ont été

classées d'une part par leur localisation sur le génome (régions codantes, non codantes, initiatrices de réplication) et d'autre part par le type de mutation qu'elles impliquaient lorsqu'elles se trouvaient dans des régions codantes (silencieuses, non sens et faux sens). Les mutations non sens font apparaître d'un codon stop dans la séquence codante tandis que les mutations faux sens entraînent, elles, un changement de l'acide aminé codé par le triplet. Le numéro ENSCAFG du gène était également indiqué quand la différence touchait une région codante. Le fichier VCF récapitulatif comprenait 510 268 lignes au total, chaque ligne correspondant à une position sur le génome.

Des cribles informatiques séparés sont réalisés dans le but d'extraire les numéros ENSCAFG des gènes portant des mutations spécifiques. Le premier sélectionne uniquement les mutations qui présentent un génotype « homozygote alternatif » et le deuxième celles qui ne sont pas des transitions. Les ENSCAFG extraits sont introduits dans l'outil en ligne Gostat (<http://gostat.wehi.edu.au/>) afin d'obtenir les catégories de gènes (GO) auxquels ils appartiennent, un gène pouvant avoir plusieurs GO. Les GO qui présentent un score de qualité supérieur à 0,05 ne sont pas retenus. Les listes des catégories de gènes obtenues pour les deux cribles sont comparées et seuls les gènes présents dans les deux cribles, c'est-à-dire ceux présentant des mutations autres que des transitions, sont conservés. Puis, les catégories de gènes obtenues sont analysées pour savoir lesquelles sont les plus fréquemment mutées.

### 2.3. Amplifications ciblées de gènes nucléaires

Dans le but d'obtenir des informations sur des mutations présentes dans des gènes nucléaires impliqués dans la digestion de l'amidon et la couleur du pelage, des amplifications ciblées sont réalisées grâce à des amorces spécifiques (Tableau 7).

**Tableau 7.** Liste des amorces utilisées pour réaliser les amplifications ciblant les gènes nucléaires avec la taille du fragment attendue (colonne 4) et la température d'hybridation amorces pour la PCR (colonne 5).

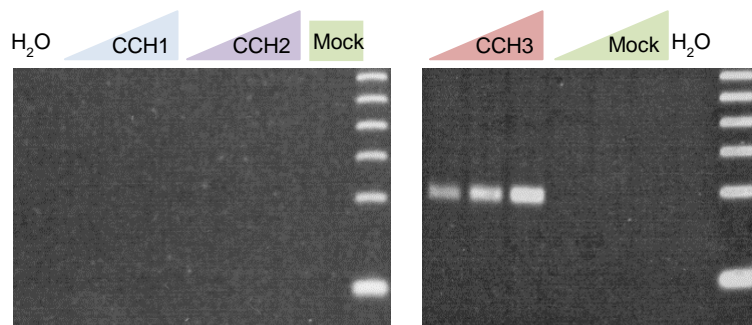
Gène	Amorce sens	Amorce anti-sens	Taille	T°C
MGAM	TCATTTCACTGACGTGGAGA	GTAGAAGTAGTTCCAAAGCTATTCAT	79 pb	49°C
SLC5A1	GAACGCCCTTCCAACGT G	ATGGGATCTCGGAAAATGTGA	99 pb	54°C
CBD103	GGCACGTCTGTTTTTTTCTTTG	GCACCGACCGCTCCTTATT	88 pb	53°C
ASIP	CAGCAGAAAAGAGGCGGAAAAGA	ACGCAGGGGGTGGGTGG	96 pb	60°C

Les réactions de PCR sont réalisées dans un volume de 50 µl, de la même manière que décrit précédemment, et le protocole de PCR consiste toujours en une étape d'activation (95°C, 8'45''), suivie de 45 cycles (95°C, 15'' ; 49-60°C, 20'' ; 70°C, 1').

### 3. RÉSULTATS ET DISCUSSION

#### 3.1. Choix de l'échantillon

Afin de caractériser les échantillons, l'ADN a été extrait et des amplifications par réactions de PCR - ciblant une portion (89 paires de bases) de la région de contrôle du génome mitochondrial du loup - ont été réalisées. Dans le but d'obtenir une caractérisation plus fine, une quantité d'ADN variable, correspondant à 0,5-2,5% de l'extrait a été incorporée au mélange de PCR (Figure 38).



**Figure 38.** Amplification par PCR d'une portion de la région de contrôle du génome mitochondrial (89 paires de bases). Les amplifications (33 cycles) ont été réalisées avec une quantité d'ADN correspondant à 0,5%, 1% ou 2,5 % de chaque extrait.

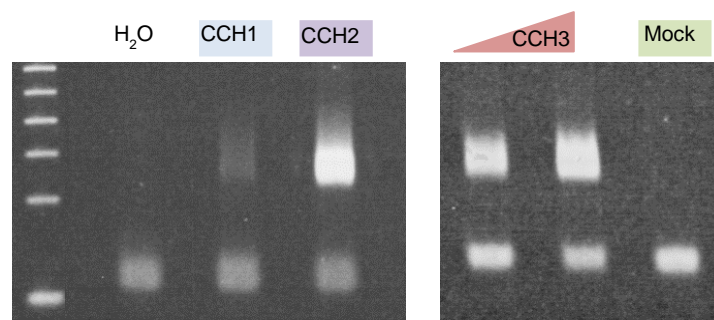
Les puits contenant de l'eau ou de l'extrait témoin (« mock extract ») sans ADN ne montrent aucune amplification d'ADN, prouvant ainsi qu'il n'y a pas eu de contamination lors de la manipulation des échantillons et lors de l'extraction d'ADN.

L'absence d'amplification pour les échantillons CCH1 et CCH2 indique qu'ils contiennent très peu, voire pas du tout d'ADN de loup. Comme les amorces utilisées permettent également d'amplifier l'ADN de coprolithes de renard et même d'hyène des cavernes (nos autres expériences réalisées sur des échantillons de différents sites), il est possible de penser que l'ADN de ces carnivores - s'il était présent - ne s'est pas non plus conservé dans les coprolithes.

L'échantillon CCH3 présente, par contre, une amplification d'ADN dès le début de la gamme et ne montre pas d'effet d'inhibition. Le séquençage des bandes observables pour CCH3 a permis de confirmer qu'il s'agissait bien d'ADN de *Canis lupus*.

Malgré les différences de résultats pour les 3 échantillons, une banque de fragments a été construite pour chacun d'eux, dans le but de savoir, après séquençage haut débit, si les coprolithes CCH1 et CCH2 n'appartenaient pas à une autre espèce.

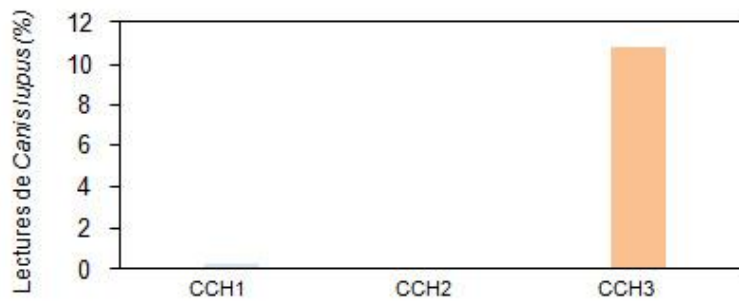
Avant ce séquençage massif, un 'aliquot' de chacune des 3 banques est amplifié par PCR pendant 12 cycles, afin de vérifier qu'elles ont été bien construites (Figure 39).



**Figure 39.** Amplification au moyen de 12 cycles de PCR d'une fraction de chaque banque. Pour CCH1 et CCH2, la banque a été construite à partir de 2,5% de l'extrait d'ADN; pour CCH 3, les banques ont été construites à partir de 2,5% ou 5% de l'extrait. Un aliquot de chaque banque (40%) est amplifié par PCR. Les bandes supérieures, spécifiques des échantillons, sont découpées pour séquencer l'ADN.

Grâce à cette PCR, il est visible que l'échantillon CCH1 contient très peu d'ADN. La quantité d'ADN amplifiée à partir des banques CCH2 et CCH3 est en revanche équivalente. L'absence d'amplification dans les puits contenant l'eau ou l'extrait contrôle sans ADN montre qu'il n'y a pas eu de contamination pendant la construction des banques. Les bandes, observables dans la partie inférieure du gel, correspondent aux adaptateurs. Une fraction de chaque banque amplifiée a ensuite été envoyée au Génoscope d'Evry, afin d'y être séquencée par haut-débit grâce à la méthode Illumina.

Le pourcentage d'ADN de canidé présent dans chaque échantillon est évalué en comparant 10 000 lectures sélectionnées de manière aléatoire aux séquences disponibles dans les banques de données *Genbank* (*nt/nr* et *wgs*). Les lectures qui présentent le même score pour plusieurs espèces sont jugées non discriminantes et ne sont pas prises en compte.



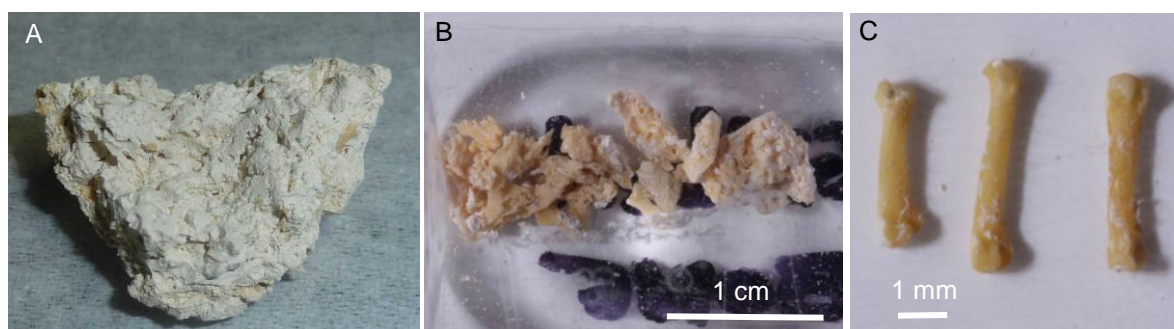
**Figure 40.** Pourcentage en ADN de canidé calculé pour chaque échantillon en comparant 10 000 lectures aléatoires aux séquences répertoriées dans *GenBank wgs*.

La figure 40 montre que les banques construites pour CCH1 et CCH2 ne présentent aucune lecture de *Canis lupus* pour les 10 000 séquences analysées et ne contiennent donc pas de quantité importante d'ADN de canidé, comme le laissaient pressentir les PCR préliminaires. Les banques de ces 2 échantillons ne montrent pas non plus de lectures correspondant à une autre espèce particulière de mammifère, y compris l'Homme. L'ADN de l'espèce qui a produit ces coprolithes n'apparaît donc pas conservé en quantité suffisante pour poursuivre l'analyse.

Par contre, la banque construite pour CCH3 contient environ 11 % d'ADN de *Canis lupus*. Nos études ultérieures se sont donc focalisées sur l'échantillon CCH3, puisque c'est celui qui contient de l'ADN de canidé.

### 3.2. Datation de l'échantillon

Dans le but de dater le coprolithe, sur lequel nos recherches vont se focaliser, un fragment de celui-ci a été détaché et envoyé à un collaborateur pour datation (Figure 41A). La dissection de ce fragment a permis de montrer que la matrice du coprolithe contenait des morceaux d'os broyés (Figure 41B), mais également des os intacts de petite taille (Figure 41C).



**Figure 41.** A. Fragment de coprolithe qui a été envoyé en datation (© J.-M. Elalouf). B et C: ossements broyés (B) et intacts (C) retrouvés après dissection du fragment de coprolithe (© T. Stafford).

Deux parties distinctes du coprolithe ont été datées : la matrice de l'échantillon et le collagène contenu dans les fragments osseux (Figure 41B) retrouvés après dissection du coprolithe. Ces deux fractions ont été datées par la méthode radiocarbone en AMS et les dates obtenues ont été calibrées grâce au logiciel OxCal en utilisant la courbe de calibration IntCal13 (Ramsey 2014). Les résultats obtenus pour la matrice du coprolithe et le collagène contenu dans les restes osseux ont été compilés dans le tableau ci-dessous en BP et en âge calibré (cal BP) (Tableau 8).

**Tableau 8.** Tableau compilant les informations sur la nature de la fraction datée, l'âge en BP et l'intervalle calibré pour 2 sigma (cal BP).

LAB NO.	Fraction datée	<sup>14</sup> C AGE, ±1 SD, BP	Intervalle CAL BP (2 sigma) 95.4% C.I.
AAR-19655	Matrice	27 900 ± 450	33 020 à 31 060
AAR-19656	Collagène des os	30 450 ± 550	35 610 à 33 620

La datation du collagène contenu dans les os donne un résultat à 30 450 ±550 BP, ce qui équivaut à un âge calendaire de 35 610 à 33 620 cal BP pour deux sigma. La datation de la matrice donne un âge plus récent : 27 900 ± 450 BP, avec un intervalle compris entre 33 020 et 31 060cal BP toujours pour deux sigma.

Les contaminations extérieures, en apportant du C14 plus récent, ont pour effet de rajeunir la date obtenue pour un échantillon. Le coprolithe CCH3 n'étant pas enfoui, sa matrice a été plus en contact avec l'air que les restes osseux présents à l'intérieur ; elle a donc pu être davantage sujette aux contaminations. C'est pourquoi, la date retenue pour notre échantillon est celle calculée grâce aux restes osseux ; ce qui est en accord avec un début de fermeture de la Grotte Chauvet-Pont d'Arc il y a 27 000 ans et sa fermeture définitive il y a 21 000 ans.



### 3.3. Les banques construites pour l'échantillon CCH3

Les deux premières banques construites à partir de l'échantillon CCH3 l'ont été au CEA de Saclay. Elles ont été réparties pour séquençage sur 19 pistes analysées au moyen de la plateforme Illumina GaIIx ; la profondeur des lectures était fixée à 76 nucléotides. Un nombre de 456 972 936 séquences - de taille comprise entre 20 et 76 nucléotides - a pu être obtenu par séquençage haut-débit.

Afin d'obtenir plus d'information sur l'échantillon CCH3, une troisième banque a été construite dans un laboratoire collaborateur à Copenhague (groupe de Thomas Gilbert, Centre for GeoGenetics, Natural History Museum of Denmark). Le séquençage, toujours au Génoscope d'Evry, sur 4 pistes avec une profondeur de 101 paires de bases a permis d'obtenir 504 145 349 lectures.

**Tableau 9.** Nombre de lectures obtenues pour chaque banque construite à partir de différents extraits d'ADN venant de différentes fractions du coprolithe CCH3. a : lectures de taille comprise entre 10 et 76 nucléotides ; b : lectures de taille comprise entre 10 et 101 nucléotides.

Échantillon	Banque	Fraction utilisée	Volume d'extrait utilisé	Nombre de lectures
CCH3	1	Périphérie + Cœur	5 µl	273932357 <sup>a</sup>
CCH3	2	Périphérie + Cœur	10 µl	292309494 <sup>a</sup>
CCH3	3	Cœur	10 µl	504 145 349 <sup>b</sup>

### 3.4. Génome mitochondrial du canidé de la Grotte Chauvet-Pont d'Arc

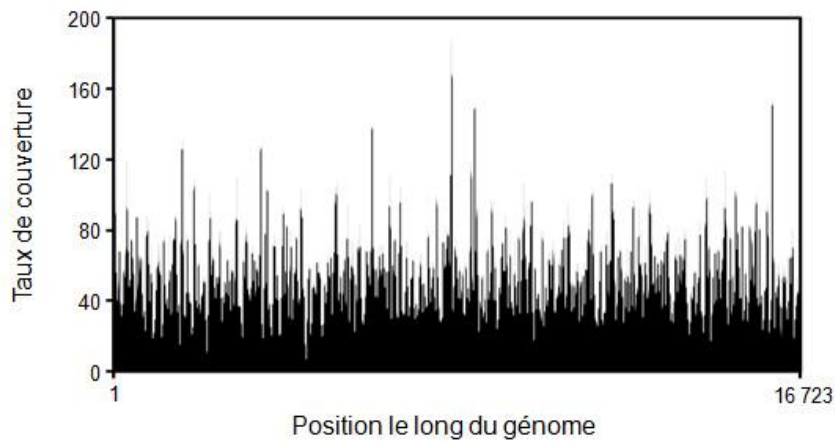
#### 3.4.1. Reconstitution et analyse du génome mitochondrial complet

Dans le but de reconstituer un génome mitochondrial complet pour le canidé de la Grotte Chauvet-Pont d'Arc, une stratégie en plusieurs temps a été mise en place.

Tout d'abord, les contigs des banques 1 et 2, construits grâce au logiciel SOAP, ont été alignés sur le génome mitochondrial d'un loup gris espagnol (Björnerfledt *et al.*, 2006 ; DQ480505) au moyen de Sequencher avec comme critère d'alignement une identité minimale de 90%. Au total, 91 contigs pour lesquels le meilleur score d'alignement par BLAST sur *GenBank nt* est systématiquement une séquence de canidé (*Canis lupus* ou *Canis lupus familiaris*) ont été obtenus. Leur taille est comprise entre 24 et 1991 nucléotides, et leur

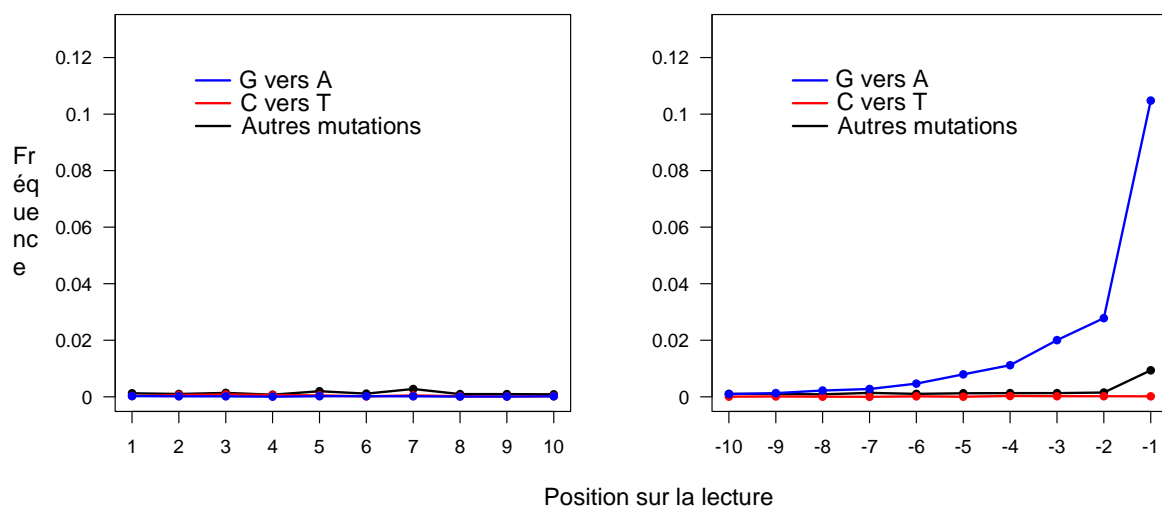
assemblage permet de couvrir l'intégralité du génome à l'exception de 204 positions, la plupart d'entre elles (198 positions) situées dans la partie de la région de contrôle constituée d'un motif répété. Une séquence provisoire pour notre échantillon a été déduite des contigs, avec pour les régions non couvertes un complément à partir du génome du loup espagnol.

Puis, les lectures des banques 1 et 2 ont été alignées sur la séquence provisoire, avec un crible tolérant un mésappariement et un *indel*. Après élimination des doublons, nous avons obtenu 22 992 lectures Illumina qui ont permis de reconstruire un premier génome mitochondrial complet de 16 723 paires de bases. Le taux de couverture moyen du génome est de 51,9X, et la redondance minimale pour chaque position est de 7 X (Figure 42).



**Figure 42.** Nombre de lectures pour chaque position du génome mitochondrial reconstitué avec les 22 992 lectures des banques 1 et 2.

En parallèle, une fois, les banques séquencées, les dommages possiblement subis par les séquences d'ADN ont été analysés. Dans ce but, la position des mutations le long des lectures a été étudiée pour les 10 premières et les 10 dernières bases (Figure 43).



**Figure 43.** Fréquence des mutations observées pour les 10 premières et les 10 dernières positions des séquences. L'analyse est réalisée en comparant la séquence des lectures utilisées pour reconstituer le génome mitochondrial à la séquence consensus obtenue.

Une augmentation des substitutions G vers A est observée, mais uniquement à l'extrémité 3' des lectures. En revanche, aucune augmentation des transitions C vers T n'est visible pour les deux extrémités. Cette observation est en accord avec ce qui avait déjà été rapporté pour des banques d'ADN anciens construites selon le même protocole (Jónsson *et al.*, 2013 ; Marsolier-Kergoat *et al.*, 2015). Une augmentation des dommages aux extrémités des séquences d'ADN est une preuve de l'authenticité de celui-ci (Briggs *et al.*, 2007).

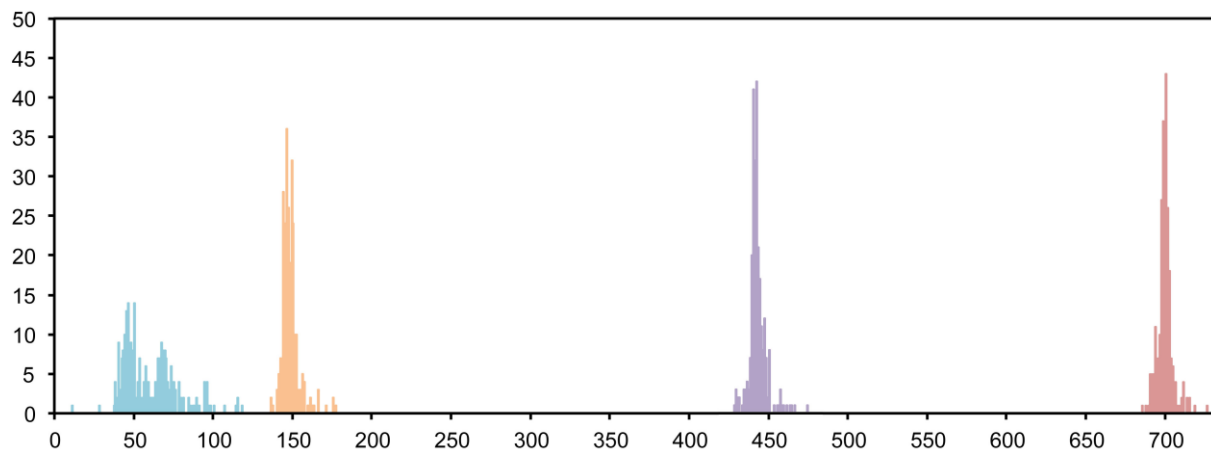
Lorsque les 4 pistes de la banque 3 ont été séquencées, un second consensus a été construit en alignant les lectures de cette banque sur le premier consensus. La comparaison des deux séquences obtenues a montré que ces deux génomes étaient identiques. La reconstruction du génome mitochondrial du canidé de Chauvet est donc reproductible.

En alignant les lectures des 3 banques sur le consensus, un nombre total de 38 885 lectures a été obtenu, ce qui donne un génome mitochondrial complet avec un taux de couverture de 89 X.

Une fois la séquence mitochondriale reconstituée, ses caractéristiques ont été analysées. Le contenu en GC (39,5 %) est similaire à celui rapporté pour des séquences de référence de *Canis lupus* (39,6 %, DQ480505) et de *Canis lupus familiaris* (39,6 %, U96639) actuels. Le génome obtenu présente les caractéristiques attendues pour un génome mitochondrial de mammifères : c'est un génome circulaire, comportant une région de contrôle, 2 gènes de rDNA, 22 gènes de tDNA, et 13 gènes codant des protéines. La qualité

de la séquence est notamment attestée par l'absence de codons stop internes dans les gènes codant des protéines. Les tailles prédites pour les séquences protéiques sont similaires à celles des protéines orthologues de canidés.

La séquence de notre génome mitochondrial a ensuite été comparée aux génomes mitochondriaux de chiens et loups actuels disponibles dans *GenBank* en construisant une matrice de différences. Les génomes mitochondriaux du loup du Tibet (*Canis lupus laniger*) et du coyote (*Canis latrans*) ont été utilisés comme groupe externe (Figure 44).



**Figure 44.** Graphique réalisé à partir de la matrice de différences construite par comparaison entre les génomes mitochondriaux complets du canidé de Chauvet (orange), de 245 chiens et 15 loups actuels (bleu), du loup du Tibet (violet) et du coyote (rouge). L'abscisse représente le nombre de différences et l'ordonnée le nombre de spécimens.

Le graphique réalisé à partir de la matrice de différences montre que le canidé de Chauvet se situe en dehors de la diversité génétique des chiens et loups actuels ; cependant, il ne présente pas autant de différences avec ces derniers que *Canis lupus laniger*, qui est une sous-espèce bien différenciée. Le nombre de différences entre notre canidé et les chiens et loups actuels est important : avec 136 à 177 différences, les valeurs sont proches de celles observées par Green *et al.* (2008) entre le génome mitochondrial d'un néandertalien et 53 génomes mitochondriaux d'*Homo sapiens* actuels (201 à 234 différences).

Les génomes mitochondriaux de canidés utilisés pour construire la matrice de différences sont présentés dans la table ci-dessous (Tableau 10).

**Tableau 10.** Génomes mitochondriaux utilisés pour la construction de la matrice de différences. Les références correspondent à 1 : Ishiguro *et al.*, 2009 ; 2 : Arnason *et al.*, 2007 ; 3 : Bjornerfeldt *et al.*, 2006 ; 4 : Pang *et al.*, 2009 ; 5 : Webb *et al.*, 2009 ; 6 : Baranowska *et al.*, 2009 ; 7 : Li *et al.*, 2011 ; 8 : Kim *et al.*, 1998 ; 10 : Meng *et al.*, 2009.

Espèce	Nom vernaculaire	GenBank	Localisation	Race	Référence
<i>Canis lupus</i>	loup gris eurasiatique	AB499818	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499819	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499820	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499821	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499822	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499823	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499824	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	AB499825	Japon		1
<i>Canis lupus</i>	loup gris eurasiatique	NC_009686 (AM711902)	Suède		2
<i>Canis lupus</i>	loup gris eurasiatique	DQ480503	Russie		3
<i>Canis lupus</i>	loup gris eurasiatique	DQ480504	Suède		3
<i>Canis lupus</i>	loup gris eurasiatique	NC_008092 (DQ480505)	Espagne		3
<i>Canis lupus</i>	loup gris eurasiatique	DQ480506	Arabie saoudite		3
<i>Canis lupus</i>	loup gris eurasiatique	DQ480507	Arabie saoudite		3
<i>Canis lupus</i>	loup gris eurasiatique	DQ480508	Canada		3
<i>Canis lupus</i>	loup gris eurasiatique	EU789787	Chine		4
<i>Canis lupus</i>	loup gris eurasiatique	EU789788	Chine		4
<i>Canis lupus familiaris</i>	chien	AB499816		Kisyu	1
<i>Canis lupus familiaris</i>	chien	AB499817		Husky	1
<i>Canis lupus familiaris</i>	chien	AY656737		Basenji	Non publié
<i>Canis lupus familiaris</i>	chien	AY656738		Jack Russell	Non publié
<i>Canis lupus familiaris</i>	chien	AY656739		Caniche	Non publié
<i>Canis lupus familiaris</i>	chien	AY656740		Terrier Kerry Blue	Non publié
<i>Canis lupus familiaris</i>	chien	AY656741		Setter irlandais	Non publié
<i>Canis lupus familiaris</i>	chien	AY656742		Vieux chien de berger anglais	Non publié
<i>Canis lupus familiaris</i>	chien	AY656743		Saint-Bernard	Non publié
<i>Canis lupus familiaris</i>	chien	AY656744		Springer anglais	Non publié
<i>Canis lupus familiaris</i>	chien	AY656745		Springer anglais	Non publié
<i>Canis lupus familiaris</i>	chien	AY656746		Schnauzer moyen	Non publié
<i>Canis lupus familiaris</i>	chien	AY656747		Welsh Springer Spaniel	Non publié
<i>Canis lupus familiaris</i>	chien	AY656748		Airedale Terrier	Non publié
<i>Canis lupus familiaris</i>	chien	AY656749		Saint-Bernard	Non publié
<i>Canis lupus familiaris</i>	chien	AY656750		Leonberg	Non publié
<i>Canis lupus familiaris</i>	chien	AY656751		Setter Gordon	Non publié
<i>Canis lupus familiaris</i>	chien	AY656752		Schnauzer moyen	Non publié
<i>Canis lupus familiaris</i>	chien	AY656753		Setter irlandais	Non publié
<i>Canis lupus familiaris</i>	chien	AY656754		Chien chinois	Non publié
<i>Canis lupus familiaris</i>	chien	AY656755		Sapsali	Non publié
<i>Canis lupus familiaris</i>	chien	AY729880		Beagle	Non publié
<i>Canis lupus familiaris</i>	chien	DQ480489		Berger allemand	2

<i>Canis lupus familiaris</i>	chien	DQ480490	Retriever à poil plat	2
<i>Canis lupus familiaris</i>	chien	DQ480491	Setter irlandais	3
<i>Canis lupus familiaris</i>	chien	DQ480492	Chien d'élan suédois	3
<i>Canis lupus familiaris</i>	chien	DQ480493	Terrier noir russe	3
<i>Canis lupus familiaris</i>	chien	DQ480494	Caniche	3
<i>Canis lupus familiaris</i>	chien	DQ480495	Cocker spaniel anglais	3
<i>Canis lupus familiaris</i>	chien	DQ480496	Terrier irlandais à poil doux	3
<i>Canis lupus familiaris</i>	chien	DQ480497	West Highland White Terrier	3
<i>Canis lupus familiaris</i>	chien	DQ480498	Schnauzer miniature	3
<i>Canis lupus familiaris</i>	chien	DQ480499	Husky de Sibérie	3
<i>Canis lupus familiaris</i>	chien	DQ480500	Berger des Shetland	3
<i>Canis lupus familiaris</i>	chien	DQ480501	Chien d'élan suédois	3
<i>Canis lupus familiaris</i>	chien	DQ480502	Jamthund	3
<i>Canis lupus familiaris</i>	chien	EU408245	Akita	5
<i>Canis lupus familiaris</i>	chien	EU408246	Cocker américain	5
<i>Canis lupus familiaris</i>	chien	EU408247	Terrier australien	5
<i>Canis lupus familiaris</i>	chien	EU408248	Berger australien	5
<i>Canis lupus familiaris</i>	chien	EU408249	Berger australien	5
<i>Canis lupus familiaris</i>	chien	EU408250	Bichon à poil frisé	5
<i>Canis lupus familiaris</i>	chien	EU408251	Bouvier australien	5
<i>Canis lupus familiaris</i>	chien	EU408252	Bichon bolonais	5
<i>Canis lupus familiaris</i>	chien	EU408253	Boxer	5
<i>Canis lupus familiaris</i>	chien	EU408254	Basset hound	5
<i>Canis lupus familiaris</i>	chien	EU408255	Basset hound	5
<i>Canis lupus familiaris</i>	chien	EU408257	Épagneul breton	5
<i>Canis lupus familiaris</i>	chien	EU408258	Cockapoo	5
<i>Canis lupus familiaris</i>	chien	EU408259	Cockapoo	5
<i>Canis lupus familiaris</i>	chien	EU408260	Cardigan Corgi	5
<i>Canis lupus familiaris</i>	chien	EU408261	Chihuahua	5
<i>Canis lupus familiaris</i>	chien	EU408262	Chihuahua	5
<i>Canis lupus familiaris</i>	chien	EU408263	Cavalier King Charles	5
<i>Canis lupus familiaris</i>	chien	EU408264	Cairn terrier	5
<i>Canis lupus familiaris</i>	chien	EU408265	Corgi	5
<i>Canis lupus familiaris</i>	chien	EU408266	Cocker anglais	5
<i>Canis lupus familiaris</i>	chien	EU408267	Cocker anglais	5
<i>Canis lupus familiaris</i>	chien	EU408268	Cocker anglais	5
<i>Canis lupus familiaris</i>	chien	EU408269	Doberman	5
<i>Canis lupus familiaris</i>	chien	EU408272	Teckel	5
<i>Canis lupus familiaris</i>	chien	EU408273	Colley de ferme	5
<i>Canis lupus familiaris</i>	chien	EU408274	Mastiff	5
<i>Canis lupus familiaris</i>	chien	EU408275	Bouledogue	5
<i>Canis lupus familiaris</i>	chien	EU408276	Dogue allemand	5
<i>Canis lupus familiaris</i>	chien	EU408277	German Shepherd	5
<i>Canis lupus familiaris</i>	chien	EU408278	Great Pyrenees	5
<i>Canis lupus familiaris</i>	chien	EU408279	Havanese	5

<i>Canis lupus familiaris</i>	chien	EU408280	Italian Greyhound	5
<i>Canis lupus familiaris</i>	chien	EU408281	Jack Russell	5
<i>Canis lupus familiaris</i>	chien	EU408282	Keeshond	5
<i>Canis lupus familiaris</i>	chien	EU408285	Labradoodle	5
<i>Canis lupus familiaris</i>	chien	EU408286	Miniature Dachshund	5
<i>Canis lupus familiaris</i>	chien	EU408287	Newfoundland	5
<i>Canis lupus familiaris</i>	chien	EU408288	Norwegian Elkhound	5
<i>Canis lupus familiaris</i>	chien	EU408289	Neapolitan Mastiff	5
<i>Canis lupus familiaris</i>	chien	EU408290	Neapolitan Mastiff	5
<i>Canis lupus familiaris</i>	chien	EU408291	Pomeranian	5
<i>Canis lupus familiaris</i>	chien	EU408292	Poodle	5
<i>Canis lupus familiaris</i>	chien	EU408293	Pit Bull Terrier	5
<i>Canis lupus familiaris</i>	chien	EU408294	Pug	5
<i>Canis lupus familiaris</i>	chien	EU408295	Rottweiler	5
<i>Canis lupus familiaris</i>	chien	EU408297	Schipperke	5
<i>Canis lupus familiaris</i>	chien	EU408300	Tibetan Mastiff	5
<i>Canis lupus familiaris</i>	chien	EU408301	Tibetan Spaniel	5
<i>Canis lupus familiaris</i>	chien	EU408302	Toy Poodle	5
<i>Canis lupus familiaris</i>	chien	EU408303		5
<i>Canis lupus familiaris</i>	chien	EU408304		5
<i>Canis lupus familiaris</i>	chien	EU408305	Viszla	5
<i>Canis lupus familiaris</i>	chien	EU408307	Walker Hound	5
<i>Canis lupus familiaris</i>	chien	EU408308	Yorkie Chihuahua	5
<i>Canis lupus familiaris</i>	chien	EU789638		4
<i>Canis lupus familiaris</i>	chien	EU789639		4
<i>Canis lupus familiaris</i>	chien	EU789640		4
<i>Canis lupus familiaris</i>	chien	EU789641		4
<i>Canis lupus familiaris</i>	chien	EU789642		4
<i>Canis lupus familiaris</i>	chien	EU789643		4
<i>Canis lupus familiaris</i>	chien	EU789644		4
<i>Canis lupus familiaris</i>	chien	EU789645		4
<i>Canis lupus familiaris</i>	chien	EU789646		4
<i>Canis lupus familiaris</i>	chien	EU789647		4
<i>Canis lupus familiaris</i>	chien	EU789648		4
<i>Canis lupus familiaris</i>	chien	EU789649		4
<i>Canis lupus familiaris</i>	chien	EU789650		4
<i>Canis lupus familiaris</i>	chien	EU789651	Shar-pei	4
<i>Canis lupus familiaris</i>	chien	EU789652	Collie	4
<i>Canis lupus familiaris</i>	chien	EU789653		4
<i>Canis lupus familiaris</i>	chien	EU789654	Kangal	4
<i>Canis lupus familiaris</i>	chien	EU789655	Galgo Español	4
<i>Canis lupus familiaris</i>	chien	EU789656	Finnish Lapphund	4
<i>Canis lupus familiaris</i>	chien	EU789657		4
<i>Canis lupus familiaris</i>	chien	EU789658	Husky	4
<i>Canis lupus familiaris</i>	chien	EU789659		4

<i>Canis lupus familiaris</i>	chien	EU789660		4
<i>Canis lupus familiaris</i>	chien	EU789661		4
<i>Canis lupus familiaris</i>	chien	EU789662	Pungsan	4
<i>Canis lupus familiaris</i>	chien	EU789663		4
<i>Canis lupus familiaris</i>	chien	EU789664		4
<i>Canis lupus familiaris</i>	chien	EU789665		4
<i>Canis lupus familiaris</i>	chien	EU789666		4
<i>Canis lupus familiaris</i>	chien	EU789667		4
<i>Canis lupus familiaris</i>	chien	EU789668		4
<i>Canis lupus familiaris</i>	chien	EU789669		4
<i>Canis lupus familiaris</i>	chien	EU789670		4
<i>Canis lupus familiaris</i>	chien	EU789671		4
<i>Canis lupus familiaris</i>	chien	EU789672		4
<i>Canis lupus familiaris</i>	chien	EU789673		4
<i>Canis lupus familiaris</i>	chien	EU789674		4
<i>Canis lupus familiaris</i>	chien	EU789675		4
<i>Canis lupus familiaris</i>	chien	EU789676		4
<i>Canis lupus familiaris</i>	chien	EU789677		4
<i>Canis lupus familiaris</i>	chien	EU789678		4
<i>Canis lupus familiaris</i>	chien	EU789679	Saluki	4
<i>Canis lupus familiaris</i>	chien	EU789680		4
<i>Canis lupus familiaris</i>	chien	EU789681		4
<i>Canis lupus familiaris</i>	chien	EU789682		4
<i>Canis lupus familiaris</i>	chien	EU789683	Border Collie	4
<i>Canis lupus familiaris</i>	chien	EU789684		4
<i>Canis lupus familiaris</i>	chien	EU789685		4
<i>Canis lupus familiaris</i>	chien	EU789686	Sloughi	4
<i>Canis lupus familiaris</i>	chien	EU789687		4
<i>Canis lupus familiaris</i>	chien	EU789688		4
<i>Canis lupus familiaris</i>	chien	EU789689		4
<i>Canis lupus familiaris</i>	chien	EU789690		4
<i>Canis lupus familiaris</i>	chien	EU789691		4
<i>Canis lupus familiaris</i>	chien	EU789692		4
<i>Canis lupus familiaris</i>	chien	EU789693		4
<i>Canis lupus familiaris</i>	chien	EU789694		4
<i>Canis lupus familiaris</i>	chien	EU789695		4
<i>Canis lupus familiaris</i>	chien	EU789696		4
<i>Canis lupus familiaris</i>	chien	EU789697	Irish Soft Coated Wheaten Terrier	4
<i>Canis lupus familiaris</i>	chien	EU789698		4
<i>Canis lupus familiaris</i>	chien	EU789699		4
<i>Canis lupus familiaris</i>	chien	EU789700		4
<i>Canis lupus familiaris</i>	chien	EU789701		4
<i>Canis lupus familiaris</i>	chien	EU789702		4
<i>Canis lupus familiaris</i>	chien	EU789703		4
<i>Canis lupus familiaris</i>	chien	EU789704		4

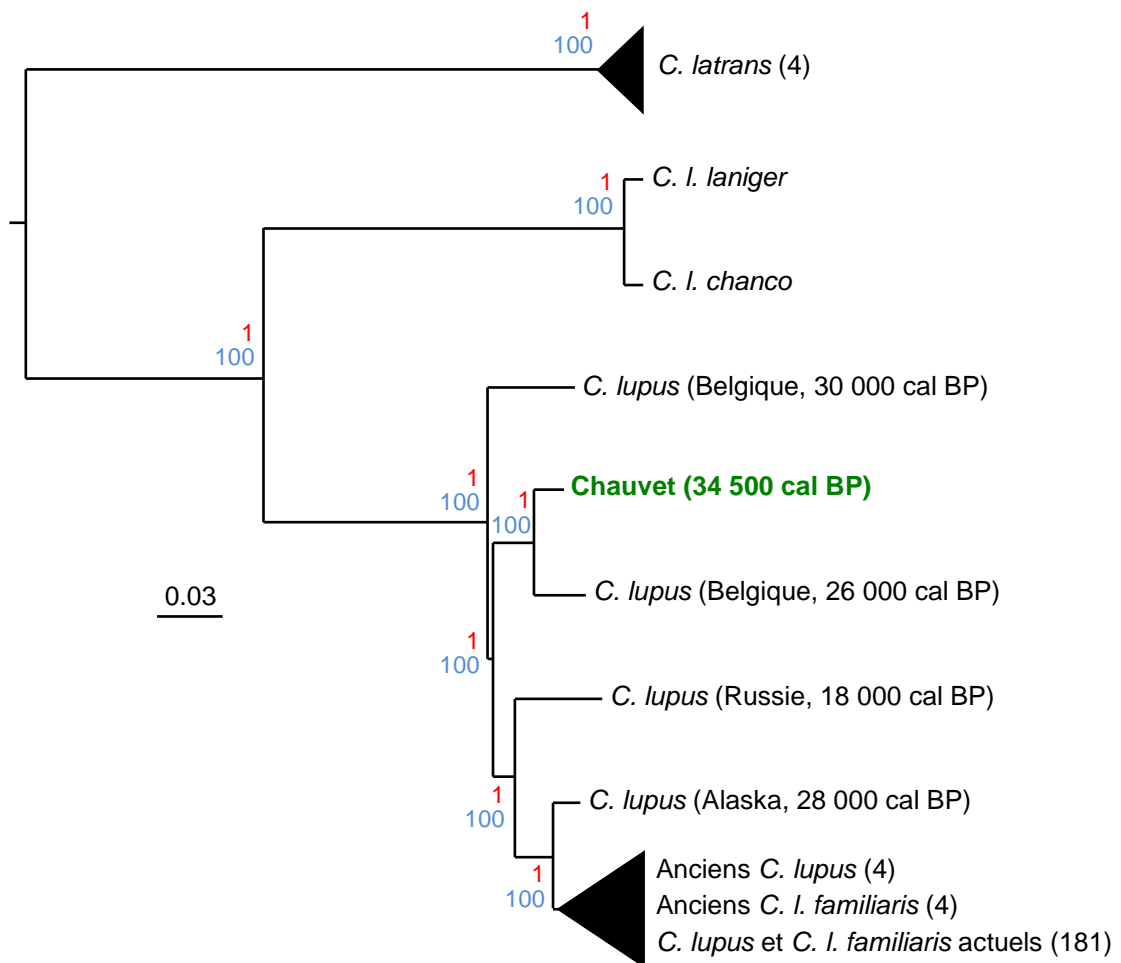


<i>Canis lupus familiaris</i>	chien	EU789705		4
<i>Canis lupus familiaris</i>	chien	EU789706	Hamiltonstövare	4
<i>Canis lupus familiaris</i>	chien	EU789707		4
<i>Canis lupus familiaris</i>	chien	EU789708		4
<i>Canis lupus familiaris</i>	chien	EU789709	Pyrenean	4
<i>Canis lupus familiaris</i>	chien	EU789710		4
<i>Canis lupus familiaris</i>	chien	EU789711		4
<i>Canis lupus familiaris</i>	chien	EU789712	Laika, East/West siberian	4
<i>Canis lupus familiaris</i>	chien	EU789713	Sloughi	4
<i>Canis lupus familiaris</i>	chien	EU789714	Galgo Español	4
<i>Canis lupus familiaris</i>	chien	EU789715		4
<i>Canis lupus familiaris</i>	chien	EU789716		4
<i>Canis lupus familiaris</i>	chien	EU789717		4
<i>Canis lupus familiaris</i>	chien	EU789718	Poodle	4
<i>Canis lupus familiaris</i>	chien	EU789719		4
<i>Canis lupus familiaris</i>	chien	EU789720		4
<i>Canis lupus familiaris</i>	chien	EU789721		4
<i>Canis lupus familiaris</i>	chien	EU789722		4
<i>Canis lupus familiaris</i>	chien	EU789723		4
<i>Canis lupus familiaris</i>	chien	EU789724		4
<i>Canis lupus familiaris</i>	chien	EU789725		4
<i>Canis lupus familiaris</i>	chien	EU789726		4
<i>Canis lupus familiaris</i>	chien	EU789727		4
<i>Canis lupus familiaris</i>	chien	EU789728	Golden Retriever	4
<i>Canis lupus familiaris</i>	chien	EU789729		4
<i>Canis lupus familiaris</i>	chien	EU789730		4
<i>Canis lupus familiaris</i>	chien	EU789731		4
<i>Canis lupus familiaris</i>	chien	EU789732		4
<i>Canis lupus familiaris</i>	chien	EU789733		4
<i>Canis lupus familiaris</i>	chien	EU789734		4
<i>Canis lupus familiaris</i>	chien	EU789735		4
<i>Canis lupus familiaris</i>	chien	EU789736		4
<i>Canis lupus familiaris</i>	chien	EU789737		4
<i>Canis lupus familiaris</i>	chien	EU789738		4
<i>Canis lupus familiaris</i>	chien	EU789739		4
<i>Canis lupus familiaris</i>	chien	EU789740		4
<i>Canis lupus familiaris</i>	chien	EU789741		4
<i>Canis lupus familiaris</i>	chien	EU789742	Laika, East/West siberian	4
<i>Canis lupus familiaris</i>	chien	EU789743		4
<i>Canis lupus familiaris</i>	chien	EU789744		4
<i>Canis lupus familiaris</i>	chien	EU789745		4
<i>Canis lupus familiaris</i>	chien	EU789746		4
<i>Canis lupus familiaris</i>	chien	EU789747		4
<i>Canis lupus familiaris</i>	chien	EU789748		4
<i>Canis lupus familiaris</i>	chien	EU789749		4

<i>Canis lupus familiaris</i>	chien	EU789750		4
<i>Canis lupus familiaris</i>	chien	EU789751		4
<i>Canis lupus familiaris</i>	chien	EU789752		4
<i>Canis lupus familiaris</i>	chien	EU789753		4
<i>Canis lupus familiaris</i>	chien	EU789754		4
<i>Canis lupus familiaris</i>	chien	EU789755	Japanese Splitz	4
<i>Canis lupus familiaris</i>	chien	EU789756		4
<i>Canis lupus familiaris</i>	chien	EU789757	Tazi	4
<i>Canis lupus familiaris</i>	chien	EU789758	Afghan	4
<i>Canis lupus familiaris</i>	chien	EU789759	Basenji	4
<i>Canis lupus familiaris</i>	chien	EU789760	Bouvier	4
<i>Canis lupus familiaris</i>	chien	EU789761	Pungsan	4
<i>Canis lupus familiaris</i>	chien	EU789762	Jindo	4
<i>Canis lupus familiaris</i>	chien	EU789763	Spinone Italiano	4
<i>Canis lupus familiaris</i>	chien	EU789764	Jamthund	4
<i>Canis lupus familiaris</i>	chien	EU789765	Bouvier	4
<i>Canis lupus familiaris</i>	chien	EU789766		4
<i>Canis lupus familiaris</i>	chien	EU789767		4
<i>Canis lupus familiaris</i>	chien	EU789768	Laika, East/West siberian	4
<i>Canis lupus familiaris</i>	chien	EU789769		4
<i>Canis lupus familiaris</i>	chien	EU789770	Braque	4
<i>Canis lupus familiaris</i>	chien	EU789771		4
<i>Canis lupus familiaris</i>	chien	EU789772	Pungsan	4
<i>Canis lupus familiaris</i>	chien	EU789773	Jamthund	4
<i>Canis lupus familiaris</i>	chien	EU789774	Jamthund	4
<i>Canis lupus familiaris</i>	chien	EU789775	Shetland Sheepchien	4
<i>Canis lupus familiaris</i>	chien	EU789776	Siberian Husky	4
<i>Canis lupus familiaris</i>	chien	EU789777	Miniature Schnauzer	4
<i>Canis lupus familiaris</i>	chien	EU789778	West Highland White Terrier	4
<i>Canis lupus familiaris</i>	chien	EU789779	Irish Soft Coated Wheaten Terrier	4
<i>Canis lupus familiaris</i>	chien	EU789780	Cocker Spaniel	4
<i>Canis lupus familiaris</i>	chien	EU789782	Black Russian Terrier	4
<i>Canis lupus familiaris</i>	chien	EU789783	Jamthund	4
<i>Canis lupus familiaris</i>	chien	EU789784	Irish Setter	4
<i>Canis lupus familiaris</i>	chien	EU789785	Flat Coated Retriever	4
<i>Canis lupus familiaris</i>	chien	EU789786	German Shepherd	4
<i>Canis lupus familiaris</i>	chien	FJ817358	Golden retriever	6
<i>Canis lupus familiaris</i>	chien	FJ817362	Golden retriever	6
<i>Canis lupus familiaris</i>	chien	FJ817363	Golden retriever	6
<i>Canis lupus familiaris</i>	chien	FJ817364	Golden retriever	6
<i>Canis lupus familiaris</i>	chien	HM048871	Tibetan Mastiff	7
<i>Canis lupus familiaris</i>	chien	NC_002008 (U96639)	Sapsaree	8
<i>Canis lupus laniger</i>	Loup du Tibet	NC_011218 (FJ032363)	Tibet	10
<i>Canis latrans</i>	coyote	NC_008093 (DQ480509)	États-Unis	3

### 3.4.2. Position phylogénétique du génome mitochondrial du canidé de Chauvet

Le fort taux de couverture de notre génome mitochondrial assurant une séquence mitochondriale fiable, la position phylogénétique du spécimen de Chauvet a pu être explorée. Les arbres phylogénétiques ont été construits au moyen de différentes méthodes : Neighbor Joining, Maximum Likelihood et Minimum Evolution en utilisant la suite Mega6, et analyse bayésienne avec le logiciel MrBayes. Le set de données comprend des séquences de chiens (121 génomes) et de loups (60 génomes) actuels, et des séquences attribuées à des spécimens anciens de loup (8 génomes) ou de chien (4 génomes). Parmi les génomes anciens de loup ou de chien publiés par Thalmann *et al.* (2013), nous n'avons pris en compte que ceux dont la couverture est supérieure à 8 X. Le loup du Tibet (*Canis lupus laniger*) et le loup de Mongolie (*Canis lupus chanco*) ont également été considérés ; le coyote (*Canis latrans*, 4 génomes) a été utilisé comme groupe externe (Figure 45).



**Figure 45.** Position phylogénétique du canidé de la Grotte Chauvet-Pont d'Arc à partir des séquences mitochondriales. Les chiffres en rouge indiquent les valeurs des probabilités a posteriori pour l'analyse bayésienne et les nombres en bleu les valeurs de *Bootstrap* pour les autres méthodes d'analyse.

La topologie de l'arbre est la même pour toutes les méthodes de construction et les valeurs des nœuds sont toujours égales à 100 pour les *Bootstrap* et à 1 pour les probabilités *a posteriori* ; ce qui apporte une grande fiabilité à l'arbre phylogénétique obtenu. Celui-ci confirme les résultats obtenus avec la matrice de différences : la séquence mitochondriale du canidé de Chauvet se situe en dehors de la diversité des chiens et loups actuels.

Son génome se positionne à proximité de ceux attribués aux loups de Belgique du Pléistocène ; cette observation est en accord avec l'âge des échantillons (26 000 et 30 000 ans) et leur origine (Europe occidentale). Ces génomes correspondent à d'anciennes lignées maternelles bien distinctes de la diversité des chiens et loups actuels, et qui ont aujourd'hui disparu.

### 3.5. Analyse du génome nucléaire

Compte tenu de la position phylogénétique particulière du canidé de la Grotte Chauvet, une stratégie mixte a été mise en place pour caractériser le génome nucléaire du spécimen. Une approche globale avec alignement des lectures sur le génome nucléaire du chien a été réalisée, et une analyse ciblée de certains gènes a également été entreprise.

#### 3.5.1. Analyse globale du génome nucléaire

En collaboration avec la plate-forme eBio de l'Université Paris XI, les lectures des 3 banques ont été alignées sur le génome nucléaire du chien (CanFam3.1). Nous avons toléré pour l'alignement 1 mésappariement et 1 indel pour les lectures courtes (taille comprise entre 25 et 33 nucléotides) et deux mésappariements ainsi que 3 *indel* pour les lectures longues (taille comprise entre 34 et 99 nucléotides). La prise en compte des *indel* est justifiée par le fait que les gènes transcrits ne comportent pas forcément le même nombre de codons dans le génome du spécimen de la Grotte Chauvet et le génome de référence du chien. Un génome nucléaire partiel avec un taux de couverture globale de 0,65 X a ainsi été obtenu. Le taux de couverture pour chaque chromosome a également été calculé (Figure 46).



**Figure 46.** Taux de couverture des 38 autosomes et du chromosome sexuel X.

Le taux de couverture de chaque chromosome correspond à celui qui a été trouvé pour le génome nucléaire global, ce qui confirme la fiabilité de l'analyse. La couverture du chromosome sexuel X étant identique à celles des autosomes, il est logique de penser que le canidé de Chauvet était une femelle. Pour confirmer cette observation, le taux de couverture

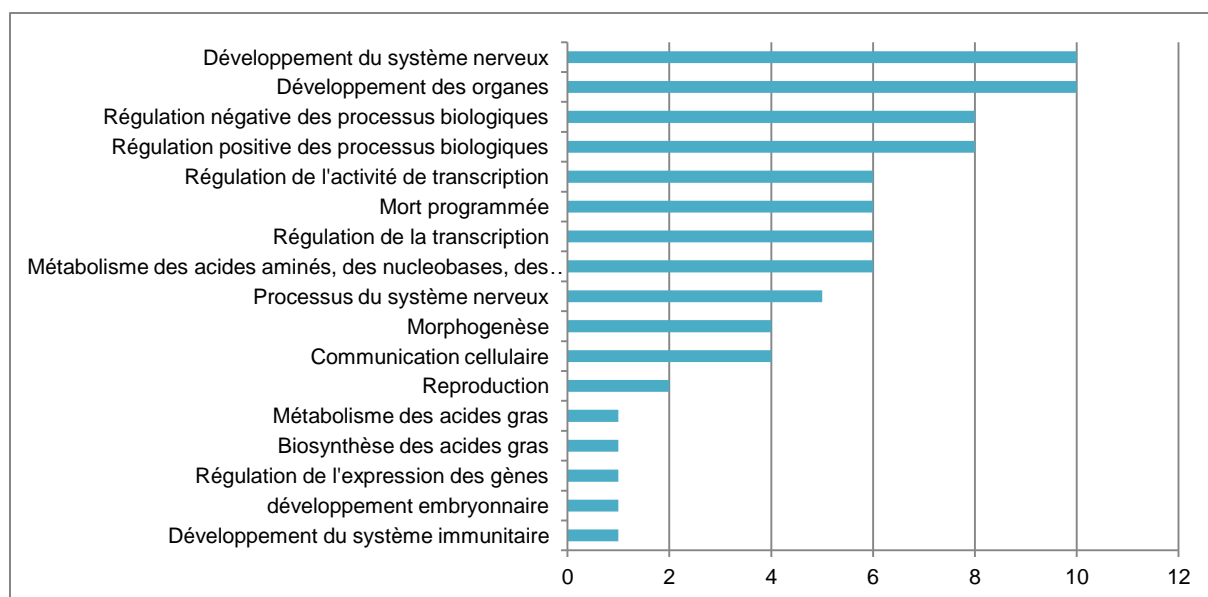
du gène *Gabra3* - présent sur le chromosome X - a été calculé par alignement des lectures des 3 banques sur la séquence du gène. Le taux de couverture obtenu est de 0,6 X ce qui est identique à la couverture observée par alignement global. Le spécimen de la grotte Chauvet était donc bien une femelle.

### 3.5.2. Analyse des *Gene ontology*

Grâce aux informations contenues dans le fichier VCF, nous avons cherché à savoir si, entre le génome nucléaire du chien et notre génome partiel, une ou plusieurs catégories précises de gènes présentaient un plus grand nombre de mutations que les autres. Pour cela, un protocole informatique ciblant uniquement les mutations présentes dans les régions codantes a été mis en place. Puis, un tri a été effectué pour ne garder que les différences entraînant des mutations soit dans les initiateurs de transcription, soit dans les régions codantes, soit faux-sens dans les phases ouvertes de lecture ; il a permis d'obtenir respectivement 17, 34 et 4 692 positions du génome nucléaire.

Les 4 743 lignes ont ensuite été soumises à deux cribles distincts. Le premier consiste à supprimer les différences créées par des transitions, car celles-ci peuvent être induites lors de la dégradation de l'ADN ancien. Ce crible nous laisse 2 526 lignes. Le second crible ne garde que les différences annotées comme génotype "homozygote alternatif" dans le fichier VCF. Ce second crible nous a permis d'obtenir 2 238 lignes.

Puis, séparément, les ENSCAFG de ces lignes ont été extraits et comparés à ceux du chien, disponibles dans la base de données *Ensembl*. Nous avons eu 2 510 ENSCAFG dont 2 117 noms de gènes connus pour le premier crible et 2 397 ENSCAFG dont 1 946 noms de gènes connus pour le second crible. Toujours séparément, les ENSCAFG obtenus pour les deux cribles ont été rentrés dans la base de données *Gene Ontology*. En ne sélectionnant que les *GO* qui ont un score de qualité inférieur à 0,05, on obtient 476 *GO* pour le crible sans transition et 391 *GO* pour le crible ne prenant en compte que les homozygotes alternatifs. Le recoupement des résultats obtenus pour nos deux cribles nous donne 278 *GO* en commun. Puis, en sélectionnant les *GO* qui avaient des fonctions bien précises et non trop générales comme par exemple « biological process », 80 *GO* d'intérêt ont été retenus. Les catégories auxquelles ces *GO* appartiennent ont été analysées (Figure 47).



**Figure 47.** Catégories de gènes pour lesquelles le génome nucléaire du spécimen de Chauvet présente des différences dans ses régions codantes par rapport à celui du chien.

Les deux grandes catégories de *GO* qui reviennent le plus souvent pour le canidé de Chauvet sont celle du développement des structures anatomiques et celle de la régulation des processus biologiques. Quand une analyse plus fine des sous-catégories est effectuée, il est possible de voir que les *GO* les plus courants sont ceux impliqués dans le développement du système nerveux, le développement des organes et la régulation positive et négative des processus biologiques. Dans un article paru en 2013, les *GO* présentant des différences entre les chiens et loups actuels avaient été analysés (Axelsson *et al.*, 2013). Sur les 25 *GO* qui présentaient des différences majeures selon l'article, nous en retrouvons 9 qui ressortent également pour le canidé de Chauvet (8 impliquant les neurones et 1 pour le métabolisme des acides gras).

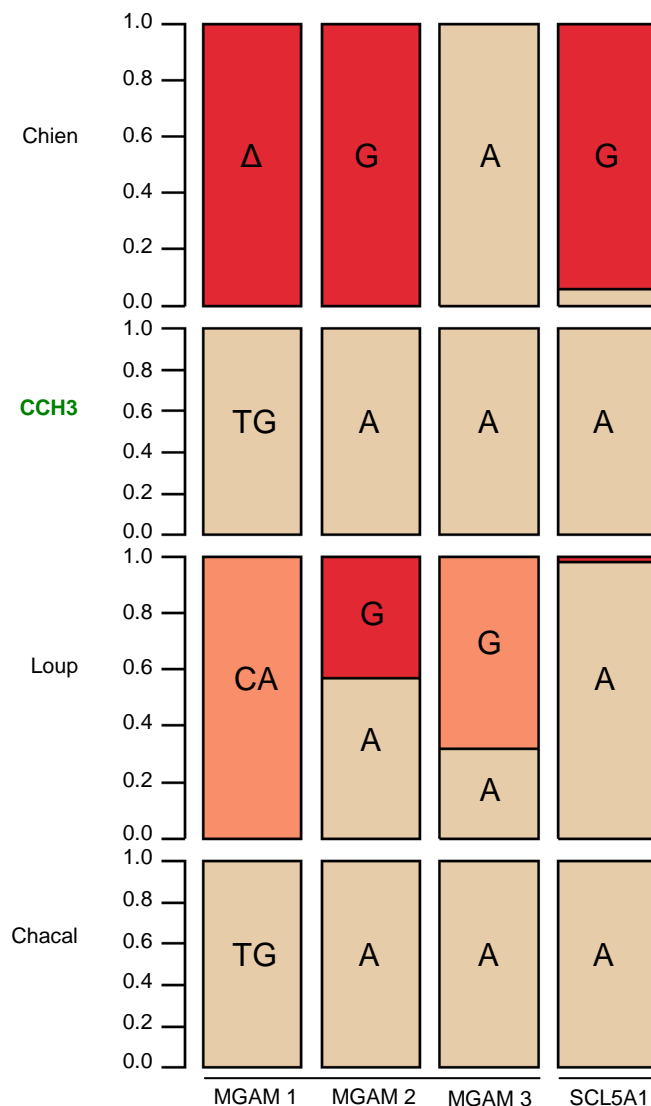
### 3.5.3. Analyse ciblée de gènes nucléaires

#### 3.5.3.1. Analyse de gènes impliqués dans la digestion

Dans ce même article (Axelsson *et al.*, 2013), il apparaissait qu'il y avait eu une sélection spécifique de gènes impliqués dans la digestion de l'amidon lors de la domestication du loup.

Nous nous sommes focalisés sur deux gènes : MGAM qui catalyse l'hydrolyse du maltose en glucose et SLC5A1 qui code pour le co-transporteur sodium/glucose 1. Les

lectures des 3 banques ont été alignées sur le génome nucléaire du chien (CanFam2) grâce au logiciel IGVTools. Trois mutations ont été analysées pour le gène MGAM (SNP 7 125 201, SNP 7 156 695 et SNP 7 164 842) et une pour le gène SLC5A1 (SNP 24 934 536). Pour chaque mutation d'intérêt, l'analyse globale a été complétée par des réactions de PCR ciblant les régions des gènes concernés afin de confirmer nos informations. Les fréquences des haplotypes de chiens et de loups actuels ont été comparées aux informations trouvées pour le canidé de Chauvet ; le chacal est utilisé comme groupe externe. Les données pour les chiens et loups ont été recueillies à partir de l'article d'Axelsson *et al.* (2013) et les données pour le chacal à partir de l'article de Freedman *et al.* (2014).



**Figure 48.** Fréquence des génotypes pour les 3 mutations de MGAM et celle de SCL5A1 (MGAM 1 : SNP 16:7 125 201 ; MGAM 2 : SNP 16:7 156 695 ; MGAM 3 : SNP 16:7 164 842 et SCL5A1 : SNP 26:24 934 536) pour le chien, le loup, le chacal et le canidé de Chauvet.



Pour les 4 SNPs étudiés, le canidé de Chauvet présente le même allèle, la version ancestrale, que le chacal. Dans deux cas, les chiens ont un allèle dérivé qui leur est propre et qui a dû être sélectionné durant la domestication (le SNP 7 125 201 et le SNP 7 156 695) selon l'article d'Axelsson. Il est à noter qu'une partie des loups possèdent le même allèle que les chiens pour le SNP 7 156 695, même s'il n'est pas majoritaire. Cette observation pourrait être due à des hybridations postérieures à l'apparition de la mutation chez les chiens. Pour le SNP 7 164 842, les loups présentent un taux d'allèle dérivé élevé alors que les chiens et le canidé de Chauvet portent la version ancestrale du gène. Enfin, pour le gène SCL5A1 (SNP 24 934 536), la séquence du spécimen de Chauvet présente un allèle identique à celui des loups et du chacal, ce qui pourrait être corrélé à une alimentation carnivore plutôt qu'omnivore.

Nous nous sommes également intéressés au gène AMY2B qui code l'amylase et dont un nombre élevé de copies correspondrait, pour certains auteurs, à une alimentation omnivore. Pour connaître le nombre de copies de ce gène chez notre spécimen, nous avons aligné les lectures des 3 banques sur la séquence du gène et calculé son taux de couverture. Celui-ci équivaut à 0,6 X, ce qui est identique à la couverture globale du génome nucléaire. Le canidé de Chauvet ne présente donc pas de multiples copies du gène de la synthèse de l'amylase.

Ces différentes observations permettent de penser que le spécimen de la Grotte Chauvet-Pont d'Arc était plus adapté à un régime carné qu'à une alimentation omnivore. L'analyse de son bol alimentaire lors d'analyses ultérieures nous l'a d'ailleurs confirmé.

### ***3.5.3.2. Analyse des gènes déterminant la couleur du pelage***

Après les gènes impliqués dans la digestion de l'amidon, nous avons analysé les gènes jouant un rôle dans la couleur du pelage (Candille *et al.*, 2007). Cette dernière est principalement définie par 3 gènes : MC1R qui code le récepteur de la mélanocortine de type 1, CBD103 qui code la  $\beta$ -défensine103 et ASIP qui code la protéine Agouti. Les mutations analysées sont : celle de MC1R où le codon 306 code une arginine (allèle sauvage) ou comporte un stop (allèle muté), celle de CBD103 où le codon 23 détermine une glycine (allèle sauvage) ou bien est absent ( $\Delta$ G23, muté), et celle de ASIP où pour l'acide aminé 96 une arginine (sauvage) est remplacée par une cystéine (muté). L'allèle portant la mutation  $\Delta$ G23 est dominant et celle-ci augmente l'affinité de la  $\beta$ -défensine103 pour le récepteur

MC1R ce qui entraîne la production d'eumélanine et donc un pelage noir. L'allèle muté pour ASIP est responsable aussi d'un pelage noir mais cette mutation est récessive. La mutation de l'allèle de MC1R n'apporte aucun changement de couleur de pelage.

Dans un premier temps, les lectures des 3 banques ont été alignées sur des séquences chimères, où les allèles mutés et non mutés des trois gènes ont été fusionnés. Pour les mutations respectives  $\Delta G23$  et AAC96R des gènes CBD103 et ASIP, une lecture unique a été trouvée en faveur des versions sauvages dans les deux cas. Aucune lecture ne s'est alignée pour la mutation du gène MC1R, que ce soit sur l'allèle sauvage ou l'allèle muté.

Afin de compléter ces données, des amorces spécifiques, encadrant les mutations d'intérêt de ces 3 gènes, ont donc été sélectionnées et des amplifications d'ADN ont été réalisées. Dans les trois cas les amplifications ont été réalisées avec succès, ce qui est remarquable pour de l'ADN ancien compte tenu qu'il s'agit de gènes présents à raison de deux copies seulement par cellule, et confirme la bonne préservation de l'ADN dans le coprolithe. Les PCR ont permis de confirmer les génotypes vus par alignement de lectures ; ce sont bien les versions sauvages des allèles. La mutation  $\Delta G23$  du gène CBD103 étant dominante, nous avons réalisé une centaine de réactions PCR – avec à chaque fois, le séquençage de 8 clones minimum - afin de nous assurer que nous avons bien uniquement des allèles de version sauvage, et non pas une version mutée et une version sauvage. Les réactions de PCR ont également permis d'obtenir des informations pour l'allèle du gène MC1R : c'est la version sauvage que l'on trouve chez le canidé de Chauvet (Figure 49).

	Chauvet	Allèle dérivé	Allèle sauvage
MC1R	AA306R	codon stop	AA306R
CBD103	G23	$\Delta G23$	G23
ASIP	AA96R	AA96C	AA96R

**Figure 49.** Récapitulatif des haplotypes pour les mutations touchant l'acide aminé 306 du gène MC1R, le triplet n°23 du gène CBD103 et l'acide aminé 96 du gène ASIP.

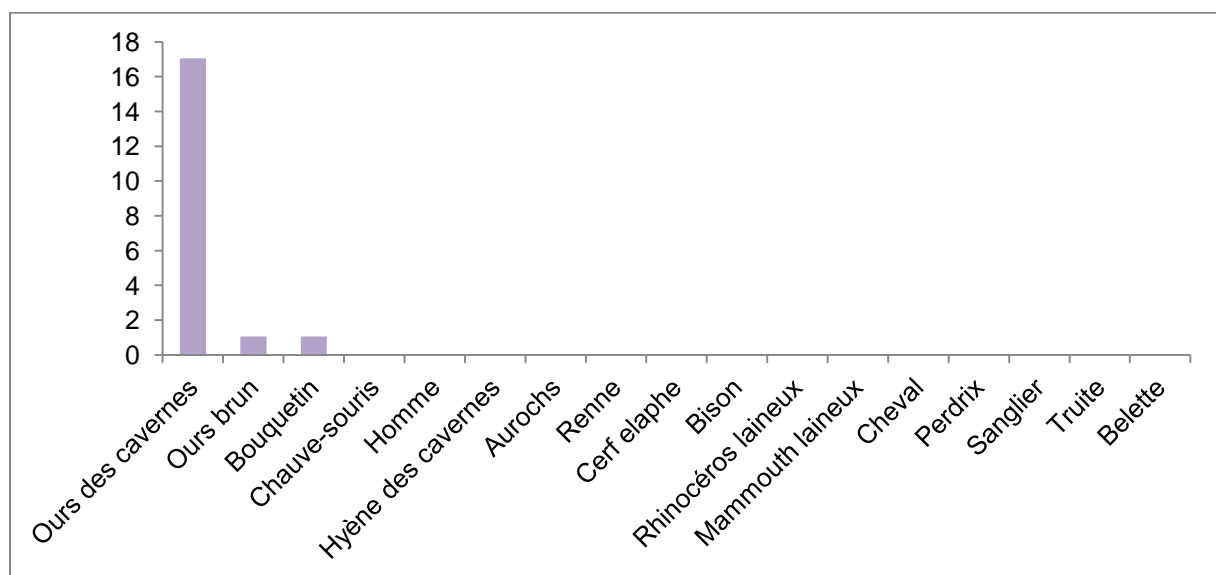
Le fait que le canidé de Chauvet possède les versions sauvages des allèles pour ces 3 gènes - MC1R, ASIP et CBD103 - permet de conclure que son pelage était de couleur claire, probablement grise. Cette observation est en accord avec une étude montrant que la mutation

pour le pelage noir est apparue ultérieurement (Ollivier *et al.*, 2013), tout d'abord chez les chiens domestiques avant de se transmettre au loup par hybridation.

### 3.6. Analyse de l'alimentation du canidé de la Grotte Chauvet-Pont d'Arc

Dans le but de caractériser l'alimentation du canidé de la grotte Chauvet, deux techniques différentes ont été mises en place : une approche globale grâce aux lectures des banques et une approche ciblée par amplification d'ADN.

En premier lieu, les 110 millions de lectures de la banque 3 ont été alignées sur une cinquantaine de génomes mitochondriaux de proies potentielles - dont plusieurs espèces éteintes - avec un crible tolérant jusqu'à 1 mésappariement. La banque 3 a été choisie pour ce cible car elle a été construite à partir de l'ADN du cœur du coprolithe ; les possibilités de trouver de l'ADN exogène sont donc diminuées. Les séquences appartenant au canidé (dont l'ADN est majoritaire dans le coprolithe) ont été retirées, ainsi que les séquences présentant le même score sur plusieurs espèces et donc jugées non discriminantes. La figure 50 présente les résultats obtenus pour 12 espèces représentatives des grands groupes animaux criblés.



**Figure 50.** Nombre de séquences de la banque présentant un alignement tolérant un mésappariement sur un échantillonnage de génomes mitochondriaux de proies potentielles. Les séquences présentant un alignement parfait ou équivalent avec le canidé de Chauvet ont été retirées. Les lectures analysées (110 millions de fragments de taille comprise entre 20 et 101 nucléotides) proviennent d'une ligne de séquences Illumina de la banque 3, produite à partir de l'ADN extrait du cœur du coprolithe.

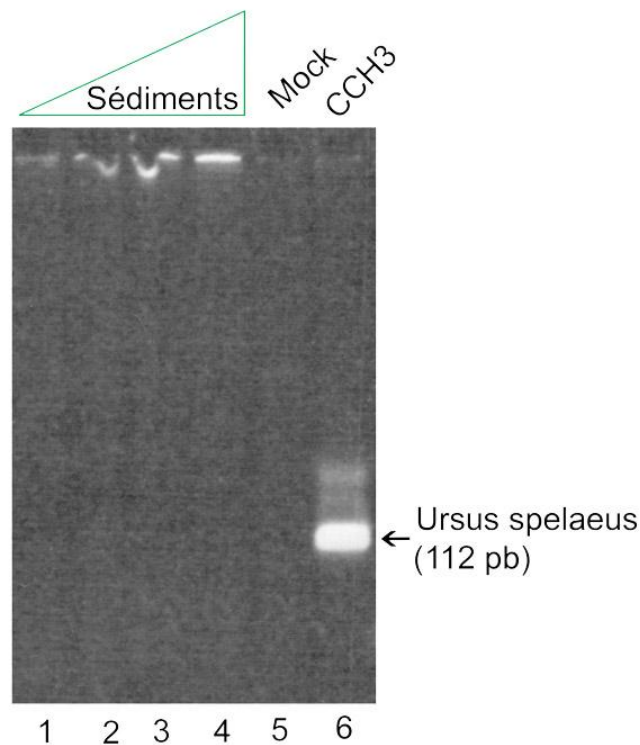
L'espèce pour laquelle nous obtenons le plus grand nombre de séquences est l'ours des cavernes avec 17 lectures. L'absence de lectures appartenant à l'Homme prouve, une fois de plus, qu'il n'y a pas eu contamination lors de la manipulation de l'échantillon et de la construction des banques de fragments. Le tableau complet comportant les résultats pour les 50 espèces analysées se trouve dans les annexes (Tableau 11).

**Tableau 11.** Nombre de lectures, entre 20 et 101 nucléotides, de la banque 3 qui présente un mésappariement sur les génomes mitochondriaux indiqués. Les lectures présentant un alignement parfait ou équivalent avec le canidé de Chauvet ont été retirées.

Nom	Nom commun	GenBank	Nombre de lectures
<i>Ursus spelaeus</i>	Ours des cavernes	EU327344	17
<i>Capra ibex</i>	Bouquetin des Alpes	FJ207526	1
<i>Rhinolophus ferrumequinum</i>	Grand rhinolophe	JX084273	1
<i>Homo sapiens</i>	Homme	JX893364	0
<i>Bos primigenius</i>	Aurochs	GU985279	0
<i>Crocuta crocuta</i>	Hyène des cavernes	JF894378	0
<i>Cervus elaphus</i>	Cerf	AB245427	0
<i>Bison bonasus</i>	Bison d'Europe	HM045017	0
<i>Mammuthus primigenius</i>	Mammouth laineux	DQ316067	0
<i>Equus przewalskii</i>	Cheval de Przewalskii	JN398403	0
<i>Coelodonta antiquitatis</i>	Rhinocéros laineux	FJ905813	0
<i>Sus crofa</i>	Cochon domestique	AP003428	0
<i>Lepus europaeus</i>	Lièvre d'Europe	AJ421471	0
<i>Salmo trutta trutta</i>	Truite	AM910409	0
<i>Alectoris chukar</i>	Perdrix choukar	FJ752426	0
<i>Alces alces</i>	Élan eurasiatique	JN632595	0
<i>Anas platyrhynchos</i>	Canard colvert	EU009397	0
<i>Antilope cervicapra</i>	Antilope	AP003422	0
<i>Apodemus agrarius</i>	Mulot rayé	JN629047	0
<i>Bison bison</i>	Bison d'Amérique	GU946990	0
<i>Buteo buteo</i>	Buse	AF380305	0
<i>Capra pyrenaica</i>	Bouquetin d'Espagne	FJ207528	1
<i>Castor fiber</i>	Castor d'Europe	FR691686	0
<i>Cepaea nemoralis</i>	Escargot des bois	U23045	0
<i>Clupea harengus</i>	Hareng	AP009133	0
<i>Coptotermes formosanus</i>	Coptotermite de Formose	AB626145	0
<i>Elaphe poryphyracea</i>	Couleuvre obscure à bandes	GQ181130	0
<i>Equus hemionus</i>	Hémione	HM118851	0
<i>Equus ovodovi</i>	Sussemione	JX312734	0
<i>Gallus gallus</i>	Poule	X52392	0
<i>Gastrimargus marmoratus</i>	Sauterelle	EU513373	0
<i>Gulo gulo</i>	Glouton	AM711901	0
<i>Lutra lutra</i>	Loutre eurasiatique	EF672696	0
<i>Martes pennanti</i>	Pékan	HQ705179	0
<i>Martes zibellina</i>	Zibeline	HM106323	0

<i>Meles meles anakuma</i>	Blaireau japonais	AB291075	0
<i>Mustela nivalis</i>	Belette d'Europe	HM106319	0
<i>Mustela putorius</i>	Putois d'Europe	HM106318	0
<i>Myoxus glis</i>	Loir gris	AJ001562	0
<i>Neovison vison</i>	Vison d'Amérique	HM106322	0
<i>Ovibos moschatus</i>	Bœuf musqué	FJ207536	0
<i>Perdix dauurica</i>	Perdrix de Daourie	FJ752431	0
<i>Procyon lotor</i>	Raton laveur	AB297804	0
<i>Rangifer tarandus</i>	Renne	AB245426	0
<i>Rousettus aegyptiacus</i>	Roussette égyptienne	AB205183	0
<i>Rupicapra rupicapra</i>	Chamois	FJ207539	0
<i>Salmo salar</i>	Saumon d'Atlantique	U12143	0
<i>Sciurus vulgaris</i>	Ecureuil roux européen	AJ238588	0
<i>Talpa europeae</i>	Taupe européenne	Y19192	0
<i>Ursus arctos</i>	Ours brun	EU497665	5

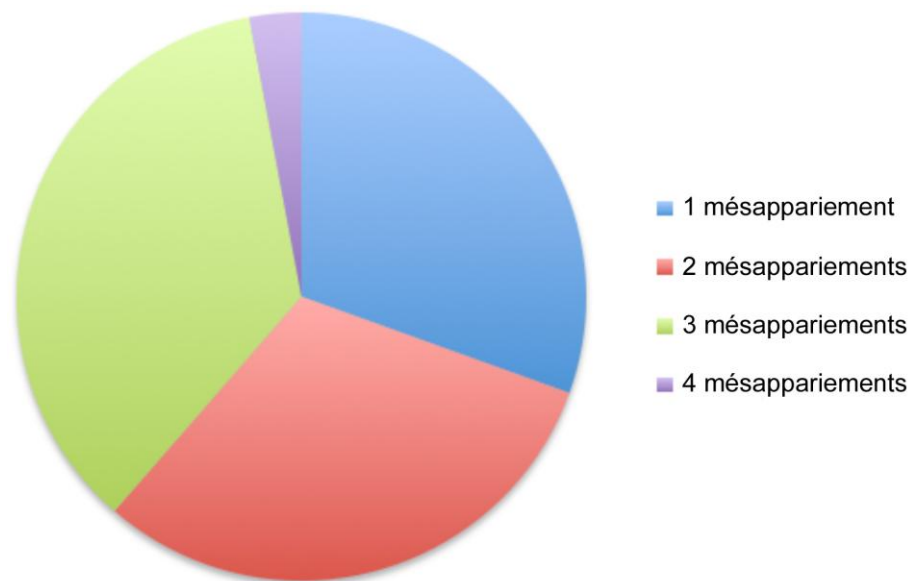
Dans un second temps, comme le nombre de séquence d'ours des cavernes reste faible malgré sa prédominance sur les autres espèces, nous avons voulu nous assurer que sa présence était réelle et non pas due à un artefact. Dans ce but, des amplifications d'un fragment de 112 paires de bases avec des amorces spécifiques du génome mitochondrial d'*Ursus spelaeus* ont donc été réalisées (Figure 51).



**Figure 51.** Amplification par PCR d'un fragment (112 paires de bases) du gène 16S du génome mitochondrial de l'ours des cavernes dans le coprolithe de canidé de Chauvet.

Le gel d'électrophorèse réalisé présente une amplification de la taille attendue du fragment pour les amorces ciblant l'ours des cavernes. Les réactions de PCR réalisées à partir de l'ADN extrait de sédiments prélevés à quelques mètres du coprolithe ne montrent aucune amplification. Le séquençage des produits de PCR par la méthode Sanger montre que l'ADN amplifié correspond bien à celui de l'ours des cavernes.

Outre la confirmation de la présence d'ADN d'ours des cavernes dans le coprolithe, les expériences de PCR apportent une série d'observations intéressantes. Nous avons analysé pour ces expériences 3 produits de PCR différents et séquencé au total 272 fragments clonés. La comparaison des séquences obtenues avec celle établie précédemment à partir d'un os d'ours des cavernes de la même grotte (Bon *et al.*, 2008), utilisée comme référence, montre qu'aucun des fragments amplifiés à partir du coprolithe n'est dépourvu de dommages. Un tiers d'entre eux présente une mutation par rapport à cette séquence de référence, et les autres en comportent 2 à 4 (Figure 52).

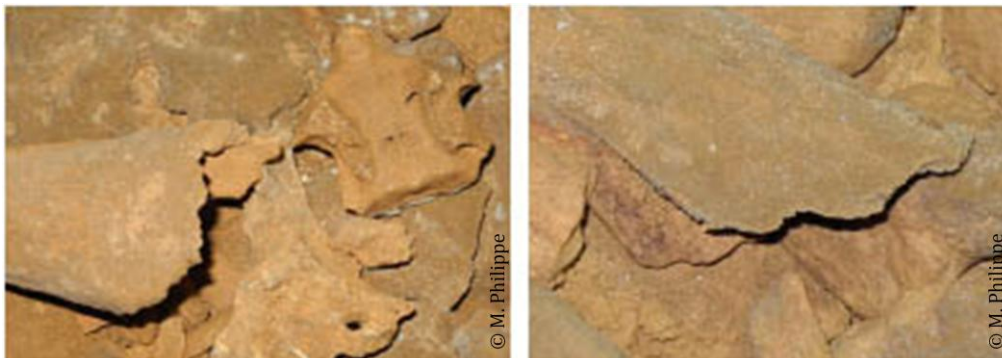


**Figure 52.** Statistique des endommagements de l'ADN pour les fragments d'ADN d'ours des cavernes amplifiés à partir du coprolithe. Les statistiques portent sur 273 fragments clonés issus de 3 PCR différentes. La séquence des fragments, correspondant à une portion du gène 16S, est comparée à celle établie pour un spécimen d'ours des cavernes de la même grotte (Bon *et al.*, 2008).

Ces résultats pourraient expliquer le faible nombre de fragments d'ADN d'ours des cavernes mis en évidence par alignement des lectures Illumina sur le génome mitochondrial, ceci pour deux raisons : premièrement, un seul mésappariement a été autorisé pour cet alignement, et les résultats de PCR, et séquençage Sanger montrent que la majorité (69%) des fragments

comporte au moins deux mésappariements avec la séquence utilisée pour l'alignement ; deuxièmement, les réactions de PCR sur le gène 16S ont été réalisées avec de la *Taq* polymérase, alors que la banque est amplifiée préalablement au séquençage au moyen de la Phusion DNA polymérase, qui ne peut utiliser comme matrice les brins d'ADN comportant une cytosine désaminée, endommagement le plus fréquent pour l'ADN ancien (Briggs *et al.*, 2007). Ainsi, la majorité des fragments d'ADN d'ours des cavernes du coprolithe ne sont probablement pas présents dans la banque amplifiée et n'auraient pu être détectés, même en tolérant des critères d'alignement plus lâches des lectures sur le génome de référence.

La présence d'ADN d'*Ursus spelaeus* dans le coprolithe peut être mise en relation avec une observation réalisée par les paléontologues de l'équipe Chauvet (M. Philippe et P. Fosse). En effet, des os d'ours des cavernes dont les extrémités sont rongées sont présents dans plusieurs salles de la grotte Chauvet-Pont d'Arc, y compris les secteurs profonds (Figure 53).



**Figure 53.** Os d'ours des cavernes de la Grotte Chauvet-Pont d'Arc dont l'épiphyse est rongée.

La présence d'os dont les épiphyses sont rongées est plus en faveur de l'activité d'un canidé que de l'hyène des cavernes, qui broie plutôt la diaphyse des os. Il est possible que le canidé soit entré dans la grotte pour consommer des ours des cavernes morts pendant leur hibernation.

## 4. CONCLUSION

L'analyse de l'ADN contenu dans trois coprolithes prélevés dans les secteurs profonds de la Grotte Chauvet-Pont d'Arc, au moyen de la méthode de séquençage haut-débit, mais aussi d'amplifications ciblées, a permis de mettre en lumière de nombreuses informations.

Il a été possible de voir que la conservation de l'ADN dans des échantillons n'est pas toujours identique ; en effet, un seul des trois coprolithes - celui collecté dans la Salle Hillaire - présentait une quantité importante d'ADN de mammifère.

La construction de 3 banques de fragments pour cet échantillon nous a permis de reconstituer un génome mitochondrial complet avec un fort taux de couverture (89 X) et de réaliser des arbres phylogénétiques. Ces derniers ont révélé que le canidé de la Grotte Chauvet-Pont d'Arc n'appartenait à aucun des haplogroupes mitochondriaux définis par les chiens et loups actuels. De plus, il se regroupe avec un autre génome mitochondrial attribué à un loup de Belgique et reconstitué à partir d'un ossement daté à 26 000 ans (Thalmann *et al.*, 2013). Cette observation est cohérente avec l'origine de notre échantillon - l'Europe de l'Ouest - et avec l'âge obtenu par la datation radiocarbone de celui-ci (34 500 cal BP). Le canidé de la Grotte Chauvet-Pont d'Arc appartenait à une lignée maternelle présente en Europe au Pléistocène et aujourd'hui éteinte.

L'analyse globale du génome nucléaire de notre spécimen a permis d'obtenir un génome partiel avec un taux de couverture de 0,65 X, ce qui a montré que le canidé était une femelle. L'analyse a également montré que les catégories de gènes qui présentaient le plus de différences entre le spécimen et le chien actuel étaient celles impliquées dans le développement des neurones et de structures anatomiques. Ces différences au niveau des gènes déterminant le développement des neurones avaient pu être observées dans un article discutant des mutations survenues lors de la domestication (Axelsson *et al.*, 2013).

L'analyse ciblée de mutations - présentes dans des gènes nucléaires impliqués dans le métabolisme de l'amidon - a montré que le canidé ne présentait pas les allèles propres aux chiens et qui ont été sélectionnés lors de la domestication (Axelsson *et al.*, 2013). De plus, l'allèle dérivé présent chez les loups actuels est absent chez notre spécimen ; ce dernier porte les mêmes allèles que le chacal pour les 4 mutations étudiées. Ces observations, combinées au fait que le gène *AMY2B* ne soit pas présent en plusieurs copies, permettent de penser que le



canidé de Chauvet était plus adapté à une alimentation carnée qu'à un régime alimentaire omnivore. Cette conclusion n'empêche pas l'hypothèse que l'animal ait pu avoir des contacts avec les Hommes du Paléolithique Supérieur. En effet, la culture du blé, apparue lors de la sédentarisation, est postérieure à notre échantillon qui est daté de 34 500 ans.

L'analyse de mutations présentes dans les gènes nucléaires déterminant la couleur du pelage a permis de voir que le canidé présentait les allèles sauvages pour ces gènes. Il a été possible de déduire que la louve de Chauvet avait un pelage de couleur claire, grise ou blanche selon l'âge de l'animal. Cette observation est en accord avec une étude qui propose que certaines mutations - donnant un pelage noir - seraient apparues dans un premier temps chez les chiens domestiques avant de se transmettre aux loups par hybridation (Ollivier *et al.*, 2013).

Lors de l'étude de l'alimentation du canidé de la Grotte Chauvet, nous avons pu montrer, par deux approches différentes, que ce dernier avait consommé de l'ours des cavernes. Cette observation est en accord avec le fait que l'animal était adapté à une alimentation carnée ; elle propose également une explication à l'intrusion du canidé dans les secteurs profonds de la grotte. Sa venue pourrait être liée à la consommation d'ours des cavernes en hibernation ou de carcasses d'individus morts durant celle-ci. L'observation - dans la caverne - par les paléontologues d'os rongés d'*Ursus spelaeus*, probablement par un canidé, supporte le fait que les loups aient pu venir dans la grotte afin de consommer de l'ours des cavernes. La dissection du coprolithe a également permis de trouver des os intacts de très petite taille, attribuables à un petit rongeur selon les paléontologues. L'analyse par séquençage Illumina de l'ADN du coprolithe n'a toutefois pas permis de retrouver d'ADN de rongeur. Nous émettons l'hypothèse que l'endommagement de l'ADN des proies, mise en évidence par nos études par réactions de PCR de l'ADN d'ours des cavernes, a empêché la détection de l'ADN de petits animaux dans le coprolithe. Notre étude donc montre à la fois les potentialités (mise en évidence de la consommation d'ours des cavernes) et les limites du séquençage métagénomique.

Dans le but de continuer les recherches sur le spécimen de la Grotte Chauvet-Pont d'Arc, il serait intéressant d'utiliser les données nucléaires pour définir sa proximité avec les loups et chiens actuels. Une tentative, réalisée en collaboration avec le centre de GeoGenetics de Copenhague, suggère que le spécimen de Chauvet est plus proche des chiens que des loups

actuels, et qu'il donne donc des informations sur les lignées à partir desquelles les chiens ont été domestiqués. Cependant, la faible couverture du génome incite à la prudence. En utilisant une partie du set de données pour différents loups actuels (loup d'Israël, de Croatie ou de Russie) pour réaliser une analyse sur un génome avec une couverture de 0,7 X, la topologie de l'arbre change et ces loups occupent alors une position phylogénétique équivalente à celle du spécimen de Chauvet (Thomas Gilbert, communication personnelle). Il apparaît donc nécessaire d'obtenir davantage de séquences pour obtenir une information solide. Cette considération méthodologique s'applique également, à notre sens, à l'étude récemment publiée par Skoglund *et al.* (2015), portant sur un génome nucléaire obtenu avec une couverture de 1X pour un spécimen de loup du nord de la Sibérie, daté de 35 000 ans. Cette étude suggère que la divergence entre les loups et chiens actuels pourrait être antérieure au dernier maximum glaciaire. Un effort de séquençage supplémentaire pour le spécimen de la Grotte Chauvet, possible compte tenu de la qualité de l'échantillon, apparaît souhaitable pour vérifier si l'hypothèse proposée par l'équipe de Skoglund s'applique également aux canidés d'Europe occidentale (Skoglund *et al.*, 2015).

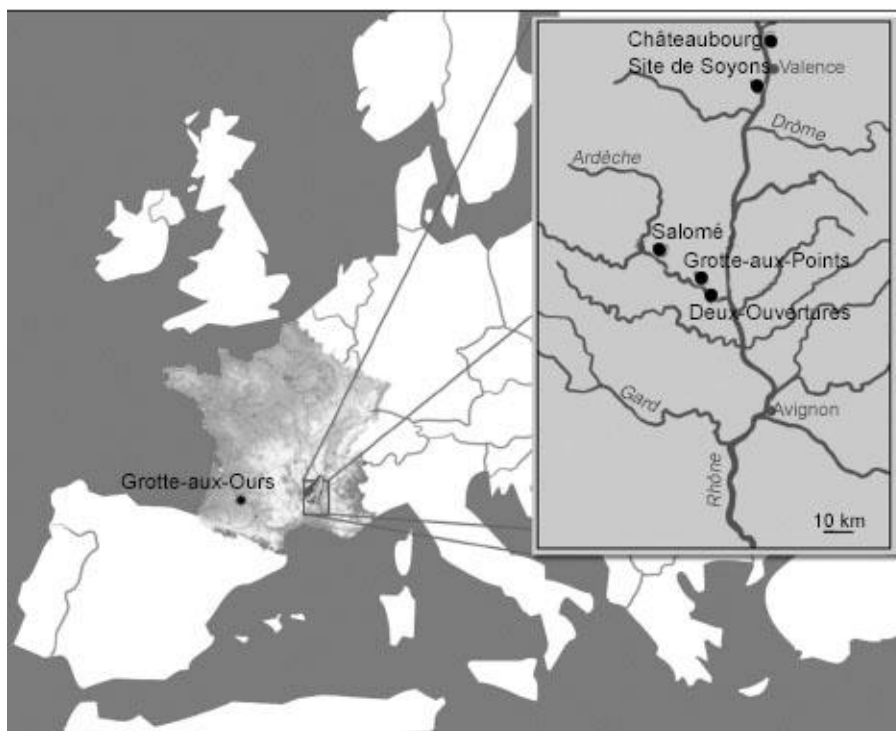


*Partie 4 :*

*Analyses de  
coprolithes d'hyènes des cavernes*

## 1. INTRODUCTION

L'hyène des cavernes, très répandue en Europe au Pléistocène, était un compétiteur de l'Homme pour les ressources carnées (Discamps, 2011), et de nombreux sites présentent des traces de son passage (ossements d'hyène et de proies chassées ou charognées, coprolithes). L'hyène étant un carnivore broyeur d'os, ses excréments ont pu se fossiliser et sont retrouvés - parfois en nombre important - dans de nombreux sites archéologiques. Dans le but d'établir un panorama des proies de l'hyène des cavernes au Pléistocène, l'ADN de 8 coprolithes d'hyène (Ardèche : 7 échantillons ; Lot : 1 échantillon) a été étudié par la méthode de séquençage à haut-débit Illumina (Figure 54).



**Figure 54.** Localisation des sites d'où proviennent les 8 coprolithes étudiés par séquençage Illumina. Le site de Soyons regroupe les grottes dites du Trou-du-Renard, de Néron, et un échantillon du Musée de Soyons.

Dans ce chapitre, nous verrons que l'analyse de l'ADN de ces échantillons a permis de mettre en évidence une alimentation variée pour l'hyène des cavernes, basée sur la consommation d'animaux de grande et de petite taille. L'observation la plus originale était toutefois la présence, dans un coprolithe de la Grotte-aux-Ours (Lot), d'ADN d'un boviné différent de l'aurochs et non représenté dans les bases de données. Cette observation nous a semblé suffisamment intéressante pour focaliser nos analyses sur les bovinés.

Outre l'aurochs, plusieurs espèces de bovinés étaient présentes en Europe de l'Ouest durant le Pléistocène supérieur.

Le bison des steppes (*Bison priscus*) était une espèce abondante qui vivait dans une vaste zone géographique s'étendant de l'Angleterre à l'Amérique du Nord et de l'Archipel de Nova Zembla dans l'océan Arctique jusqu'au sud de l'Espagne (Lorenzen *et al.*, 2011). Cette espèce était très commune au Pléistocène comme l'attestent de nombreux restes retrouvés dans des gisements anglais, italiens, palestiniens ou encore sibériens (Kurtén, 1968). En France, l'espèce est mise en évidence grâce à des spécimens exhumés dans un grand nombre de sites tels que celui de Lunel-Viel (Kurtén, 1968), ou la grotte des Trois-frères (Bégouën, 2014). Des ossements de bison des steppes sont également retrouvés dans les sites allemands de Süssenborn, Mosbach, Hundsheim ou encore Heppenloch. Il est communément admis que c'est cette espèce qui a été représentée par les artistes du Pléistocène dans les grottes ornées telles que la Grotte d'Altamira dans le Nord de l'Espagne, la Grotte Chauvet-Pont d'Arc, la Grotte de Lascaux, de Font-de-Gaume ou encore celle des Trois-Frères en France.

Le bison des steppes était plus grand et plus massif que les bisons actuels. Sa hauteur au garrot était de près de 2m, et sa longueur voisine de 2,70m. Ses larges cornes, dont la distance de pointe à pointe pouvait atteindre 1,20m, étaient encombrantes dans les espaces couverts ; c'est sans doute pour cette raison qu'il vivait dans de grands espaces ouverts, à l'image des bisons d'Amérique actuels. L'espèce *Bison priscus* s'est éteinte à la fin de la dernière ère glaciaire, il y a environ 10 000 ans (Kurtén, 1968).

Le bison des forêts (*Bison schoetensacki*), dont les restes sont beaucoup moins fréquents que ceux du bison des steppes, était présent en Europe de l'ouest durant le Pléistocène moyen et supérieur. L'espèce s'est éteinte durant le Pléistocène supérieur. L'espèce *Bison schoetensacki* a été caractérisée à partir d'ossements retrouvés en Allemagne sur les sites, appartenant au Pléistocène moyen, de Mosbach et de Mauer. Des restes ont également été retrouvés dans d'autres sites allemands, ainsi qu'en Grande-Bretagne (Forest-Bed), en Moldavie (Tiraspol) et en Russie (Sinaja Balka, péninsule de Taman). En France, des restes de *Bison schoetensacki* attribués au Pléistocène moyen ancien ont été retrouvés à Dufort (Gard), à Saint-Prest (Eure et Loir), à la Nautérie (Gers), à Solihac et à Sainzelles (Haute-Loire). Des ossements ont également été retrouvés dans des sites du Pléistocène moyen tel que Châtillon-Saint-Jean dans la Drôme (Mourer-Chauviré, 1972). Pour le Pléistocène supérieur, l'espèce *Bison schoetensacki* est présente sur le site de Siréjol dans le

Lot (Guérin & Philippe, 1971 ; Joannin, 2002), à Chasse-sur-Rhône et à Poleymieux dans le Rhône, ainsi que dans la région de la Vienne en Isère et dans la grotte de Lherm en Ariège. Le bison des forêts, comme son nom l'indique, était une forme adaptée au biotope forestier et souvent considérée à tort comme de petite taille (Sher, 1997). En effet, si sa taille est bien inférieure à celle du grand bison des steppes *Bison priscus priscus* qui peuplait l'Europe à la fin du Pléistocène moyen et au début du Pléistocène supérieur, elle est en revanche proche de celle de *Bison priscus mediator* qui vivait au Pléistocène supérieur (Joannin, 2002). Cependant, les os longs et les métapodes de *Bison schoetensacki* présentent des proportions bien différentes et sont plus graciles que ceux du bison des steppes. De plus, les spécimens de bison des forêts se caractérisent par des cornes courtes, la distance des chevilles pointes à pointes ne dépassant pas 0,87 m. Dans leur premier tiers, ces cornes sont faiblement tournées vers l'avant puis partent vers le haut et enfin dans leur tiers distal, sont faiblement inclinées vers l'arrière.

La phylogénie des bovidés est complexe et difficile à appréhender. En effet, le bison actuel d'Europe (*Bison bonasus*) a un génome mitochondrial proche de celui des spécimens du genre *Bos* alors que son génome nucléaire est plus proche de celui du bison actuel d'Amérique (*Bison bison*). Cette observation pourrait être due soit à une divergence très ancienne des lignées de bisons d'Amérique et d'Europe, soit à une introgression d'ADN nucléaire de bisons mâles dans une ancienne population eurasiennne de *Bos* (Verkaaret *al.*, 2004, Nowak & Olech, 2008). En parallèle, pour certains auteurs, *Bison priscus* serait l'ancêtre du bisons d'Amérique et *Bison schoetensacki*, celui du bison d'Europe (Nowak & Olech, 2008). Cependant, l'absence totale d'information génétique sur *Bison schoetensacki* et des données se limitant à la région de contrôle du génome mitochondrial pour *Bison priscus* (Shapiro *et al.*, 2004) ne permettent pas d'établir la position phylogénétique précise de ces espèces éteintes.

Dans notre étude, nous comparons la séquence du génome mitochondrial de boviné reconstruite à partir d'un coprolithe d'hyène de la Grotte-aux-Ours (Lot) à celles que nous avons pu obtenir à partir d'ossements de *Bison priscus* et de *Bison schoetensacki*. Le spécimen de *Bison priscus*, qui provient de la grotte des Trois-Frères dans les Pyrénées, a été analysé dans le but de reconstruire un premier génome mitochondrial complet pour cette espèce. En effet, malgré de nombreux restes présents dans les sites archéologiques, seulement une partie de la région de contrôle du génome mitochondrial avait été séquencée

préalablement (Shapiro *et al.*, 2004; Zazula *et al.*, 2009). Puis des ossements de l'espèce *Bison schoetensacki* - provenant de la Grotte de Siréjol dans le Lot - ont été étudiés afin d'obtenir des informations génétiques sur la séquence mitochondriale du bison des forêts.

## 2. MATÉRIEL ET MÉTHODES

### 2.1. Séquençage Illumina de l'ADN extrait de coprolithes d'hyène des cavernes

Après sélection des échantillons et extraction de l'ADN qu'ils contiennent, des banques de fragments ont été construites comme décrit précédemment (voir 2.7., Méthodes). Les banques de fragments sont ensuite envoyées au Génoscope d'Évry afin d'y être séquencées au moyen de la méthode à haut-débit Illumina (Tableau 12). Le séquençage a été réalisé sur la plateforme Illumina GaIIX (coprolithe de la Grotte des Deux-Ouverture) ou HiSeq2000 (autres coprolithes). Ceci explique le faible nombre de lectures obtenues pour le coprolithe de la Grotte des Deux-Ouvertures comparativement aux autres échantillons.

**Tableau 12.** Nombre de lectures (supérieures ou égales à 20 nucléotides) dans les banques obtenues grâce au séquençage haut débit par la méthode Illumina. \* 4 pistes ont été séquencées pour cet échantillon.

Site	Nombre de lectures de la banque
Néron	130817614
Trou du Renard	121682729
Salomé	128289143
Châteaubourg	73656565
Soyons	118356461
Grotte-aux-Ours *	601509879
Grotte-aux-Points	161677014
Deux-ouvertures	27803334

### 2.2. Reconstitution des génomes mitochondriaux d'hyènes des cavernes

Dans le but de reconstituer les génomes mitochondriaux d'hyène des cavernes pour chaque échantillon, les doublons des différentes banques sont éliminés puis les lectures ont été alignées sur le génome de référence de l'hyène des cavernes (NC\_020670) en tolérant 0



*indel* et 0,1 mésappariement. Ce dernier paramétrage permet de tolérer un nombre variable de mésappariement en fonction de la longueur des lectures, de 1 mésappariement pour les fragments de taille comprise entre 20 et 26 nucléotides, à 4 mésappariements pour ceux de 88 à 101 nucléotides. Le nombre de bases couvertes ainsi que le taux de couverture des génomes mitochondriaux reconstitués sont compilés dans le tableau ci-dessous (Tableau 13).

**Tableau 13.** Nombre de bases des séquences mitochondriales couvertes et taux de couverture du génome obtenu pour l'hyène des cavernes pour chaque site indiqué.

<b>Échantillon</b>	<b>Nombre de bases couvertes du génome mitochondrial</b>	<b>Couverture du génome mitochondrial</b>
Néron	17 016	130 X
Trou du Renard	17 138	163 X
Salomé	17 138	295 X
Châteaubourg	17 138	100 X
Soyons	17 136	40 X
Grotte-aux-Ours	16 960	30 X
Grotte-aux-Points	15 709	5 X
Deux-Ouvertures	17 132	13 X

## 2.3. Reconstitution des génomes mitochondriaux de bovinés

### 2.3.1. Bison des steppes

La stratégie de reconstitution est la même que celle décrite précédemment (voir 2.8., Méthodes). Le génome utilisé pour aligner les lectures de la banque est la séquence mitochondriale de référence de *Bison bison* (NC\_012346). Les lectures couvrent la majorité du génome mitochondrial mais il demeure des positions non couvertes. Des couples d'amorces spécifiques ont donc été sélectionnés à partir de la séquence provisoirement déduite des lectures afin d'obtenir par PCR des informations sur les positions non couvertes (Tableau 14).

**Tableau 14.** Liste des amorces utilisées pour amplifier l'ADN de *Bison priscus* de l'échantillon de la grotte des Trois-Frères afin de combler les positions non couvertes par les lectures Illumina. La taille des fragments amplifiés par PCR est exprimée en paire de bases. Les couples d'amorces 1 à 10 ont été utilisés pour caractériser l'échantillon. Les couples d'amorces 11 à 38 ont, eux, été utilisés pour combler les positions non couvertes.

N°	Position	Amorce sens	Amorce anti-sens	Taille
1	15 259F 15 362R	CCTCGGAGACCCAGATAACT	TTGATCGTAAGATTGCGTATG	104
2	10 052F 10 175R	GAGCCTCACAAACAACAACTT	GTACCATATTTCGGTTCATTCTAGTC	124
3	14 711F 14 812R	ACCCATATCTGCCGAGAC	TCGTCCTACGTGCATATATAA	102
4	14 768F 14 867R	GGAGCTTCAATGTTTTTATCTG	GGATTACTCCAATATTTTCATGTTT	100
5	14 818F 14 919R	ATATTACGGGTCTTACACTTTTC	CCTCATGGTAGGACGTATCC	102
6	14 936F 15 053R	GGAGCAACAGTCATCACCAA	AAGCGAAGAATCGGGTAAGG	118
7	15 357F 15 438R	GATCAATCCCCAACAAAC	GTGTGTAGTAGGGGAATTAGA	82
8	15 883F 16 012R	AACCAAATATTACAAACACCACTA	GGGGCATATAATTTAATGTACT	130
9	16 010F 16 148R	CCCATGCATATAAGCAAGT	TGGTAATTAAGCTCGTGATCTA	139
10	16 162F 16 266R	CAGCAACCCGCTAGGCA	GCCCTGAAGAAAAGAACCAGATGT	105
11	156F 254R	CATAATGGTAGGCATGCACA	AGTGGTAAATATTTAAGGGGGAA	99
12	156F 270R	CATAATGGTAGGCATGCACA	AAAATCTGTTAAAAGCAGTGGT	115
13	304F 399R	CAATACTCAAATTAGCACTCCA	ATTTTCAGTGCCTTGCTT	96
14	1394F 1483R	AGCTTAAACAAAGCATCCAGT	AGTTAAGAGGGTATTTTGGGCTA	90
15	1457F 1559R	AATCTAGCCCAAAATACCCTC	CGCCATATTCTAGATTTCTATCTCC	103
16	1900F 1998R	GGTACAGCCTTTTAGAAACGGAT	TTAACGCTTCTCAATTGGTGG	99
17	2013F 2132R	AATTAACAGATCCCAATAACA	CTTATGCAAGGAGAAAAGTT	120
18	3060F 3177R	AACTTTTATACCCAGAGATTCA	TTTCGTTCTACTAATGTGAGG	118
19	4024F 4134R	ATCCTTACATCAGGCATCCCA	CCTATAGTTCTAGAAATAAGAGGGTT	111
20	4037F 4156R	GCATCCCACCACAAACAT	ATTCTTAGGAGTAGGTTCAATTC	120
21	4251F 4359R	CCCTTCCCGTACTAATAAATCCAA	CCGATTACAGACAAGTAGTCAGT	109
22	4439F 4539R	TATTTTCTAACCCAATCAACAGC	GGGTAAATAATTTTATTACGGTTCA	101
23	5179F 5258R	CCCCTCTACAAACAACAT	ATACGGCTATTGTTGGTAGG	80
24	5493F 5605R	ACCCCAATTAATCAGGCTTC	AGTTTACACGTTGAATTGCAAA	113
25	6431F 6513R	CCTGGATTTGGAATAATCTCCC	GCTCAGACTATTCCTATATATCCG	83
26	7264F 7357R	TTCAAGCCAACATCATAACCTC	TTTCACTTGTAACCTTAACCTTGACA	94
27	8219F 8301R	CACAACCTTCTACTACAATCCAG	ATTTTCGTTTCAATTTGTTTCTCA	83
28	8290F 8403R	ATGAACGAAAATTTATTTGCCTCT	CACTAATCGATTTGATGTTGGGA	114
29	11068F 11173R	ATACTGAGTACAACCTGTGCAT	TCAAAGGTGGAGGCCGTAT	106
30	11838F 11939R	TATCGCTGCACATCTTACCT	CACAATCTAATGTTTTTGTAAACT	102
31	11851F 11960R	CTTACCTTTACTACTCTTATCCCT	ATGAGTTTCTATTGTTAGATTCAC	110
32	12077F 12178R	CGAAAAATTGGTGCAACTCC	AGGCTTGTTATTATAATGGGTAC	102
33	12320F 12393R	CCAAACCCTTAAACTATCCCTC	CGAATAGTGCTACTGGGACA	74
34	12876F 12977R	ATCTTCCTATTAATCCGCTTT	TATCGCTGTAATAACGTGGT	102
35	14087F 14173R	TCAAACACAATCTCTACCTCC	TTAGGGGCATTTATTACTGG	87
36	14 353F 14457R	CCCATAAATAGGTGAAGGC	AGATTCCATGTAAGAATAATGA	105
37	15547F 15639R	CCCATATATCATCATCGGACA	AGTTTGTTTCAATTTGTGCC	93
38	15922F 16016R	CCCCAAAATGCATTACCCAA	GCATGGGGCATATAATTTAATGT	95

Les réactions de PCR sont réalisées dans un volume de 50 µl, de la même manière que décrit précédemment, et le protocole de PCR consiste toujours en une étape d'activation (95°C, 8'45''), suivie de 45 cycles (95°C, 15'' ; 49-60°C, 20'' ; 70°C, 1').

Le couple d'amorce 1 a été utilisé pour caractériser les ossements de *Bison priscus*. Les couples d'amorces 2 à 10 ont, eux, permis une caractérisation plus fine de l'échantillon de bison des steppes choisi pour construire la banque de fragments. Les couples 11 à 38 ont été

dessinés afin de combler les positions non couvertes par les lectures Illumina pour le génome mitochondrial obtenu pour *Bison priscus*.

### 2.3.2. Boviné mis en évidence dans le coprolithe de la Grotte-aux-Ours

Comme pour le génome mitochondrial du bison des steppes, les lectures de la banque de fragments ont été alignées sur 4 génomes de référence de bovidés (*Bos primigenius*, *Bison bison*, *Bison bonasus* et *Bos taurus*), puis une seconde fois sur le consensus provisoirement déduit du premier alignement. Afin de combler les positions non couvertes par les lectures Illumina, des amorces spécifiques ont été choisies à partir de la séquence mitochondriale provisoire (Tableau 15). Les réactions de PCR sont réalisées comme décrit pour *Bison priscus*.

**Tableau 15.** Liste des amorces utilisées pour amplifier l'ADN de boviné du coprolithe de la Grotte-aux-Ours afin d'obtenir des informations sur les positions non couvertes du génome mitochondrial par le séquençage Illumina.

Position	Amorce sens	Amorce anti-sens	Taille (pb)
170F 268R	TGGACATTACAGTCAATGGT	AAGCAACTAGGGAAAAGTCTG	99
1420F 1539R	CCTAGAAGACTTCATTCAT	CTCCTATACTTTAAATTGGGAA	120
1 548F 1 634R	CTAAGTACGGCGCTATAGAGA	GTACAAGGGGTAATCTTTGCT	87
1 992F 2 080R	GCGTTAAAGCTCAACAACA	ATTCTATAATAGATTAGTCCAGT	89
2 154F 2 235R	ATTCTGACCACTAACAGCTA	TGCACTCCTGTGTTGGATT	82
3 064F 3 156R	TTTACATCCAGAGATTCAAATCC	GAATGCTACGGCCAATAGGA	93
3 564F 3 648R	AATAAGTGGATCCTTTACCCT	TATTGCTAGAGGTCATGCTGG	85
4 040F 4 109R	TCCCGCCACTGACATAAGAA	AGTTCTAGAAATAAGAGGGTT	90
4 250F 4 360R	ACCCTTCCCGTACTAATAAACC	TCCGATTGACACAAGTAGTCAG	111
4 831F 4 935R	AATAACAGCAGTACTACCATA	GTGGAATTGGCCATGAAT	105
5 032F 5 134R	CCTATCTGGGTTTCATACAA	TAGGTTTAGTAGAGCTGTAATTGC	103
5 201F 5 297R	ATAAAATGACAATTTCCCCTT	CTGATAGTATTGGTGTGAGT	97
6 432F 6 531R	TGGATTTGGAATAATCTCCATA	AAATCCGATTGACATTATAGCC	100
7 039F 7 130R	ACTATCTCATCAATAGGCTCAT	ACTTCTCGTTTAGATGCAAA	92
7 411F 7 485R	ATCACCAATCATAGAGGAACT	ACTAATGAGCTAATTAAGAAGACA	75
7 571F 7 687R	ATTTTGCCCGCTATTATCTT	GTATCATTGATGTCCTATGGTT	117
11 044F 11 133R	TCCCTAAACTTCCTAATACTCCA	AGCTATTATACATGCTAGTCA	90
11 682F 11 769R	CTTTTCATGATCTAACATTACAAT	AGTTATAATTAGCATGTATAGGGA	88
11 876F 11 977R	ACCCAAAAATTATTCTAGGACCTC	TCGGTAAATAAGAAGGTAATGAGTT	102
12 110F 12 202R	AACATATTCGCCTCATTCTC	GTAATTGGAAGATTTGTAGGTGT	93
12 337F 12 437R	CCTCAGCTTTAAAATAGACTACTTC	CTGAGTGTATGTATCATATTGAGA	101
13 847F 13 941R	AAAGGCCTAATCAAACATACTTC	GGTTATTATAGAAATTACTCGTGGA	95
14 067F 14 165R	ATCTCCCAAACCATTAACCTCA	GCATTTGTTACTGGCTTGTTGA	99
14 380F 14 472R	AACCCTACAAAACCTATCACA	ATCATTAGTCATGGTTAGATTCC	93
15 910F 16 019R	GCACACACCCCATACACA	GGGGCATATAATTTAATGTACT	110
16 012F 16 149R	TATGCCCCATGCGTATAAGCAA	TCAAGCTCGCGATCTAATGGA	138

## 2.4. Analyse par PCR de l'ADN mitochondrial du *Bison schoetensacki*

La caractérisation des ossements de *Bison schoetensacki* provenant de la Grotte de Siréjol a montré que ces derniers ne contenaient pas une quantité d'AN suffisante pour construire une banque de fragments. Les analyses ont donc été réalisées uniquement à l'aide de réactions de PCR. Les couples d'amorces utilisées pour amplifier l'ADN de *Bison schoetensacki* ont été déduits à partir de la séquence mitochondriale finale obtenue pour le boviné de la Grotte-aux-Ours (Tableau 16).

**Tableau 16.** Liste des amorces utilisées pour amplifier l'ADN de *Bison schoetensacki*. Les amorces ont été sélectionnées en utilisant la séquence du génome mitochondrial de bison reconstitué à partir du coprolithe de la Grotte-aux-Ours. Les caractères gras indiquent les 6 couples d'amorces également utilisés pour valider la séquence de ce génome (voir tableau précédent). La taille des fragments amplifiés est exprimée en paire de bases (pb).

Position	Amorce sens	Amorce anti-sens	Taille (pb)
1 451F 1525R	AACTAAACCTAGCCAAA	AATTGGGAATAAATGTTTTG	75
<b>1 547F 1 634R</b>	<b>CTAAGTACGGCGCTATAGAGA</b>	<b>GTACAAGGGGTAATCTTTGC</b>	<b>88</b>
<b>1 992F 2 080R</b>	<b>GCGTAAAGCTCAACAACA</b>	<b>ATTCTATAATAGATTAGTCCAGT</b>	<b>89</b>
<b>2 154F 2 235R</b>	<b>ATTCTGACCACTAACAGCTA</b>	<b>TGCACTCCTGTGTTGGATT</b>	<b>82</b>
5 083F 5162R	CAACAGCATCATTCTACC	TAGAGTATGTGAGTCGTA	80
<b>7 039F 7 130R</b>	<b>ACTATCTCATCAATAGGCTCAT</b>	<b>ACTTCTCGTTTAGATGCAA</b>	<b>92</b>
<b>7 411F 7485R</b>	<b>ATCACCAATCATAGAGGAACT</b>	<b>ACTAATGAGCTAATTAAGAAGACA</b>	<b>75</b>
9 831F 9 890R	ATACTAGCCCTCCTGACCAA	TCAGAATGCGATGATGACAAG	60
<b>11 044F 11 133R</b>	<b>TCCCTAAACTTCCTAATACTCCA</b>	<b>AGCTATTATACATGCTAGTCA</b>	<b>90</b>
15 033F 15 104R	CCTTACCCGATTTTTTCGC	AGTAGATGAACTATGGCAAT	72
15 103F 15 177R	CTATTCTCCACGAAACAGGT	GGGGTGAATGGAATTTTGTCT	75
15 356 F 15 347R	GATCAATCCCCAACAAAC	GTGTGTAGTAGGGGAATTAGA	83
16 012F 16 076R	TATGCCCCATGCGTATAA	TAGTAATTGTATGTATTATGT	65
16 090F 16 176R	CTTTATGTCAAGCTCATTCTT	GGTTGCTGGTTTCACGC	87
16 137F 16 211R	ATCGCGAGCTTGATTACC	GGCCCGAGCGAGAAGAG	75
16 169F 16 244R	CAGCAACCCGCTAGGCAA	AAGTTCATTAATAGCGACCCC	76

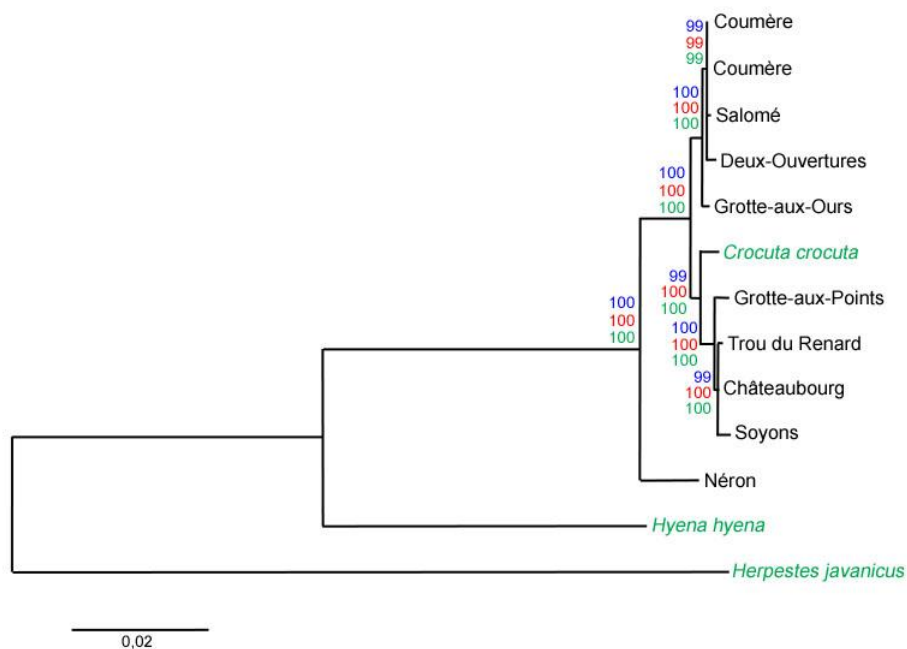
## 3. RÉSULTATS ET DISCUSSION

### 3.1. Panorama de l'alimentation et des génomes de l'hyène des cavernes

#### 3.1.1. Diversité génétique de l'hyène des cavernes au Pléistocène

Après la reconstitution des génomes mitochondriaux d'hyène des cavernes pour chaque échantillon (Tableau 13), la position phylogénétique des différents spécimens a pu être explorée. Les arbres phylogénétiques ont été construits au moyen de la méthode Neighbor Joining, Minimum Evolution, Maximum Likelihood. Pour ne pas biaiser l'analyse, les

génomés mitochondriaux devaient avoir un taux de couverture supérieur ou égal à 5 X pour être utilisés ; ce qui est le cas de tous les génomes mitochondriaux. De plus, les positions non couvertes pour un des génomes ont été retirées des autres séquences ; au final, les génomes utilisés pour construire l'arbre font 15 482 nucléotides. Le set de données comporte également 2 génomes mitochondriaux d'hyène des cavernes, d'un site pyrénéen obtenus précédemment au laboratoire (Bon *et al.*, 2012), ainsi que les séquences mitochondriales de l'hyène tachetée (*Crocuta crocuta*) et de l'hyène rayée (*Hyena hyena*) africaines actuelles. La mangouste (*Herpestes javanicus*) est utilisée comme groupe externe.



**Figure 55.** Arbre phylogénétique réalisé avec les génomes mitochondriaux d'hyène des cavernes des différents sites étudiés. L'analyse a été effectuée avec les portions du génome (15 482 nucléotides, soit 90% de la séquence complète) couvertes pour tous les échantillons. Les séquences en vert correspondent à des spécimens actuels. Les nombres en bleu indiquent les valeurs de *Bootstrap* pour la méthode Maximum Likelihood, ceux en rouge pour la méthode Neighbor Joining, et ceux en vert pour la méthode Minimum Evolution. Les séquences de Coumère ont été obtenues dans une précédente étude (Bon *et al.*, 2012). L'arbre présenté est celui obtenu avec la méthode Maximum Likelihood.

La topologie de l'arbre est la même pour toutes les méthodes de construction. Les valeurs des nœuds sont toujours égales à 100 ou à 99 ; l'arbre obtenu présente donc une grande fiabilité. La séquence mitochondriale de l'hyène tachetée actuelle (*Crocuta crocuta*) se trouve parmi les différents génomes mitochondriaux obtenus pour l'hyène des cavernes. La position de l'hyène tachetée actuelle semble aller dans le sens de l'hypothèse qui propose que

l'hyène des cavernes n'est pas une espèce distincte mais doit être considérée comme l'hyène tachetée du Pléistocène (Rohland *et al.*, 2005). Cependant, la séquence obtenue pour le coprolithe de la Grotte de Néron se positionne en dehors du cluster formé par l'hyène tachetée actuelle et les autres génomes mitochondriaux d'hyène des cavernes. La datation des différents échantillons - réalisée par T. Stafford (Université d'Aarhus, Danemark) - montre que la position originale de la séquence de la Grotte de Néron n'est pas due à l'âge de l'échantillon (Tableau 17).

**Tableau 17.** Dates obtenues pour les échantillons des différents sites grâce à la datation par la méthode radiocarbone. Les dates ont été réalisées par T. Safford (Université d'Aarhus, Danemark).

UCIAMS	Site	$\delta^{13}\text{C}$	Âge $^{14}\text{C}$	$\pm$	2 $\sigma$ cal BP	Commentaire
143983	Salomé	- 25,8	11 370	35	13 295 - 13 124	
143986	Deux-Ouvertures	- 25,5	25 030	90	29 384 - 28 780	
	Grotte-aux-Ours					Non datable
143988	Grotte-aux-Points	- 24,9	21 240	60	25 781 - 25 371	
143991	Néron	- 20,4	26 270	130	30 909 - 30 254	
143993	Châteaubourg	- 22,0	33 930	240	38 988 - 37 718	
143995	Trou du Renard	- 21,2	21 240	90	25 807 - 25 325	
143997	Soyons	- 26,6	17 670	60	21 634 - 21 099	

Les différentes datations réalisées montrent que l'échantillon de la Grotte de Néron n'est pas plus ancien ou plus récent que les autres coprolithes ; en effet, son âge est similaire à ceux des autres échantillons. Sa position originale est peut-être due à un haplogroupe différent.

Le vernissage de certains échantillons - ceux de la Grotte-aux-Points, du Trou-du-Renard et du site de Soyons - ne semble pas porter préjudice à la datation de ceux-ci. Tous les échantillons présentent un âge similaire, sauf le coprolithe du Réseau Salomé qui est plus récent.

Un des échantillons, le coprolithe provenant de la Grotte-aux-Ours, n'a pas pu être daté à cause de sa teneur en carbone qui s'est révélée trop faible (< 10  $\mu\text{g}$ ).

La date obtenue pour le coprolithe de la Grotte des Deux-Ouvertures est cohérente avec des datations précédemment réalisées à partir d'ossements d'ours des cavernes de cette même grotte (Bon *et al.*, 2011).

### 3.1.2. Panorama de l'alimentation de l'hyène des cavernes

Après l'analyse des séquences d'hyènes des cavernes présentes dans les différents échantillons, les fragments d'ADN d'origine animale qu'ils contiennent ont été répertoriés. Après avoir éliminé les lectures appartenant à l'hyène des cavernes afin de ne pas biaiser les analyses, les résultats ont été normalisés pour 100 millions de lectures et compilés dans le tableau ci-dessous (Tableau 18).

**Tableau 18.** Nombres de séquences des banques de fragments - construites à partir des coprolithes d'hyène des cavernes - qui s'alignent sur les génomes mitochondriaux des espèces indiquées pour 100 millions de lectures (+ : 10-99 lectures ; ++ : 100-999 lectures ; +++ ≥ 1 000 lectures).

	<i>Bos primigenius</i>	<i>Equus przewalskii</i>	<i>Ursus spelaeus</i>	<i>Talpa europaea</i>	<i>Bison bonasus</i>	<i>Bison bison</i>	<i>Capra ibex</i>	<i>Sus scrofa</i>
Deux-Ouvertures	+			++				
Néron		+++						
Trou-du-Renard			++					
Salomé			+					
Grotte-aux-Points							++	
Grotte-aux-Ours	++				+++	++		
Soyons								+

Les résultats obtenus pour 7 sites montrent une alimentation variée pour l'hyène des cavernes, composée de gros comme de petits animaux. Il est possible de noter que la quantité d'ADN animal - autre que celui de l'hyène - n'est pas la même dans tous les échantillons ; le nombre de séquences varie de quelques lectures à plus d'un millier de lectures. De plus, le nombre de lectures d'hyène contenu dans la banque ne semble pas avoir de rapport avec le nombre de séquences appartenant à d'autres espèces animales. En effet, certaines banques qui présentent plusieurs milliers de lectures d'hyènes des cavernes sont paradoxalement très pauvres en ADN d'autres animaux. Ces différences peuvent dépendre de la partie de l'animal ingérée par l'hyène des cavernes, les tissus mous contiennent par exemple une quantité plus importante d'ADN que les ossements (Hofstetter *et al.*, 1997 ; Hochmeister *et al.*, 1991).

Pour l'un des échantillons (Châteaubourg, Chât1Q) qui contient une quantité importante de séquences d'hyène des cavernes - plus de 60 000 pour 100 millions de lectures - nous n'avons pas pu mettre en évidence d'AD mitochondrial de proies. Il est possible d'envisager que l'ADN de l'alimentation ne se soit pas conservé ou que l'hyène ait mangé un congénère, le cannibalisme étant avéré chez cette espèce.

L'analyse des séquences animales contenues dans le coprolithe de la Grotte-aux-Ours, située dans le Lot, présente un profil intéressant (Figure 56).



**Figure 56.** Coprolithe d'hyène des cavernes de la Grotte-aux-Ours, située dans le Lot.

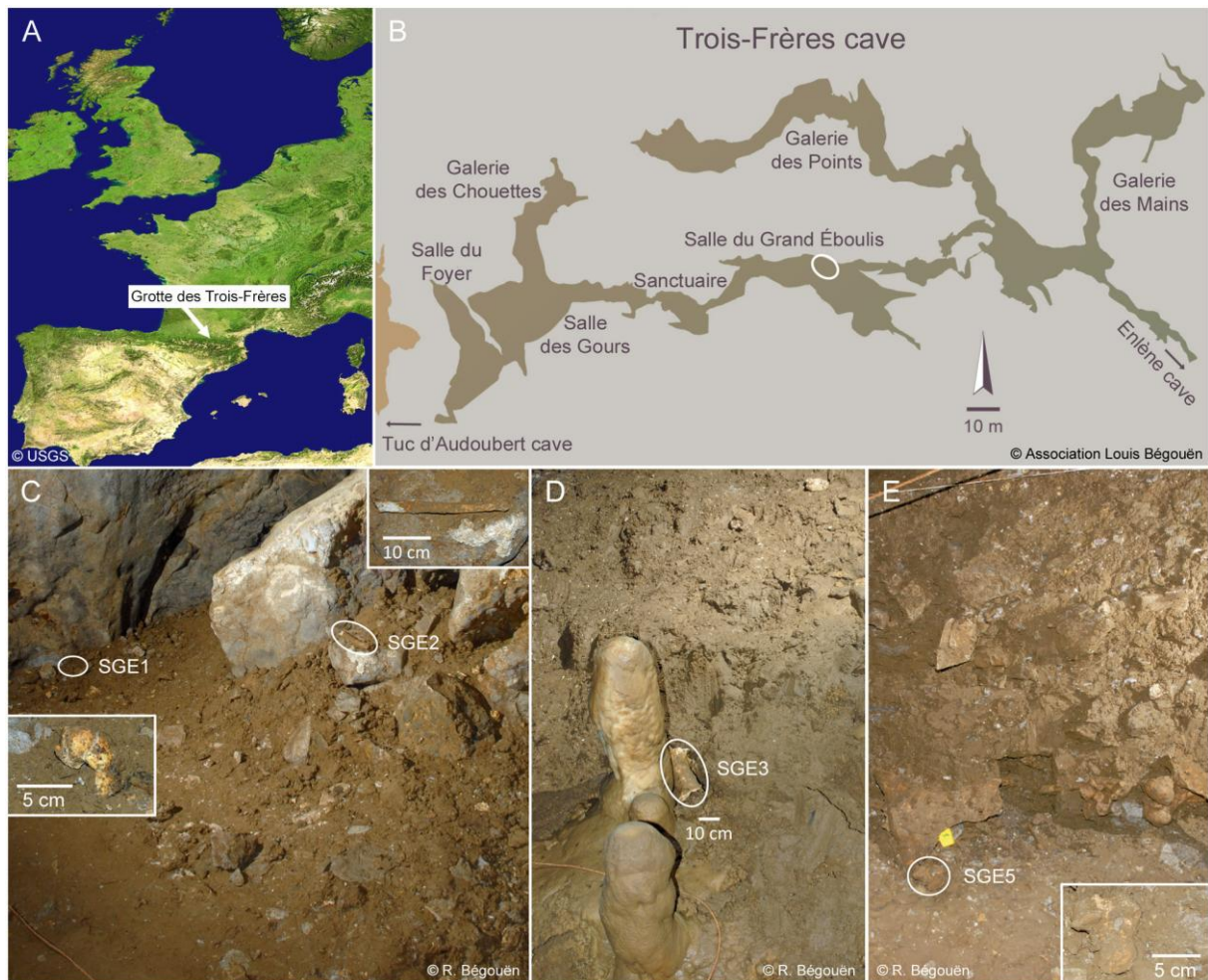
Des séquences de bovinés et plus vraisemblablement de bison sont mises en évidence. En effet, les lectures obtenues présentent trop de différences avec le génome mitochondrial complet de l'aurochs (*Bos primigenius* ; GU985279) pour que le spécimen appartienne à cette espèce. Les informations génétiques contenues dans les banques de données publiques ne permettent cependant pas de trancher sur l'espèce de bison dont il s'agit. Une stratégie basée sur la reconstitution du génome mitochondrial de deux espèces éteintes de bison - *Bison priscus* et *Bison schoetensacki* - a donc été mise en place pour définir celle dont l'ADN est détecté dans le coprolithe de la Grotte-aux-Ours. Cette stratégie est présentée dans les paragraphes qui suivent.

## **3.2. Reconstitution d'un génome complet de *Bison priscus***

### **3.2.1. Choix de l'échantillon**

Dans le but de reconstituer le génome mitochondrial complet du bison des steppes, des ossements attribués à l'espèce *Bison priscus* ont été collectés par notre équipe dans la Grotte des Trois-Frères, située en Ariège (Figure 57). Les échantillons ont tous été prélevés dans la Salle du Grand Éboulis qui était autrefois un puits dont l'ouverture s'est fermée au début de l'Holocène. Trois des échantillons se trouvaient sur le sol à proximité du cheminement balisé ; seul SGE2 se trouvait à l'écart posé sur un rocher (Figure 57).

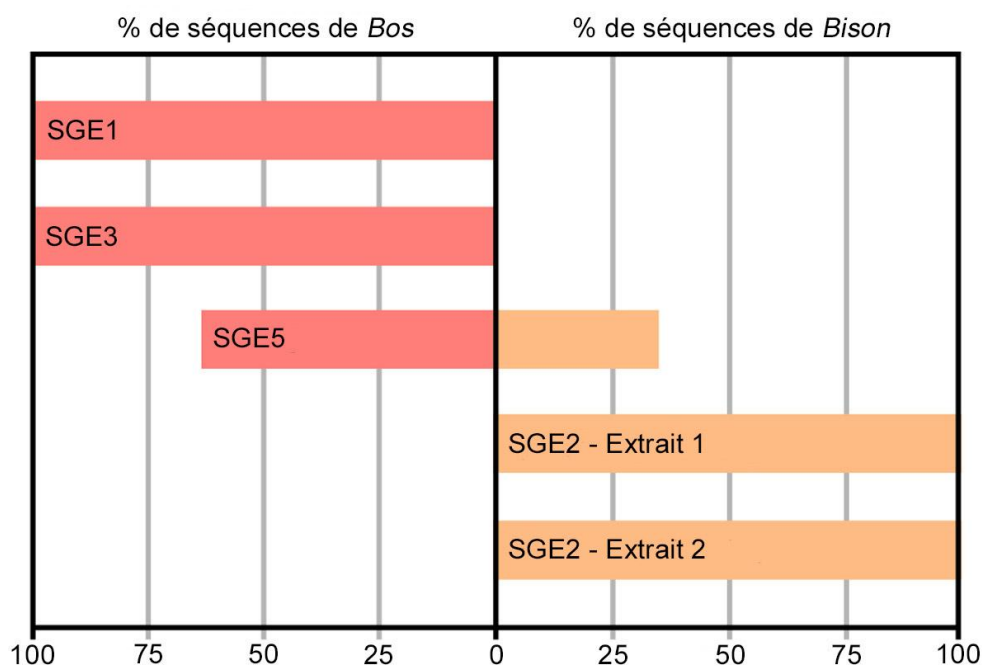




**Figure 57.** Position des échantillons dans la Grotte des Trois-Frères. A : Localisation géographique de la Grotte. B : Topographie de la caverne des Trois-Frères. Le cercle blanc indique l'endroit où les ossements ont été prélevés. C-E : Échantillons utilisés pour les extractions d'ADN (SGE1 : fragment distal d'ulna ; SGE2 : côte ; SGE3 : fragment d'humérus ; SGE5 : vertèbre). D'après (Marsolier-Kergoat *et al.*, 2015).

Afin de caractériser ces échantillons, l'ADN a été extrait et des réactions de PCR, ciblant un fragment de 104 paires de bases du gène mitochondrial du cytochrome B ont été réalisées en utilisant le couple d'amorce n°1 du Tableau 14. Les amorces s'alignent sans aucun mésappariement aux génomes de références de *Bos taurus* et *Bos primigenius*, et présentent 0 à 1 mésappariement avec les génomes de référence de *Bison bison* et *Bison bonasus*. Les amorces encadrent un fragment de 63 paires de bases dont la séquence est identique chez *Bos taurus* et *Bos primigenius*. Cette portion du génome permet de différencier les genres *Bos* et *Bison*. Elle comporte en effet 5 différences (dont une transversion) entre les génomes de *Bos* et ceux de *Bison bison* et *Bison bonasus*, qui tous deux ne diffèrent dans cette région que par 2 transitions. Chaque produit de PCR a été caractérisé par le séquençage de 3 à 33 fragments d'ADN cloné. Ces séquences ont ensuite été comparées au gène du cytochrome

B de référence pour les genres *Bos* et *Bison*. Elles ont été attribuées à l'un ou l'autre des genres en fonction du nombre de différences qu'elles présentaient avec ceux-ci (Figure 58).



**Figure 58.** Résultat des analyses des séquences obtenues par amplification d'un fragment du gène mitochondrial *cytB*. Les produits de PCR ont été clonés et séquencés puis attribués au genre *Bos* ou au genre *Bison* en fonction du nombre de différence avec les génomes de référence. D'après (Marsolier-Kergoat *et al.*, 2015).

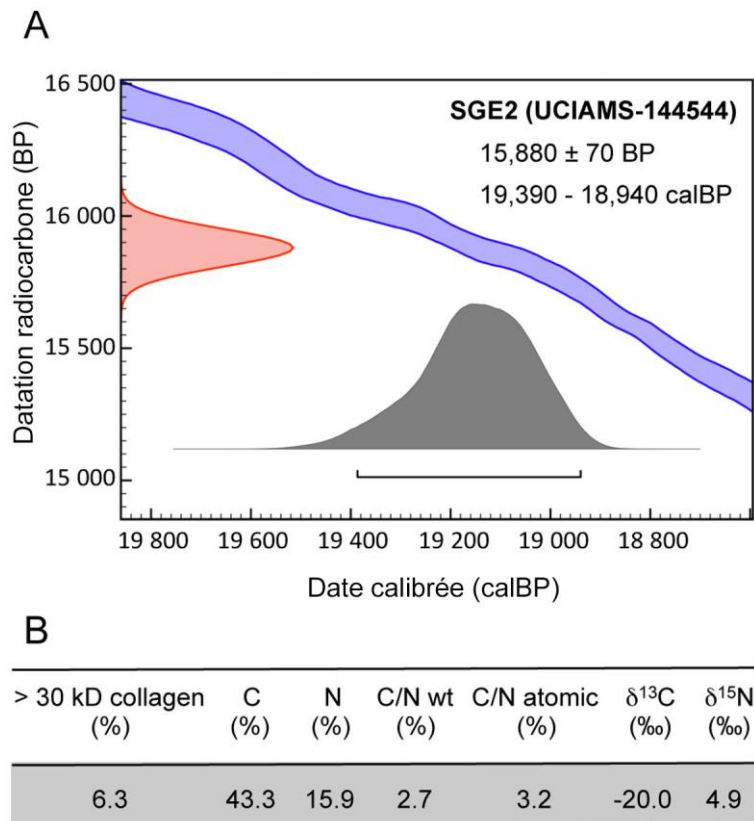
L'analyse des clones pour les échantillons SGE1 (12 clones) et SGE3 (14 clones) montre que ceux-ci contiennent tous des séquences d'ADN qui s'alignent sans aucun mésappariement sur les génomes du genre *Bos*. Ces données suggèrent que les ossements SGE1 et SGE3 correspondent à des spécimens de *Bos*. Une explication alternative est toutefois possible au vu des résultats obtenus sur l'échantillon SGE5 (14 clones) qui fournit un mélange de séquences : une partie d'entre elles s'alignent parfaitement aux génomes de *Bos*, alors que les autres s'alignent sans aucun mésappariement sur le génome de *Bison bison*. On est tenté de conclure que les échantillons SGE1, SGE3 et SGE5, qui sont tous situés dans la zone de cheminement actuel, ont été contaminés par des proportions variables d'ADN depuis la découverte de la Grotte des Trois-Frères. Quoiqu'il en soit, aucun de ces échantillons n'apparaît utilisable pour étudier le génome de bison. L'échantillon SGE2, une côte prélevée à l'écart de la zone de cheminement, apparaît beaucoup plus approprié pour notre étude. Les deux extraits analysés par PCR (extrait 1 : 47 clones ; extrait 2 : 39 clones) contiennent des séquences qui présentent un meilleur alignement sur le génome de *Bison bison* que sur ceux de *Bos taurus* et *Bos primigenius*. La majorité (79%) des séquences

obtenues s'aligne sans aucun mésappariement sur le génome de *Bison bison*, et les autres présentent 1-3 et 6-8 mésappariements avec les génomes de *Bison bison* et de *Bos*.

Dans le but de garantir que l'échantillon SGE2 est bien un ossement de bison des steppes, des réactions de PCR additionnelles ont été réalisées avec des couples d'amorces spécifiques au genre *Bison* (couples 2 et de 7 à 10 ; Tableau 14) et au genre *Bos* (couples 3 à 6 ; Tableau 14). Tous les fragments d'ADN obtenus avec ces réactions de PCR présentent des scores d'alignement supérieurs pour les génomes mitochondriaux de *Bison* par rapport à ceux de *Bos*. De plus, pour les paires d'amorces choisies à partir de la région de contrôle du bison des steppes, les fragments obtenus présentaient à chaque fois un meilleur score sur *Bison priscus* par rapport aux autres espèces. Toutes ces observations permettent de penser que l'échantillon SGE2 est bien un ossement de *Bison priscus*. C'est donc sur cet échantillon que les études ont été poursuivies.

### **3.2.2. Datation de l'échantillon SGE2**

Dans le but de connaître l'âge de l'échantillon, un fragment a été détaché et envoyé pour datation. La caractérisation du matériel montrant une quantité de collagène suffisante (Figure 59B), l'échantillon a pu être daté par T. Stafford (Université d'Aarhus, Danemark) au moyen de la méthode radiocarbone en AMS (Figure 59A). La date obtenue a été calibrée grâce au logiciel OxCal en utilisant la courbe de calibration IntCal13 (Ramsey, 2014).



**Figure 59.** Datation de l'échantillon SGE2. A : âge non calibré et calibré de l'ossement. La date en cal BP a été obtenue avec 2 sigma en utilisant la courbe de calibration IntCal13. B : caractéristique du matériel daté. D'après (Marsolier-Kergoat *et al.*, 2015).

La datation donne un âge de 15 880 +/- 70 BP, ce qui correspond à un âge calendaire de 19 390 à 18 940 cal BP pour deux sigma. Cette date est cohérente avec les précédentes dates réalisées sur des ossements de *Bison priscus*, prélevés eux-aussi dans la Salle du Grand Éboulis (Bégouën *et al.*, 2014). L'analyse des produits de PCR et la date obtenue valident l'échantillon SGE2 pour l'analyse du génome de *Bison priscus*.

### 3.2.3. Caractéristiques de la banque de fragments construite pour SGE2

Le séquençage haut-débit par la méthode Illumina de la banque de fragment a permis d'obtenir 1 milliard de lectures. Dans le but d'explorer la diversité des espèces dont l'ADN est présent dans la banque, 1 million de lectures - choisies aléatoirement - ont été comparées aux séquences disponibles dans la banque de données *GenBank wgs* : 99 684 séquences ont permis d'obtenir des résultats. La majorité de ceux-ci correspondent à des bactéries (89%) alors qu'il y en a seulement 9% pour les mammifères. Parmi les génomes bactériens, les classes les plus présentes sont celles qui regroupent les *Actinobacteria* (32% des séquences

bactériennes), les *Betaproteobacteria* (20%), les *Gammaproteobacteria* (18%) et enfin les *Alphaproteobacteria* (16%). Chez les *Eukaryota*, la famille la plus représentée est celle des *Poaceae* (10%), vient ensuite la famille des *Bovidae* (6%) et enfin celle des *Noelaerhabdaceae* (3%). Les 540 fragments de la banque qui s'alignent sur la famille des *Bovidae* correspondent à 26 856 nucléotides pour 1 million de lectures Illumina ; cette observation permet de prédire un total d'environ 27 754 600 nucléotides de *Bovidae* sur le milliard de fragments séquencés.

Ensuite, afin de vérifier que notre échantillon n'avait pas été contaminé par de l'ADN moderne, les fragments de la banque - d'une taille supérieure ou égale à 30 nucléotides - ont été alignés sur les génomes mitochondriaux de référence de *Bison bison*, *Bison bonasus*, *Bos primigenius* et *Bos taurus*. Les critères d'alignement ne toléraient aucun mésappariement (Tableau 19).

**Tableau 19.** Nombres de lectures s'alignant sans mésappariement sur les génomes mitochondriaux de référence des bovidés indiqués.

Espèce	Code d'accès Genbank	Nombre de lectures alignées
<i>Bison bison</i>	GU985279	1825
<i>Bison bonasus</i>	HM045017	324
<i>Bos primigenius</i>	EU177871	347
<i>Bos taurus</i>	AY526085	331

C'est sur le génome mitochondrial du bison d'Amérique que le plus grand nombre de lectures (1 825) s'alignent. Des plus, un grand nombre des lectures qui s'alignent sur les génomes de *Bison bonasus* et ceux des deux espèces de *Bos* le font également sans mésappariement sur celui de *Bison bison*. C'est le cas pour 296 des fragments obtenus pour *Bison bonasus*, 305 pour ceux de *Bos primigenius* et 290 pour ceux de *Bos taurus*. Cette observation permet d'affirmer que les lectures retrouvées sur les génomes, autres que celui du bison d'Amérique, correspondent à des fragments qui s'alignent sur des régions conservées du génome mitochondrial entre ces 4 espèces, très proches génétiquement.

Puis, nous avons voulu connaître le nombre de fragments de la banque correspondant à des séquences du génome nucléaire de *Bison priscus*. Cette caractérisation a été réalisée grâce à deux méthodes d'évaluation.

Dans un premier temps, les lectures de la banque ont été alignées - en tolérant un mésappariement - sur le gène de l'ARN ribosomal 18S qui est présent en 200 copies environ dans le génome nucléaire. Sur un milliard de fragments, environ 110 s'alignent sur le gène de

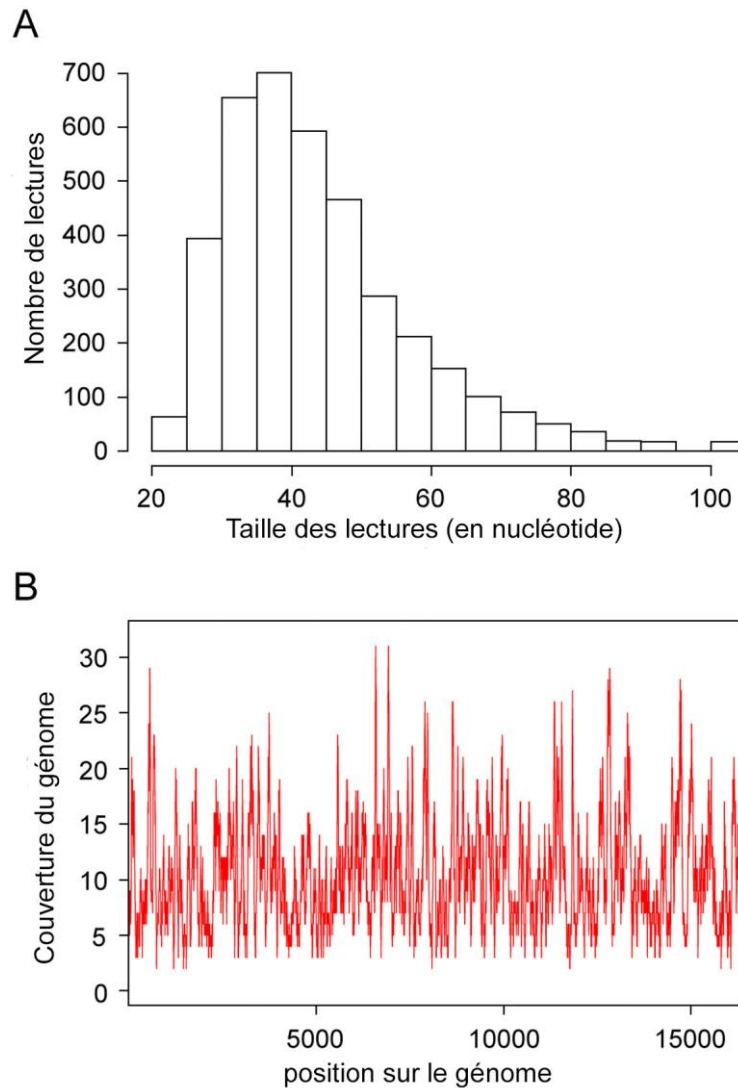
l'ARN 18S de *Bos taurus*. Cela correspond à une couverture d'environ 2,2 pour 200 copies, ce qui équivaut à une couverture de 0,01 pour un gène à copie unique.

Dans un second temps, les lectures ont été alignées sur la totalité du génome nucléaire de *Bos taurus*. Il y a 741 848 fragments qui s'alignent sur ce génome avec un score de qualité supérieur à 20. En totalité, cela représente 26 117 728 paires de bases qui correspondent à une couverture de 0,01 pour 2,7 Mb du génome nucléaire de *Bos taurus*. Cette observation est cohérente avec le résultat précédemment obtenu pour le gène de l'ARN ribosomal du 18S. En raison de la faible couverture du génome nucléaire, son analyse n'a pas été poursuivie.

#### **3.2.4. Reconstitution d'un génome mitochondrial complet de *Bison priscus***

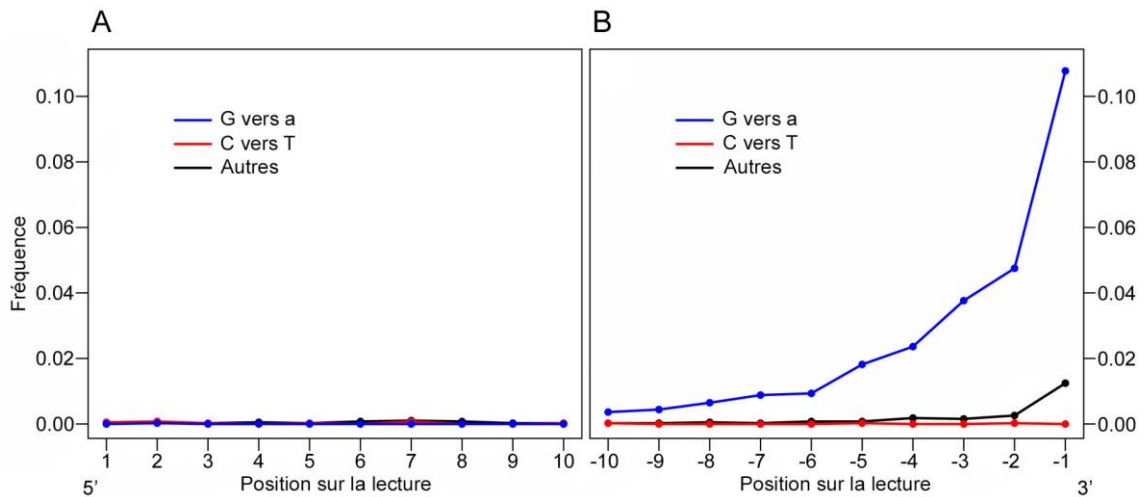
Afin de reconstituer un génome mitochondrial complet, les lectures de la banque sont alignées soit sur le génome du bison d'Amérique - qui est apparu comme le plus approprié dans nos expérimentations précédentes - soit sur la séquence de la région de contrôle de bison des steppes disponibles dans les bases de données (AY748539). Ces alignements sont réalisés avec des paramètres moins stricts que les précédents cribles, 2 mésappariements sont tolérés. Cette analyse donne 3 816 lectures uniques qui permettent d'obtenir un premier consensus avec une couverture de 9,8 X. Cette séquence provisoire est utilisée pour cribler de nouveau la banque avec des critères d'alignement tolérant 0,1 mésappariement (voir 2.3., Méthodes). Ce deuxième alignement permet d'obtenir 3 868 lectures uniques.

Cependant, il reste des parties du génome mitochondrial qui ne sont pas couvertes (39 positions), qui le sont par une lecture unique (176 positions) ou bien des lectures ne permettant pas de déduire une séquence consensus (5 positions). Afin d'obtenir une séquence complète et fiable, des amorces de PCR encadrant ces positions ont été dessinées en utilisant le consensus obtenu précédemment, et les fragments amplifiés ont été clonés et séquencés. Le séquençage des 81 fragments obtenus grâce aux réactions de PCR ciblées a permis de compléter le consensus dérivé des 3868 lectures Illumina. Les lectures de la banque ont été réalignées sur la nouvelle séquence obtenue en tolérant à nouveau 0,1 mésappariement. Au total, 3 851 fragments s'alignent sur la séquence finale, pour une couverture de 10,4 X (Figure 60B). La répartition de taille de ces dernières montre qu'elles sont en majorité entre 30 et 50 paires de bases (Figure 60A) ; elles sont relativement courtes ce qui est caractéristique de l'ADN ancien.



**Figure 60.** Caractéristiques des séquences utilisées pour reconstituer le génome mitochondrial de *Bison priscus*. A : Distribution de la taille des 3 851 lectures Illumina uniques. B : Couverture de chaque position du génome pour le génome mitochondrial obtenu avec les lectures Illumina et les fragments de PCR. D'après (Marsolier-Kergoat *et al.*, 2015).

Le score d'identité entre le consensus et les 3 851 lectures uniques est élevé (99,1 %) ; il y a seulement 1 443 mésappariements sur les 166 849 bases alignées. La majorité de ces différences est due à l'augmentation du taux de substitution G vers A à l'extrémité 3' des fragments (Figure 61). Ce profil est caractéristique des banques d'ADN ancien générées avec notre protocole (Jónsson *et al.*, 2013).



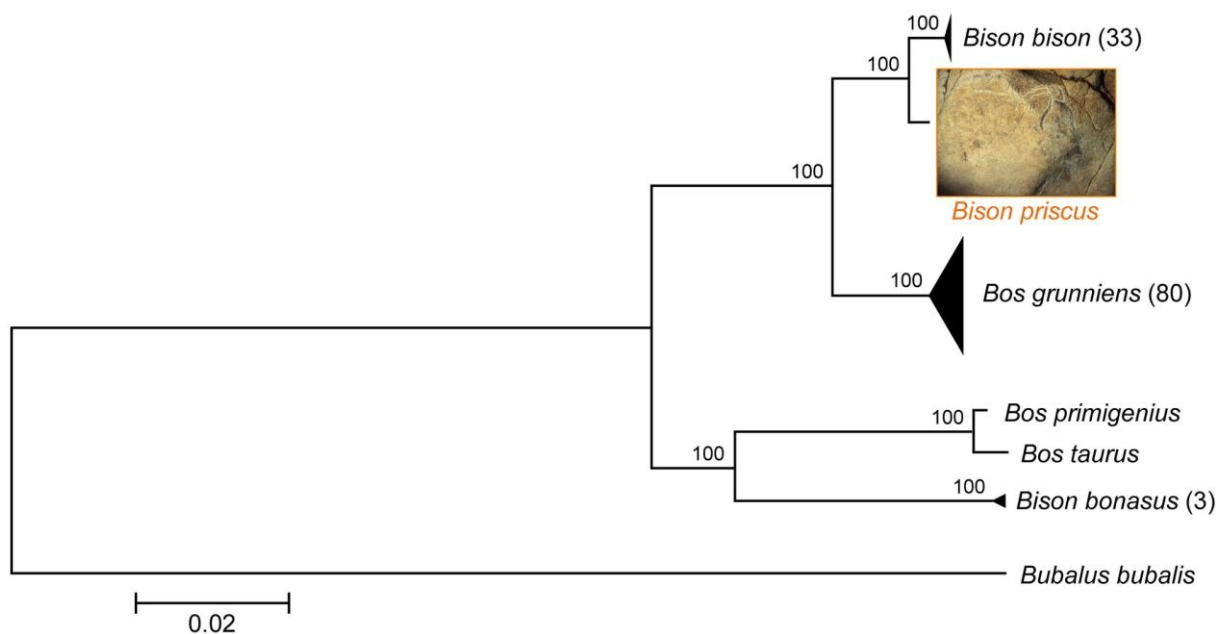
**Figure 61.** Fréquence des mutations observées pour les 10 premières et les 10 dernières positions des séquences. L'analyse est réalisée en comparant la séquence des lectures utilisées pour reconstituer le génome mitochondrial à la séquence consensus obtenue. Comme certaines lectures font 20 nucléotides, seulement les 10 premières et dernières bases sont montrées.

La séquence mitochondriale que nous avons obtenue pour *Bison priscus* a une taille de 16 318 paires de bases. Cette longueur est similaire à celles des génomes mitochondriaux de *Bison bison* connus qui sont comprises entre 16 318 et 16 323 nucléotides. Le génome mitochondrial obtenu est - comme attendu - constitué de 13 régions codantes, 22 gènes d'ARN de transfert, 2 gènes d'ARN ribosomique et 1 région de contrôle. L'absence de codon stop interne dans les gènes codant des protéines atteste de la qualité de la séquence obtenue. Il y a entre 106 et 117 différences entre notre génome mitochondrial et les 33 séquences de bison d'Amérique - disponibles dans les banques de données - qui diffèrent entre elles de 24 nucléotides au maximum. Comparé à la séquence de référence pour *Bison bison* (EU177871), notre génome de *Bison priscus* présente 114 différences dont une insertion de 1 paire de bases dans la région de contrôle et une délétion de 2 paires de bases à la fin du tRNA qui code la sérine. Cette délétion est également présente dans certains génomes mitochondriaux de *Bison bison* (Douglas *et al.*, 2011). Les autres différences correspondent à 105 transitions et 6 transversions ; 25 de ces différences sont situées dans la région de contrôle et 73 dans les régions codant des protéines. Pour les mutations qui sont présentes dans les gènes, 16 touchent le premier nucléotide du triplet, 3 le deuxième et 54 le troisième ; ces mutations entraînent le changement de 11 acides aminés seulement. Sur les 114 différences observées, 93 correspondent à des polymorphismes uniques de l'espèce *Bison priscus*.



### 3.2.5. Position phylogénétique du génome mitochondrial de *Bison priscus*

Afin de déterminer la position phylogénétique de *Bison priscus* au sein la famille des bovinés, une série d'analyses a été effectuée. Le set de données utilisé est composé des génomes mitochondriaux de *Bison bison*, *Bison bonasus*, *Bos grunniens* (le yak) et avec les génomes de référence de *Bos primigenius*, de *Bos taurus* et de *Bubalus bubalis* (le buffle). Les arbres phylogénétiques ont été construits avec différentes méthodes : Maximum Likelihood, Minimum Evolution et Neighbor Joining (Figure 62).



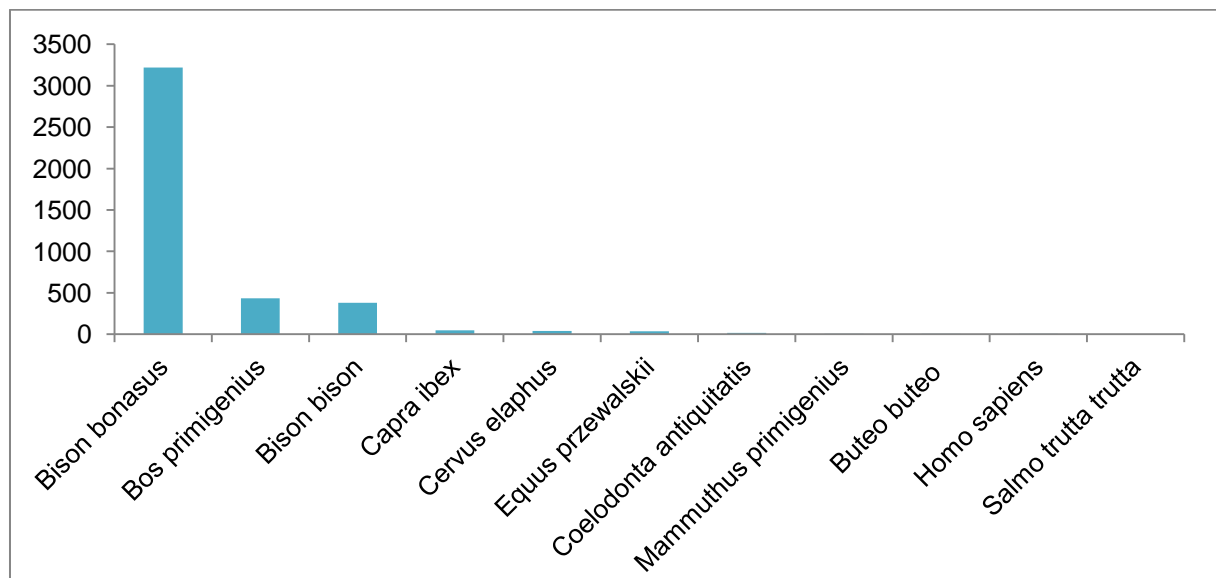
**Figure 62.** Position phylogénétique du génome mitochondrial de *Bison priscus*. Les valeurs entre parenthèses indiquent le nombre de séquences de l'espèce considérée prises en compte pour l'analyse. La topologie de l'arbre et les longueurs de branche sont celles obtenues avec la méthode *Maximum likelihood*. Les valeurs de *bootstrap*, déterminées à partir de 1 000 répliquats) sont indiquées au niveau des nœuds de l'arbre. Des topologies identiques et des valeurs de *bootstrap* égales à 100 pour tous les nœuds sont obtenues avec les méthodes *Neighbor-Joining* et *Minimum Evolution*. D'après (Marsolier-Kergoat *et al.*, 2015).

La topologie de l'arbre est la même pour toutes les méthodes utilisées et les valeurs de *bootstrap* sont égales à 100 pour tous les nœuds et toutes les méthodes de construction. La topologie obtenue présente donc une grande fiabilité. La séquence mitochondriale que nous avons obtenue à partir de l'ossement de *Bison priscus* est basale au clade formé par les 33 séquences mitochondriales de *Bison bison*. Ce résultat est en accord avec l'hypothèse qui place le bison des steppes comme l'ancêtre des bisons d'Amérique.

### 3.3. Reconstitution d'un génome mitochondrial de boviné à partir du coprolithe de la Grotte-aux-Ours

#### 3.3.1. Analyses des ADN d'origine animale présents dans le coprolithe de la Grotte-aux-Ours

Grâce à un script informatique développé au laboratoire à l'occasion de ce projet, les lectures de la banque sont alignées en parallèle sur le génome mitochondrial de l'hyène des cavernes et de 50 proies potentielles en n'autorisant aucun mésappariement. Ce crible a pour but de déterminer l'espèce de bison dont l'ADN du coprolithe est le plus proche, et de détecter l'ADN d'éventuels autres animaux consommés par l'hyène des cavernes. Les lectures qui s'alignent avec un score identique sur le génome de plusieurs espèces sont jugées non pertinentes et sont éliminées afin de ne pas biaiser l'analyse. La figure 63 présente les résultats obtenus pour 8 espèces représentant les grands groupes criblés et 3 bovinés, et les résultats pour 50 espèces sont reportés dans le Tableau 20.



**Figure 63.** Nombre de séquences des 4 pistes de la banque présentant un alignement parfait sur un échantillonnage de génomes mitochondriaux de proies potentielles.

L'ADN d'hyène des cavernes est l'ADN majoritaire avec 7 550 fragments alignés. L'absence de lectures pour *Homo sapiens* indique qu'il n'y a pas eu de contaminations décelables durant l'extraction ou lors de la construction et du séquençage de la banque de fragments.

Le boviné pour lequel nous avons obtenu le plus de séquences est le bison d'Europe. En effet, un nombre plus important de lectures s'alignent sur le génome de *Bison bonasus* (3 220 lectures) que sur ceux des autres bovinés (*Bos primigenius* : 432 lectures ; *Bison priscus* : 381 lectures).

Les seules espèces pour lesquelles nous avons obtenu plusieurs dizaines de lectures appartiennent toutes à la famille *Bovidae*. L'alignement des lectures sur ces génomes provient probablement d'un biais dû aux dommages subis par l'ADN ancien et au fait que les lectures sont alignées sur des génomes connus qui, comme nous le verrons, présentent un nombre important de différences avec celui de l'espèce de bison réellement consommée par l'hyène des cavernes. Ainsi, à ce stade de l'analyse, le nombre de fragments d'ADN de bison présent dans le coprolithe est sous estimé, et celui d'autres espèces de bovidés est surestimé.

Le faible nombre de lectures alignées sur les autres génomes mitochondriaux semble indiquer qu'aucun autre ADN - hormis celui d'hyène et de bison - n'est présent en quantité importante dans l'échantillon.

**Tableau 20.** Nombre de lectures des 4 pistes de la banque présentant un alignement parfait sur les génomes mitochondriaux indiqués. Les séquences qui s'alignent avec le même score sur plusieurs espèces ont été éliminées.

Nom scientifique	Nom vernaculaire	GenBank	Nb de lectures
<i>Crocota crocota spelaea</i>	Hyène des cavernes	JF894378	7550
<i>Bison bonasus</i>	Bison d'Europe	HM045017	3220
<i>Bos primigenius</i>	Aurochs	GU985279	432
<i>Bison priscus</i>	Bison des steppes	KM593920	381
<i>Rupicapra rupicapra</i>	Chamois	FJ207539	127
<i>Ovibos moschatus</i>	Bœuf musqué	FJ207536	60
<i>Saiga tatarica</i>	Antilope de Mongolie	JN632700	53
<i>Capreolus capreolus</i>	Chevreuril	JN632610	51
<i>Capra pyrenaica</i>	Bouquetin des Pyrénées	FJ207528	48
<i>Capra ibex</i>	Bouquetin des Alpes	FJ207526	46
<i>Alces alces</i>	Élan européen	JN632595	46
<i>Cervus elaphus</i>	Cerf	AB245427	45
<i>Martes foina</i>	Belette	HM106325	40
<i>Rangifer tarandus</i>	Renne	AB245426	39
<i>Ursus spelaeus</i>	Ours des cavernes	EU327344	35
<i>Equus przewalskii</i>	Cheval de Przewalskii	JN398403	34
<i>Equus asinus</i>	Âne	X97337	30
<i>Canis lupus</i>	Loup gris	DQ480505	26
<i>Sus scrofa</i>	Cochon domestique	AP003428	23
<i>Equus caballus</i>	Cheval	X79547	20
<i>Sorex unguiculatus</i>	Musaraigne	AB061527	16
<i>Coelodonta antiquitatis</i>	Rhinocéros laineux	FJ905813	16

<i>Lepus europaeus</i>	Lièvre européen	AJ421471	15
<i>Oryctolagus cuniculus</i>	Lapin	AJ001588	13
<i>Talpa europeae</i>	Taupe européenne	Y19192	11
<i>Otis tarda</i>	Grande outarde	FJ751803	10
<i>Sciurus vulgaris</i>	Écureuil rouge européen	AJ238588	9
<i>Apodemus agrarius</i>	Mulot rayé	JN629047	7
<i>Mammuthus primigenius</i>	Mammouth laineux	DQ316067	6
<i>Rhinolophus ferrumequinum</i>	Grand rhinolophe	JX084273	6
<i>Elaphe poryphyracea</i>	Couleuvre	GQ181130	5
<i>Gallus gallus</i>	Poule	X52392	5
<i>Accipiter gentilis</i>	Autour des palombes	AP010797	4
<i>Alectoris chukar</i>	Perdrix choukar	FJ752426	3
<i>Falco peregrinus</i>	Faucon pèlerin	AF090338	3
<i>Erinaceus europaeus</i>	Hérisson européen	X88898	3
<i>Apis mellifera scutellata</i>	Abeille africanisée	KJ601784	2
<i>Buteo Buteo</i>	Buse commune	AF380305	2
<i>Microtus kikuchii</i>	Campagnol de Taiwan	AF348082	2
<i>Anas platyrhynchos</i>	Canard colvert	EU009397	2
<i>Podarcis muralis</i>	Lézard des murailles	FJ460597	2
<i>Corvus frugilegus</i>	Corbeau freux	Y18522	1
<i>Homo neanderthalensis</i>	Homme de Néanderthal	AM948965	1
<i>Homo sapiens</i>	Homme	JX893364	1
<i>Salmo salar</i>	Saumon de l'Atlantique	U12143	0
<i>Salmo trutta trutta</i>	Truite	AM910409	0
<i>Alytes obstetricans</i>	Alyte accoucheur	AY585337	0
<i>Anguilla anguilla</i>	Anguille européenne	AP007233	0
<i>Esox lucius</i>	Brochet	AP004103	0

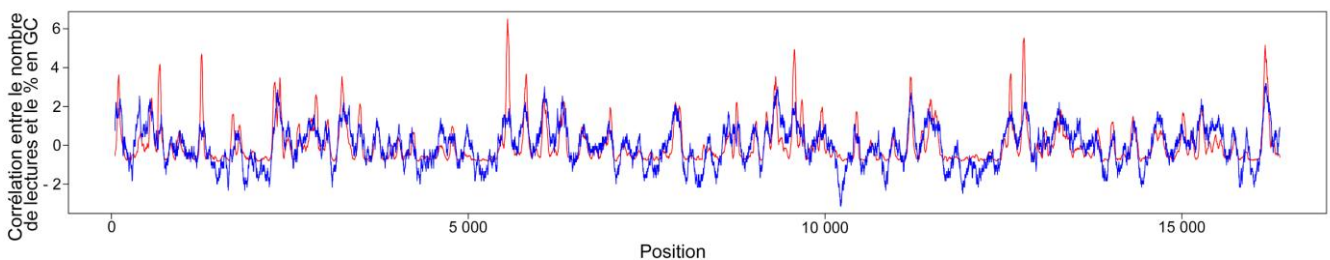
---

### 3.3.2. Reconstitution d'un génome mitochondrial pour le boviné de la Grotte-aux-Ours

Dans le but de reconstituer un génome mitochondrial complet pour le spécimen de bison de la Grotte-aux-Ours, les lectures sont alignées sur les génomes mitochondriaux de bovinés disponibles dans les banques de séquences : *Bison bonasus* (HM045017), *Bison bison* (GU947000), *Bison priscus* (KM593920) et *Bos primigenius* (GU985279). L'alignement est réalisé séparément sur chaque génome mitochondrial en tolérant 0,1 mésappariement et 1 indel pour extraire de la banque le plus de lectures possibles de bovinés. Afin de ne pas biaiser l'analyse et comme son ADN est majoritaire dans le coprolithe, les 13 559 lectures appartenant à l'hyène des cavernes, caractérisées lors d'un crible préliminaire, sont éliminées.

Après alignement des lectures, les doublons sont éliminés et 17 862 fragments uniques sont obtenus. Ces séquences permettent d'établir un consensus provisoire pour le spécimen de la Grotte-aux-Ours. Cette première séquence provisoire est utilisée pour cribler une nouvelle fois la banque, en tolérant toujours 0,1 mésappariement et 1 *indel*. Cette nouvelle analyse permet d'obtenir 19 095 séquences et d'établir un consensus de deuxième génération.

Cependant, malgré un fort taux de couverture (57,2 X), le consensus de troisième génération, assemblé avec 19 455 lectures uniques, n'est pas complet. En effet, des régions demeurent non couvertes alors que d'autres présentent un fort taux de couverture. Lors des analyses effectuées pour comprendre ces différences, la corrélation entre les régions riches en GC et le taux de couverture a pu être établie (Figure 64). Cette observation avait déjà été faite par d'autres auteurs qui proposent que ce biais soit essentiellement dû à l'étape de PCR (Aird *et al.*, 2011).



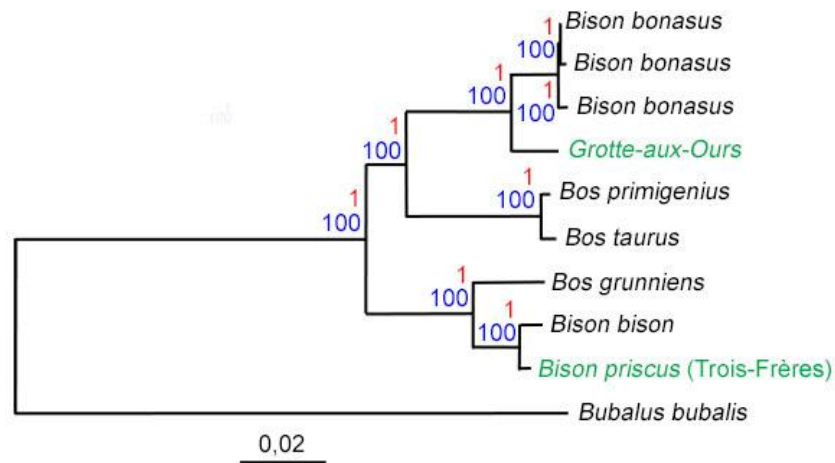
**Figure 64.** Corrélation entre le pourcentage en GC et le taux de couverture pour le génome du bovidé contenu dans le coprolithe de la Grotte-aux-ours.

Afin de combler les positions non couvertes, 25 couples d'amorces spécifiques sont dessinés à partir du consensus de dernière génération et des réactions de PCR ciblées sont effectuées. Les séquences obtenues à partir de ces réactions de PCR ont permis de reconstituer un génome mitochondrial complet pour le spécimen de la Grotte-aux-Ours.

### 3.3.3. Position phylogénétique du spécimen de boviné de la Grotte-aux-Ours

Une fois le génome mitochondrial complet obtenu pour le spécimen de la Grotte-aux-Ours, des analyses phylogénétiques ont été réalisées afin d'explorer sa position dans la phylogénie des bovinés. Les arbres phylogénétiques ont été construits au moyen de différentes méthodes : Neighbor Joining, Maximum Likelihood et Minimum Evolution, et une analyse bayésienne. Le set de données comprend les 6 génomes mitochondriaux de référence de bovinés et les 3 génomes mitochondriaux complets de *Bison bonasus* disponibles dans les

banques de données. Le buffle (*Bubalus bubalis*) est utilisé comme groupe externe (Figure 65).

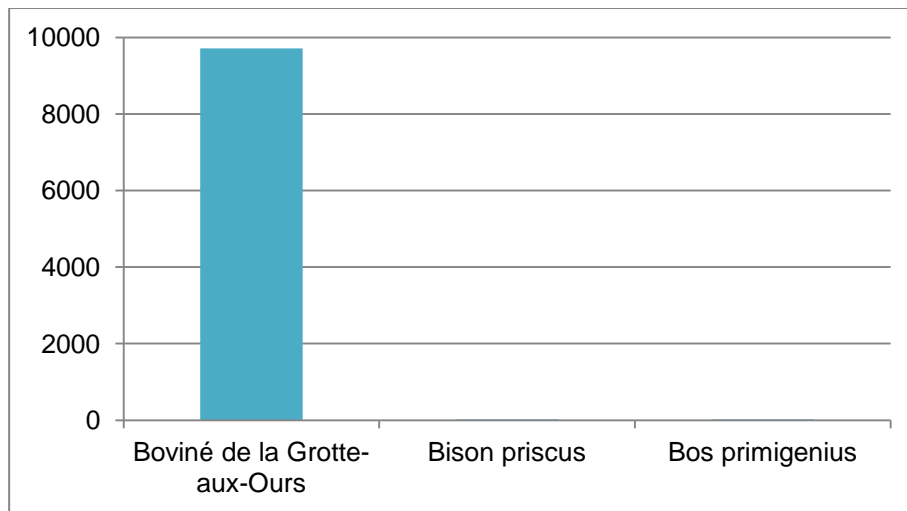


**Figure 65.** Arbre phylogénétique établi avec les génomes mitochondriaux complets obtenus à partir des échantillons de la Grotte des Trois-Frères (ossement) et de la Grotte-aux-Ours (coprolithe). Les chiffres en rouge indiquent les valeurs des probabilités a posteriori pour l'analyse bayésienne et les nombres en bleu les valeurs de *Bootstrap* pour les autres méthodes d'analyse. La topologie et les longueurs de branche sont celles obtenues avec la méthode Maximum Likelihood. Une topologie identique est obtenue avec les autres méthodes.

La topologie de l'arbre est la même pour toutes les méthodes de construction utilisées. Les valeurs des nœuds sont toujours égales à 100 pour les *Bootstrap* ; l'arbre phylogénétique obtenu présente donc une grande fiabilité. Le génome mitochondrial du spécimen de la Grotte-aux-Ours occupe une position phylogénétique éloignée de celle du bison des steppes et des bisons d'Amérique. En effet, ce génome mitochondrial se positionne à proximité des séquences de *Bison bonasus* ; cependant, il est à l'extérieur du cluster formé par ces séquences, indiquant qu'il correspond à un haplogroupe original, voire à une autre espèce. L'arbre phylogénétique permet ainsi de conclure que le génome mitochondrial du spécimen de la Grotte-aux-Ours ne correspond pas à l'espèce *Bison priscus*. La séquence de boviné reconstruite à partir du coprolithe pourrait être celle d'un spécimen de l'espèce *Bison schoetensacki*, ou bison des forêts, proposée comme l'ancêtre du bison d'Europe par certains auteurs, *Bison schoetensacki*, aujourd'hui éteint, était en effet présent en Europe à l'époque où celle-ci comportait des spécimens d'hyène des cavernes.

Dans le but de s'assurer que la séquence mitochondriale de boviné reconstituée à partir du coprolithe n'est pas une chimère due à la consommation de plusieurs espèces de bovidés par l'hyène des cavernes, les lectures de la banque sont alignées sur le génome mitochondrial

obtenu ainsi que sur celui de l'aurochs, celui du bison des steppes et celui de l'hyène des cavernes. Comme pour le crible précédent, les lectures qui présentent un score d'alignement identique sur deux espèces sont jugées non pertinentes et éliminées. La figure 66 présente les résultats obtenus pour les 3 bovinés.



**Figure 66.** Nombre de séquences des 4 pistes de la banque présentant un alignement parfait sur les génomes mitochondriaux de bovidés indiqués.

Le génome mitochondrial pour lequel nous avons obtenu le plus grand nombre de séquences est celle reconstituée à partir du coprolithe avec 9 711 lectures. En comparaison, le nombre de lectures alignées sur les deux autres bovidés est très faible (*Bos primigenius* : 23 lectures ; *Bison priscus* : 26 lectures). Pour l'hyène des cavernes, le nombre de lectures obtenu avec le second crible est similaire à celui obtenu lors du premier crible : 7 550 lectures pour le premier et 7 690 lectures pour le second. Cette observation permet de montrer la solidité du second crible. Comme nous le pensions lors du premier crible, le nombre de fragments d'ADN de bison présents dans le coprolithe était sous estimé ; en effet, le plus grand nombre de séquences obtenu était 3 220 avec le génome de *Bison bonasus*. L'utilisation du génome mitochondrial reconstitué a permis de tripler le nombre de lectures qui s'alignent. Ce second crible a permis de montrer que la séquence mitochondriale de boviné reconstruite à partir du coprolithe de la Grotte-aux-Ours n'est pas une chimère causée par la consommation de plusieurs bovinés par l'hyène des cavernes.

### 3.4. Analyse du génome mitochondrial de *Bison schoetensacki*

#### 3.4.1. Sélection de l'échantillon

Comme les arbres phylogénétiques ont montré que le spécimen de la Grotte-aux-Ours n'était pas un bison des steppes, des ossements attribués à l'espèce *Bison schoetensacki* ont été analysés. Nous avons pour cette étude, utilisé des spécimens de la Grotte de Siréjol, située à 15 km environ de la Grotte-aux-Ours (Figure 67).

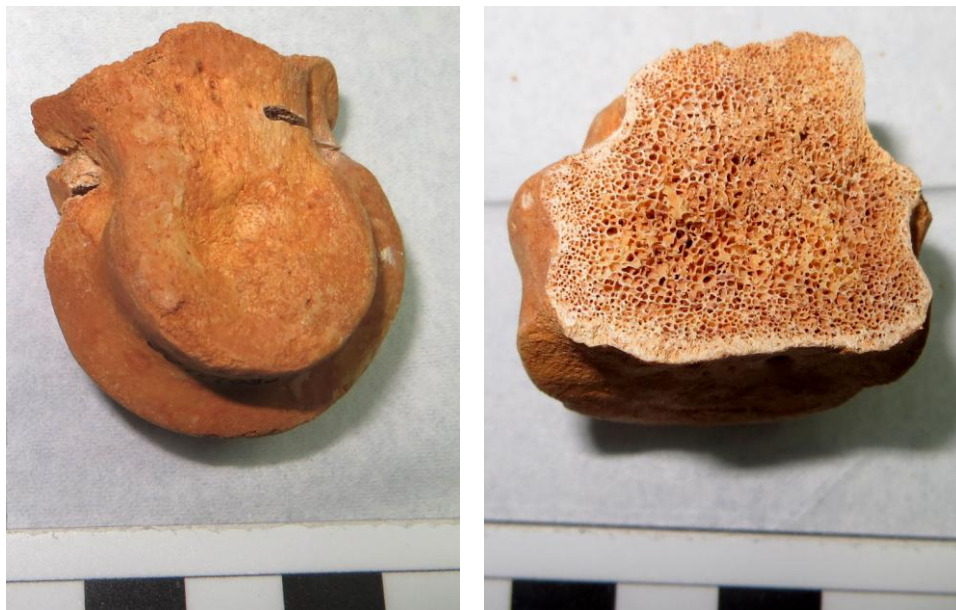


**Figure 67.** Localisation de la Grotte-aux-Ours et de la Grotte de Siréjol, situées dans le Lot.

Des extractions d'ADN ont été effectuées à partir de 5 échantillons. Ces échantillons, issus de fouilles réalisées il y a une quarantaine d'années (Philippe, 1997), étaient recouverts de vernis. Un criblage a été réalisé par PCR au moyen d'une paire d'amorces (15356F - 15437F, Tableau 16) qui s'alignent parfaitement sur le génome mitochondrial reconstitué à partir du coprolithe de la Grotte-aux-Ours et celui de *Bison bonasus* ; elles présentent chacune 1 mésappariement avec les génomes de *Bos taurus* et *Bos primigenius*. Ces amorces



amplifient un fragment de 82 pb du gène mitochondrial *cytB*. Trois échantillons n'ont pas permis d'amplification, et un a fourni 100% de séquences de *Bos taurus*, ce qui n'est pas surprenant pour un échantillon vernis, dont la probabilité de contamination par de l'ADN moderne est élevée. Enfin l'échantillon SirB4, un fragment d'os canon (Figure 68), a fourni des séquences qui s'alignent parfaitement sur le génome mitochondrial reconstitué à partir du coprolithe de la Grotte-aux-Ours, mais présentent plusieurs mésappariements avec celui des autres bovinés (*Bison bonasus* : 2 mésappariements ; *Bos taurus* et *primigenius* : 4 ; *Bison priscus* : 5).



**Figure 68.** Os canon de la Grotte de Siréjol attribué à un spécimen de *Bison schoetensacki*.

Les analyses ont donc été poursuivies sur cet échantillon. Il n'est pas indifférent de noter que l'extraction de l'ADN de SirB4 a pu être réalisée à partir de la moelle, ce qui a probablement limité les risques de contamination.

### **3.4.2. Datation de l'os de *Bison schoetensacki***

Dans le but de dater l'os canon de *Bison schoetensacki* de la Grotte de Siréjol, un fragment de celui-ci a été envoyé à T. Stafford (Université d'Aarhus, Danemark) pour datation par la méthode radiocarbone. Afin d'obtenir une date fiable, deux fractions chimiques différentes ont été datées. Les résultats obtenus pour les deux fractions sont compilés dans le tableau ci-dessous (Tableau 21).

**Tableau 21.** Dates obtenues pour l'os canon de *Bison schoetensacki* de la grotte de Siréjol. Deux fractions chimiques différentes ont été datées. Les datations ont été réalisées par T. Stafford (Université d'Aarhus, Danemark).

UCIAMS	Échantillon	Fraction	$\delta^{13}\text{C}$	Âge $^{14}\text{C}$	$\pm$	$2\sigma$ cal BP
012204	SirB4	1	- 24,2	32 316	215	36 722 - 35 696
012205	SirB4	2	- 29,1	32 623	200	37 228 - 36 060

Les deux fractions chimiques datées donnent un âge de  $32\,316 \pm 215$  BP et  $32\,623 \pm 200$  BP ce qui équivaut à un âge calendaire de 36 722 à 35 696 cal BP pour la première et 37 228 à 30 060 cal BP pour la seconde. La date moyenne pour les deux datations donne un résultat à  $32\,469 \pm 147$ , ce qui équivaut à un âge calendaire de 36 765 à 36 001 cal BP pour deux sigma. L'âge obtenu pour l'os canon de Siréjol est en accord avec les deux datations précédemment réalisées dans la grotte pour les couches contenant des ossements de *Bison schoetensacki*. Les dates obtenues correspondaient à 31 300 (+ 1 800 / - 1 600) BP et à 29 100 (+1 600 / - 1 300) BP (Philippe *et al.*, 1980). Les produits de PCR précédemment analysés et l'âge obtenu valide l'échantillon SirB4 pour l'analyse du génome mitochondrial de *Bison schoetensacki*.

### 3.4.3. Analyse de fragments du génome mitochondrial de *Bison schoetensacki*

La caractérisation du spécimen de la Grotte de Siréjol a été poursuivie par PCR au moyen d'amorces déduites de la séquence de bison de la Grotte-aux-Ours, utilisées préalablement avec succès sur le coprolithe de cette grotte. Seules les amorces amplifiant des fragments de moins de 100 paires de bases ont fonctionné sur l'os de Siréjol, rarement dans 100% des tests, indiquant que cet échantillon contient de l'ADN très fragmenté, en petite quantité. Nous n'avons donc pas tenté de construire de banque de fragments à partir de l'os de Siréjol. Nous avons en revanche sélectionné, toujours à partir de la séquence du bison de la Grotte-aux-Ours, une nouvelle série d'amorces pour l'amplification de fragments de 59 à 86 paires de bases (Tableau 16). Au total, les réactions de PCR réalisées avec 16 couples d'amorces ciblant différentes portions du génome mitochondrial ont fourni 609 paires de bases pour le spécimen SirB4 de la Grotte de Siréjol.

Dans le but de savoir de quel boviné le spécimen de Siréjol est le plus proche, une matrice de différence a été construite en comparant les 609 paires de bases obtenues aux génomes de référence (Figure 69).

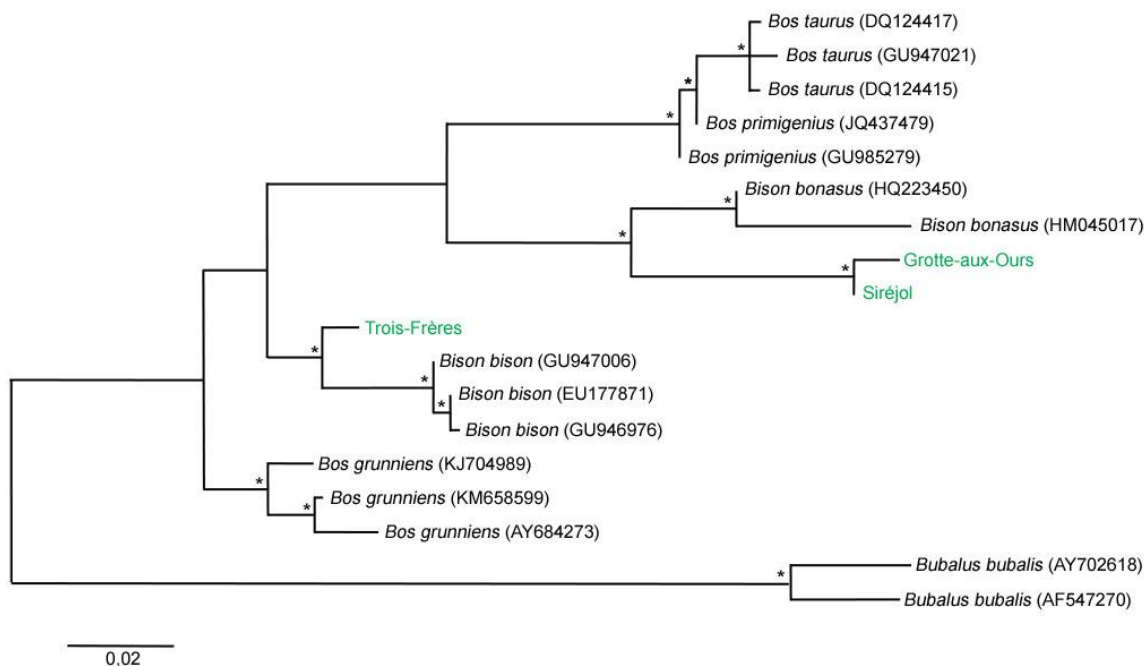
	<i>Bos taurus</i>	<i>Bos primigenius</i>	<i>Bison bison</i>	<i>Bison priscus</i>	<i>Bison bonasus</i>	Bovidé de la Grotte-aux-ours
SirB4	62	60	68	63	54	6

**Figure 69.** Nombre de différences entre les 609 paires de bases obtenues pour l'ossement de la Grotte de Siréjol et les génomes de référence des espèces indiquées.

La matrice montre qu'il y a une soixantaine de différence entre le spécimen de Siréjol et les *Bos* (*Bos taurus* : 62 ; *Bos primigenius* : 60), les bisons actuels (*Bison bison* : 68 ; *Bison bonasus* : 54) et le bison des steppes (*Bison priscus* : 63). En revanche, il n'y a que 6 différences - 5 dans la région de contrôle et 1 dans la séquence du gène du cytochrome B - entre le génome de Siréjol et le génome de bovidé obtenu à partir du coprolithe de la Grotte-aux-Ours. Ces premières observations semblent suggérer que le spécimen obtenu à partir du coprolithe appartient bien à l'espèce *Bison schoetensacki*.

#### 3.4.4. Position phylogénétique de *Bison schoetensacki*

Les 609 paires de bases obtenues par PCR pour le spécimen de *Bison schoetensacki* de Siréjol ont été alignées sur les génomes mitochondriaux de bovinés, et des analyses phylogénétiques ont été effectuées. Des arbres phylogénétiques ont été construits à l'aide des méthodes Neighbor-Joining, Maximum Likelihood et Minimum Evolution. Le set de données est constitué de génomes mitochondriaux de *Bison bison* (3 séquences), de *Bos grunniens* (3 séquences), de *Bos primigenius* (2 séquences), de *Bos taurus* (2 séquences), de *Bison bonasus* (3 séquences), de *Bubalus bubalis* (2 séquences) et du génome mitochondrial de bison des steppes ainsi que celui du spécimen de la Grotte-aux-Ours. Le buffle est utilisé comme groupe externe (Figure 70).



**Figure 70.** Arbre phylogénétique réalisé avec les génomes partiels (609 positions). Les étoiles correspondent aux valeurs de *Bootstrap* supérieures à 70 pour les trois méthodes de construction. Les séquences en vert correspondent aux génomes anciens obtenus à partir des ossements et du coprolithe de la Grotte-aux-Ours. La topologie représentée est celle obtenue avec la méthode Maximum Likelihood.

La topologie obtenue pour l'arbre est la même avec les 3 méthodes de construction. Les valeurs de *Bootstrap* pour les différentes méthodes de construction sont, pour une grande majorité, supérieures à 70. La séquence mitochondriale partielle du bison des steppes se place en position basale du cluster formé par les bisons d'Amérique, comme pour l'arbre construit avec les génomes mitochondriaux complets. La séquence partielle reconstituée à partir de l'os de bison des forêts de la grotte de Siréjol se regroupe avec celle obtenue pour le spécimen de la Grotte-aux-Ours, à proximité du cluster formé par les bisons d'Europe. Cette observation permet de proposer que l'ADN de boviné, présent dans le coprolithe, appartient à l'espèce *Bison schoetensacki* plutôt qu'à *Bison bonasus*.

#### 4. CONCLUSION

L'analyse, au moyen de la méthode de séquençage à haut-débit, de coprolithes provenant de sites archéologiques de 2 régions de France, Ardèche et Lot, a permis de mettre en lumière certaines des espèces qui constituaient l'alimentation de l'hyène des cavernes au Pléistocène. Le régime alimentaire observé est varié, constitué d'animaux de grande et de

petite taille, et principalement d'herbivores. De plus, cette étude a souligné, une nouvelle fois, la pertinence de se procurer à chaque fois plusieurs échantillons pour chaque site. En effet, la conservation de l'ADN n'étant pas la même pour tous les coprolithes, et il est important d'en analyser plusieurs afin de sélectionner celui qui contient le plus d'ADN. En revanche, il a été démontré que l'application de vernis sur les coprolithes n'empêchait pas l'amplification de l'ADN contenu dans ces derniers. Les collections de musées sont donc parfaitement utilisables pour de futures recherches sur les coprolithes.

Le séquençage haut-débit par la méthode Illumina de banques de fragments a permis d'obtenir, pour la première fois, la séquence complète du génome mitochondrial de deux espèces éteintes de bison : *Bison priscus* et *Bison schoetensacki*. Le génome mitochondrial de *Bison priscus* a été reconstitué d'un os provenant de la Grotte des Trois-Frères, déjà célèbre pour son art pariétal Magdalénien, et qui offre donc des possibilités pour les études paléogénomiques. Le génome mitochondrial de l'autre espèce de bison, sur laquelle aucune information génétique n'avait été obtenue préalablement, a été décrypté à partir de l'ADN provenant d'un coprolithe d'hyène des cavernes. La comparaison de la séquence de ce génome avec des fragments d'ADN issus d'un os de l'espèce *Bison schoetensacki* nous a permis de nous assurer qu'il correspond à celui de cette espèce

En plus d'apporter des informations sur deux espèces éteintes de bison, les analyses ont permis de réaliser des arbres éclairant la phylogénie de la famille des bovidés. Les analyses phylogénétiques montrent que la séquence mitochondriale complète du bison des steppes se positionne de manière basale par rapport au cluster formé par les bisons d'Amérique, tandis que les génomes - partiel et complet - de bison des forêts sont plus proches des génomes mitochondriaux des bisons d'Europe. Les hypothèses proposées par les paléontologues étaient que le bison des steppes est l'ancêtre des bisons d'Amérique et le bison des forêts celui des bisons d'Europe. Les résultats de notre étude sont en accord avec ces hypothèses. De plus, les études génétiques ont permis de confirmer - avec une autre approche - ce qui était proposé par les paléontologues : la Grotte des Trois-Frères contient des spécimens de *Bison priscus* et *Bison schoetensacki* est présent dans la grotte de Siréjol.

Ces analyses ont également montré l'importance de ne pas négliger les coprolithes comme source d'informations sur les génomes d'espèces éteintes. En effet, les coprolithes contiennent souvent une quantité d'ADN beaucoup plus importante que les ossements. Dans

cette étude, le coprolithe de la Grotte-aux-Ours contenait suffisamment d'ADN de bison des forêts pour reconstituer un génome complet alors que l'ossement de *Bison schoetensacki* n'a permis d'obtenir qu'un génome mitochondrial partiel. De plus, la couverture du génome reconstitué à l'aide du coprolithe était plus importante que celle du génome complet de *Bison priscus* obtenu à partir d'un ossement.

Afin d'aller plus loin dans l'étude sur les bovidés du Pléistocène, une analyse des séquences nucléaires contenu dans le coprolithe de la Grotte-aux-Ours vient d'être initiée au laboratoire en collaboration avec M.-C. Marsolier-Kergoat. Les premières analyses montrent que l'échantillon contient un nombre important de séquences nucléaires ; en effet, on obtient une couverture d'environ 1X pour le génome nucléaire. Ces études pourraient permettre d'éclairer l'introggression présumée d'ADN nucléaire de bisons mâles dans une ancienne population eurasiennne de *Bos* (Verkaar *et al.*, 2004).

Dans le but de compléter ce panorama de l'alimentation de l'hyène des cavernes au Pléistocène, il serait intéressant d'obtenir les génomes mitochondriaux complets pour les espèces obtenues dans les différents coprolithes d'hyène des cavernes.

Le coprolithe provenant de la Grotte-aux-Points contient de l'ADN de bouquetin. La présence de bouquetins dans les gorges de l'Ardèche au Pléistocène est depuis longtemps connue des paléontologues, mais les discussions portent encore sur l'identification de l'espèce: bouquetin des Alpes (*Capra ibex*) ou Bouquetins des Pyrénées (*Capra pyrenaica*). Notre étude montre que l'espèce *Capra ibex* était assurément présente dans les Gorges de l'Ardèche au Pléistocène.

Le coprolithe provenant de la Grotte de Néron contient quant à lui des séquences de cheval. Il serait également intéressant de déterminer de quelle espèce de cheval il s'agit. Selon les paléontologues, il y avait trois espèces de chevaux présentes dans les gorges de l'Ardèche au Pléistocène : *Equus hydruntinus*, *Equus germanicus* et *Equus caballus*. De plus, le génome mitochondrial d'hyène des cavernes, reconstitué à partir du coprolithe de la Grotte de Néron, présente une position phylogénétique différente des autres génomes d'hyènes qui serait intéressante à explorer.



## *Conclusion Générale*



Les travaux présentés dans cette thèse montrent l'importance des coprolithes comme source d'information sur l'ADN ancien. Grâce à l'analyse des fragments d'ADN contenus dans les échantillons étudiés nous avons pu obtenir des données sur plusieurs espèces, notamment des espèces éteintes, qui étaient présentes en Europe durant le Pléistocène.

Dans un premier temps, grâce à un coprolithe prélevé dans les secteurs profonds de la Grotte Chauvet-Pont d'Arc, nous avons pu reconstituer le génome mitochondrial complet d'un loup du Pléistocène. Les analyses phylogénétiques réalisées placent le spécimen de Chauvet en dehors de la diversité mitochondriale des chiens et loups actuels, à l'image des génomes mitochondriaux reconstitués à partir d'ossements trouvés en Belgique et datés de la même période (Thalmann *et al.*, 2013). Nous avons également initié l'étude du génome nucléaire du canidé de la Grotte Chauvet. L'analyse globale du génome nucléaire a permis de déterminer que le spécimen était une femelle. Ensuite, des analyses ciblées de gènes impliqués dans la digestion ont permis de montrer que le génome nucléaire du canidé ne présentait pas de mutations pouvant correspondre à des marqueurs de domestication. Cette observation est cohérente avec l'analyse de l'alimentation qui montre un régime carné avec la consommation d'ours des cavernes.

Dans un second temps, afin de fournir un panorama de l'alimentation de l'hyène des cavernes, nous avons étudié des coprolithes attribués à ce carnivore provenant de huit sites répartis dans deux régions de France (Lot et Ardèche). Ces analyses ont permis de mettre en évidence une alimentation variée composée d'animaux de grande comme de petite taille. L'analyse approfondie de l'un des échantillons, qui contenait des séquences inconnues de bovidés, a permis de reconstituer le premier génome mitochondrial pour une espèce éteinte : le bison des forêts, *Bison schoetensacki*. Un lien avec la paléontologie a été fait en étudiant dans le même temps un ossement de bison des forêts. En parallèle, le premier génome mitochondrial complet de bison des steppes, une espèce de bison elle aussi éteinte, a été reconstitué au laboratoire à partir d'un ossement. L'obtention de ces deux génomes mitochondriaux complets a permis de pallier au manque d'informations disponibles dans les bases de données et d'établir la position phylogénique respectives du bison des steppes et du bison des forêts dans la phylogénie des bovinés.

Nos résultats montrent que les coprolithes sont des échantillons importants et non négligeables dans l'étude des génomes anciens pour plusieurs raisons. Tout d'abord, de part leur nature, ils permettent d'accéder à plusieurs types d'ADN : celui du producteur et celui de son alimentation. Ce mélange de différents ADN ne représente pas un problème pour les cribles bio-informatiques qui permettent de discerner l'espèce à laquelle les séquences d'ADN appartiennent. Un autre avantage des coprolithes est que ce sont des échantillons étanches ; ce qui les rend peu sensibles aux contaminations et protège l'ADN lors de la pose de vernis comme c'est le cas dans les musées. De plus, ces fèces fossilisés contiennent, pour les mieux conservés d'entre eux, une quantité beaucoup plus importante d'ADN que les ossements. Les collections de musées ou les repaires d'hyènes contenant de grande quantité de coprolithes sont des mines d'informations sur ce carnivore mais également ses proies, qui comme l'ont montré nos études sont diverses. Cependant, lorsque le génome de l'espèce étudiée n'est pas encore disponible dans les bases de données - comme c'était le cas pour le bison des forêts - il s'avère nécessaire de disposer d'un ou plusieurs ossements attribués pour faire le lien avec la paléontologie.

L'augmentation constante du nombre de données fournies par les techniques de séquençage à haut-débit, laisse espérer que l'étude des séquences nucléaires pour les échantillons anciens va devenir systématique. Cet apport permettrait de pallier au manque d'informations disponibles sur les génomes nucléaires. Ces paléo-génomes permettraient de d'aborder des questions auxquelles les génomes mitochondriaux, même quand ils sont complets, ne fournissent pas de réponse. Dans le cas de la domestication du loup, les marqueurs de domestications se situent sur le génome nucléaire. L'obtention de plusieurs paléo-génomes de loups ou proto-chiens permettrait d'apporter des éléments nouveaux concernant le lieu et la date de la domestication du loup, qui sont aujourd'hui encore sujet à débat. Pour les bovidés du Pléistocène, la reconstitution de génomes nucléaires, en cours au laboratoire, serait une avancée pour comprendre les mécanismes qui ont donné naissance aux espèces actuelles. Enfin, l'analyse de génomes nucléaires d'hyène des cavernes permettrait de déterminer le statut de ce carnivore : espèce à part ou sous-espèce de l'hyène tachetée africaine actuelle. Cependant, les bases de données sont pauvres en génomes nucléaires, même d'espèces actuelles, ce qui pourrait être un frein pour la reconstitution des paléo-génomes nucléaires. Pour le loup et le bison, l'analyse est facilitée, respectivement, par la comparaison au génome de canidés (loup, chien) et de bœufs actuels. Pour l'hyène des

cavernes, comme aucun génome d'une espèce proche n'est disponible, un séquençage *de novo* serait nécessaire.

De nombreuses connaissances restent à découvrir pour comprendre les mécanismes génétiques qui ont façonné les espèces actuelles. L'analyse systématique de plusieurs échantillons pour chaque site archéologique étudié ainsi qu'une étroite collaboration avec les paléontologues sont quelques une des clés pour y parvenir.

## Bibliographie

Adler, C. J., Haak, W., Donlon, D. & Cooper, A. (2011) Survival and recovery of DNA from ancient teeth and bones. *J of Archeol Sci* 38:956-964.

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. F. Nusbaum, C. & Gnirke, A. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*, 12(2):R18.

Amory, S., Keyser, C., Crubezy, E. & Ludes, B. (2007) STR typing of ancient DNA extracted from hair shafts of Siberian mummies. *Forensic Sci Int* 166, 218-29.

Argant, A. (1989) Carnivores quaternaires de Bourgogne, Université Claude BERNARD, Lyon, Thèse de Doctorat, 338 p.

Arnason, U., Gullberg, A., Janke, A. & Kullberg, M. (2007) Mitogenomic analyses of caniform relationships. *Mol Phyl Evol* 45, 863-74.

Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A. & Thomas, R. H. (1997) Problems of reproducibility--does geologically ancient DNA survive in amber-preserved insects? *Proc Biol Sci* 264, 467-74.

Axelsson, E., Ratnakumar, A., Arendt, M. L., Magbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, A. & Lindblad-Toh, K. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360-4.

Baffier, D. (2005) La Grotte Chauvet : conservation d'un patrimoine. *Bulletin de la société préhistorique française* 102(1):11-16.

Baranovska, I., Jäderlund, K. H., Nennesmo, I., Holmqvist, E., Heidrich, N., Larsson, N. G., Andersson, G., Wagner, E. G., Hedhammar, A., Wibom, R. & Andersson, L. (2009) Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNA<sup>Tyr</sup> gene. *PLoS Genet.* 5, E1000499.

Barnett, R., Shapiro, B., Barnes, I., Ho, S. Y., Burger, J., Yamaguchi, N., Higham, T. F., Wheeler, H. T., Rosendahl, W., Sher, A. V., Sotnikova, M., Kuznetsova, T., Baryshnikov, G. F., Martin, L. D., Harington, C. R., Burns, J. A. & Cooper, A. (2009) Phylogeography of lions (*Panthera leo ssp.*) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Mol Ecol* 18, 1668-77.

Baryshnikov, G. (1999) Chronological and geographical variability of *Crocota spelaea* (Carnivora, Hyenidae) from the Pleistocene of Russia. *Mammoths and the mammoth fauna: Studies of an extinct ecosystem*, G. Haynes, J. Klimovicz & J. W. F. Reumers (Natural

History Museum, Rotterdam).

Bégouën R, Clottes J, Feruglio V, Pastoors A. (2014) *La caverne des Trois-Frères*. Paris: Somogy éditions d'art.

Bengtsson, C. F., Olsen, M. E., Brandt, L. O., Bertelsen, M. F., Willerslev, E., Tobin, D. J., Wilson, A. S. & Gilbert, M. T. (2011) DNA from keratinous tissue. Part I: hair and nail. *Ann Anat* 194, 17-25.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Echin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53-9.

Binladen, J., Wiuf, C., Gilbert, M. T., Bunce, M., Barnett, R., Larson, G., Greenwood, A. D., Haile, J., Ho, S. Y., Hansen, A. J. & Willerslev, E. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* 172, 733-41.

Björnerfledt, S., Webster, M. T. & Vilà, C. (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 16, 990-4.

Bocherens, H., Drucker, D. G., Billiou, D., Geneste, J. M. & van der Plicht, J. (2006) Bears and humans in Chauvet Cave (Vallon-Pont-d'Arc, Ardeche, France): insights from stable isotopes and radiocarbon dating of bone collagen. *J Hum Evol* 50, 370-6.

Bon, C., Berthonaud, V., Maksud, F., Labadie, K., Poulain, J., Artiguenave, F., Wincker, P., Aury, J. M. & Elalouf, J. M. (2012) Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* 279, 2825-30.

Bon, C., Berthonaud, V., Fosse, P., Gély, B., Maksud, F., Vitalis, R., Philippe, M., van der Plicht, J. & Elalouf, J.-M. (2011) Low regional diversity of late cave bears mitochondrial DNA at the time of Chauvet Aurignacian paintings. *J Archeol Sci* 38:1886-95.

Bon, C., Caudy, N., de Dieuleveult, M., Fosse, P., Philippe, M., Maksud, F., Beraud-Colomb, E., Bouzaid, E., Kefi, R., Laugier, C., Rousseau, B., Casane, D., van der Plicht, J. & Elalouf, J. M. (2008) Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proc Natl Acad Sci U S A* 105, 17447-52.

Bonifay, M.-F. (1971) Carnivores quaternaires du Sud-Est de la France, Paris, Tome XXI, Fas. 2, Edition du Muséum, 377 p.

Boudadi-Maligne, M. (2012) Une nouvelle sous-espèce de loup (*Canis lupus maximus nov. subsp.*) dans le Pléistocène supérieur d'Europe occidentale. *C.R. Palevol* 11(7): 475-484.

Boudadi-Maligne M. (2010) *Les Canis pléistocènes du Sud de la France : approche biosystématique, évolutive et biochronologique*. Bordeaux : Université Bordeaux 1, 451 p., Thèse de doctorat.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. & Paabo, S. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104, 14616-21.

Brugal, J. P. & Croitor. R. (2007) Evolution, ecology and biochronology of herbivore associations in Europe during the last 3 million years (Colloque International Q5: Le Quaternaire, Limites et Spécificités, 1-3 Fev. 07). *Quaternaire*, 18(2):129-152.

Brunnel, E., Chailloux, D., Chauvet, J.-M., Dugas, A., Hillaire, C., Raimbault, M., Renda, M. & Terres, S. (2008) La grotte aux Points (commune d'Aiguèze, Gard). *Ardèche Archeol.* 28:23-28.

Buckland, W. (1829) On the discovery of coprolites, or fossil feces, in the lias at Lyme Regis, and in other formations. *Geol Soc Lond, Transaction series 2* 3:223-36.

Candille, S. I., Kaelin, C. B., Cattanaach, B. M., Yu, B., Thompson, D. A., Nix, M. A., Kerns, J. A., Schmutz, S. M., Millhauser, G. L. & Barsh, G. S. (2007) A  $\beta$ -defensin mutation causes black coat color in domestic dogs. *Science* 318, 1418-23.

Cano, R. J., Borucki, M. K., Higby-Schweitzer, M., Poinar, H. N., Poinar, G. O., Jr. & Pollard, K. J. (1994) Bacillus DNA in fossil bees: an ancient symbiosis? *Appl Environ Microbiol* 60, 2164-7.

Clottes, J. (2001) *La grotte Chauvet : l'art des origines* (Seuil, Paris).

Clottes, J. & Lewis-William, J. D. (1996) *Les chamanes de la préhistoire : transe et magie dans les grottes ornées* (Seuil, Paris).

Cooper, A. & Poinar, H. N. (2000) Ancient DNA: do it right or not at all. *Science* 289,1139.

Cuzange, M. T. Delqué-Kolic, E., Goslar, T., Meiert Grootes, P., Higham, T., Kaltnecker, E., Nadeau, M.-J., Oberlin, C., Paterne, M., van der Plicht, J., Bonk Ramsey, C., Valladas, H., Clottes, J., Geneste, J.-M. (2007) Radiocarbon intercomparison program for Chauvet Cave. *Radiocarbon* 49(2):339-347.

d'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., Woodgate, R., Paabo, S. & Holliger, P. (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat Biotechnol* 25, 939-43.

Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Grocke, D. R., Matheus, P., Zazula, G., Guthrie, D., Froese, D., Buigues, B., de Marliave, C., Flemming, C., Poinar, D., Fisher, D., Southon, J., Tikhonov, A. N., MacPhee, R. D. & Poinar, H. N. (2008) Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr Biol* 18, 1320-6.

Defleur, A. (1995) Nouvelles découvertes de restes humains Moustériens dans les dépôts de la Baume-Guercy (Soyons, Ardèche). *Bull. et Mém. De la Soc. D'anthropologie de Paris*, n.s., 7, 3-4:185-190.

Defleur, A., Dutour, O. & Vandermeersch, B. (1992) Etude de deux dents humaines provenant des niveaux moustériens de la Baume Néron (Soyons, Ardèche). *Bull. et Mém. De la Soc. D'anthropologie de Paris*, n.s., 4, 1-2:127-134.

Delpozzo, G. & Guardiola, J. (1989) Mummy DNA Fragment Identified. *Nature* 339(6224):431-432

DeSalle, R., Gatesy, J., Wheeler, W. & Grimaldi, D. (1992) DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257, 1933-6.

Ding, Z. L., Oskarsson, M., Ardalán, A., Angleby, H., Dahlgren, L. G., Tepeli, C., Kirkness, E., Savolainen, P. & Zhang, Y. P. (2012) Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* (Edinb) 108, 507-14.

Discamps, E. (2011) Hommes et hyènes face aux recompositions des communautés d'Ongulés (MIS 5-3) : Éléments pour un cadre paléoécologique des sociétés du Paléolithique moyen et supérieur ancien d'Europe de l'Ouest. *Archaeology and Prehistory*. Université Sciences et Technologies – Bordeaux I, 2011.

Douglas, K. C., Halbert, N. D., Kolenda, C., Childers, C., Hunter, D. L. & Derr, J. N. (2011) Complete mitochondrial DNA sequence analysis of *Bison bison* and bison-cattle hybrids : function and phylogeny. *Mitochondrion*, 11:166-75.

Edwards, C. J., Magee, D. A., Park, S. D., McGettigan, P. A., Lohan, A. J., Murphy, A., Finlay, E. K., Shapiro, B., Chamberlain, A. T., Richards, M. B., Bradley, D. G., Loftus, B. J. & MacHugh, D. E. (2010) A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS One* 5, e9255.

Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M. B. & Rollo, F. (2008) Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18, 1687-93.

Fosse, P. (1997) Variabilité des assemblages osseux créés par l'hyène des cavernes. *Paléo* 9:15-54.

Fosse, P. & Philippe, M. (2005) La faune de la grotte Chauvet : paléobiologie et anthropozoologie. *Bulletin de la société préhistorique française* 102(1):82-102.

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., Lee, C., Tadiogla, V., Wilton, A., Siepel, A., Bustamante, C. D., Harkins, T. T., Nelson, S. F., Ostrander, E. A., Marques-Bonet, T., Wayne, R. K. & Novembre, J. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10, e1004016.

Garcia, M. A. (2005) Ichnologie générale de la grotte Chauvet. *Bulletin de la société préhistorique française* 102(1):103-108.

Gardien, V., Lécuyer, C. & Derym, F. (2009) Reconstitution de l'évolution du climat recent (- 120 000 ans à nos jours) dans la vallée du Rhône à partir de l'étude du site de Soyons. *Ardèche Archéologie* 26, 27-29.

Garrido, G., & Arribas, A. (2008) *Canis accitanus nov. sp.*, a new small dog (Canidae, Carnivora, Mammalia) from the Fonelas P-1 Plio-Pleistocene site (Guadix, Grenade, Espagne). *GEOBIOS*, 41:751-761.



Gély, B. & Porte, J.-L. (1996) Les gravures paléolithiques de la grotte des Deux- Ouvertures à Saint-Martin-d'Ardèche. Bulletin de la Société Préhistorique Ariège-Pyrénées LI:81-98.

Gilbert, M. T., Drautz, D. I., Lesk, A. M., Ho, S. Y., Qi, J., Ratan, A., Hsu, C. H., Sher, A., Dalen, L., Gotherstrom, A., Tomsho, L. P., Rendulic, S., Packard, M., Campos, P. F., Kuznetsova, T. V., Shidlovskiy, F., Tikhonov, A., Willerslev, E., Iacumin, P., Buigues, B., Ericson, P. G., Germonpre, M., Kosintsev, P., Nikolaev, V., Nowak-Kemp, M., Knight, J. R., Irzyk, G. P., Perbost, C. S., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Miller, W. & Schuster, S. C. (2008a) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc Natl Acad Sci U S A 105, 8327-32.

Gilbert, M. T., Kivisild, T., Gronnow, B., Andersen, P. K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P. F., Rasmussen, M., Metspalu, M., Higham, T. F., Schwenninger, J. L., Nathan, R., De Hoog, C. J., Koch, A., Moller, L. N., Andreasen, C., Meldgaard, M., Villems, R., Bendixen, C. & Willerslev, E. (2008b) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320, 1787-9.

Gilbert, M. T., Jenkins, D. L., Gotherstrom, A., Naveran, N., Sanchez, J. J., Hofreiter, M., Thomsen, P. F., Binladen, J., Higham, T. F., Yohe, R. M., 2nd, Parr, R., Cummings, L. S. & Willerslev, E. (2008c) DNA from pre-Clovis human coprolites in Oregon, North America. Science 320, 786-9.

Gilbert, M. T., Binladen, J., Miller, W., Wiuf, C., Willerslev, E., Poinar, H., Carlson, J. E., Leebens-Mack, J. H. & Schuster, S. C. (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. Nucleic Acids Res 35, 1-10.

Gilbert, M. T., Wilson, A. S., Bunce, M., Hansen, A. J., Willerslev, E., Shapiro, B., Higham, T. F., Richards, M. P., O'Connell, T. C., Tobin, D. J., Janaway, R. C. & Cooper, A. (2004) Ancient mitochondrial DNA from hair. Curr Biol 14, R463-4.

Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D. & Zurawski, G. (1990) Chloroplast DNA sequence from a miocene Magnolia species. Nature 344, 656-8.

Götherström, A., Collins, M. J., Angerbjörn, A. & Liden, K. (2002) Bone preservation and DNA amplification. Archaeometry 44:395-404.

Gray, M. M., Sutter, N. B., Ostrander, E. A. & Wayne, R. K. (2010) The IGF1small dog haplotype is derived from Middle Eastern grey wolves. BMC Biol 8,16.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspina, A. S.,

Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. & Paabo, S. (2010) A draft sequence of the Neandertal genome. *Science* 328, 710-22.

Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prufer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Wikstrom, M., Laakkonen, L., Kelso, J., Slatkin, M. & Paabo, S. (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134, 416-26.

Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., Du, L., Egholm, M., Rothberg, J. M., Paunovic, M. & Paabo, S. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330-6.

Guérin, C. & Philippe, M. (1971) Les gisements de vertébrés pléistocènes du Causse de Martel. *Bull. Soc. hist. archéol.* 93, 31-46

Hagelberg, E., Thomas, M. G., Cook, C. E., Jr., Sher, A. V., Baryshnikov, G. F. & Lister, A. M. (1994) DNA from ancient mammoth bones. *Nature* 370, 333-4.

Hanni, C., Laudet, V., Stehelin, D. & Taberlet, P. (1994) Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequencing. *Proc Natl Acad Sci U S A* 91, 12336-40.

Heyn, P., Stenzel, U., Briggs, A. W., Kircher, M., Hofreiter, M. & Meyer, M. (2010) Road blocks on paleogenomes--polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res* 38, e161.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282-4.

Hochmeister, M. N., Budowle, B., Borer, U. V., Eggman, U., Comey, C. T. & Dirnhofer, R. (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. *Journal of Forensic Sciences* 36, 1649-61

Hofreiter, M., Serre, D., Rohland, N., Rabeder, G., Nagel, D., Conard, N., Munzel, S. & Paabo, S. (2004) Lack of phylogeography in European mammals before the last glaciation.

Proc Natl Acad Sci U S A 101, 12963-8.

Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Paabo, S. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29, 4793-9.

Hofstetter, J. R., Zhang, A., Mayeda, A. R., Guscar, T., Nurnberger J. I. Jr. & Lahiri, D. K. (1997) Genomic DNA from mice : A comparison of recovery methods and tissues sources. *Biochemical and Molecular Medicine* 62, 197-202

Hoss, M., Paabo, S. & Vereshchagin, N. K. (1994) Mammoth DNA sequences. *Nature* 370, 333.

Hughes, S., Hayden, T. J., Douady, C. J., Tougaard, C., Germonpre, M., Stuart, A., Lbova, L., Carden, R. F., Hanni, C. & Say, L. (2006) Molecular phylogeny of the extinct giant deer, *Megaloceros giganteus*. *Mol Phyl Evol* 40, 285-91.

Huynen, L., Gill, B. J., Millar, C. D. & Lambert, D. M. (2010) Ancient DNA reveals extreme egg morphology and nesting behavior in New Zealand's extinct moa. *Proc Natl Acad Sci U S A* 107, 16201-6.

Ishiguro, N., Inoshima, Y. & Shigehara, N. (2009) Mitochondrial DNA analysis of the Japanese wolf (*Canis lupus hodophilax* Temminck, 1839) and comparison with representative wolf and domestic dog haplotypes. *Zool. Sci.* 26, 765-70.

Joannin, S. (2002) Implications paléo-écologiques et biostratigraphiques des ruminants du gisement pléistocène supérieur de Siréjol (Lot). Mémoire de maîtrise Sciences de la Terre et de l'Univers, Université Claude Bernard-Lyon I.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. (2013) mapDamage2.0 : fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682-4.

Kim, K.S., Lee, S.E., Jeong, H.W. & Ha, J.H. (1998) The complete nucleotide sequence of the domestic dog (*Canis familiaris*) mitochondrial genome. *Mol. Phylogenet. Evol.* 10, 210-220.

King, C., Debruyne, R., Kuch, M., Schwarz, C. & Poinar, H. (2009) A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. *Biotechniques* 47, 941-9.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. & Pääbo, S. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30.

Kurtén, B. (1968) *Pleistocene Mammals of Europe* (Aldine Pub. Co., Chicago)

Kurtén, B. (1974) *A history of coyote-like dogs (Canidae, Mammalia)*. *Acta Zoologica*

Fennica, 140:1-38.

Koufos, G. D. & Kostopoulos, D. S. 1997. New carnivore material from the Plio-Pleistocene of Macedonia (Greece) with the description of a new canid. *Münchner Geowiss*, 34:33-63.

Landry, J.-M. (2006) *Le loup. Biologie, Mœurs, Mythologie, Cohabitation, Protection*, Delachaux et Niestlé, 240 p.

Leroi-Gourhan, A. (1966) L'analyse pollinique des coprolithes. *Bulletin de la Société Préhistorique française*, 63(5):153-54.

Li, Y., Li, Q., Zhao, X., Xie, Z. & Xu, Y. (2011) Complete sequence of the Tibetan Mastiff mitochondrial genome and its phylogenetic relationship with other Canids (*Canis*, Canidae). *Animal* 5, 18-25.

Lindhal, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715

Loreille, O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P. & Hanni, C. (2001) Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr Biol* 11, 200-3.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K., Froese, D., Zazula, G., Stafford, T. W., Aaris-Sørensen, K., Batra, P. & Haywood, A. M. (2011) Species-specific responses of late quaternary megafauna to climate and humans. *Nature* 479:359-64.

MacHugh, D. E., Edwards, C. J., Bailey, J. F., Bancroft, D. R. & Bradley, D. G. (2000) The extraction and analysis of ancient DNA from bone and teeth : a survey of current methodologies. *Anc Biomol* 3:81-103.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P.,

Begley, R. F. & Rothberg, J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-80.

Markova, A. K., Smirnov, N. G., Kozharinov, A. V., Kazantseva, N. E., Simakova, A. & Kitaev, L. M. (1995) Late Pleistocene distribution and diversity of mammals in northern Eurasia (PALEOFAUNA Database). *Paleotologia I Evolucio* 28-29:5-134.

Marsolier-Kergoat, M.-C., Palacio, P., Berthonaud, V., Maksud, F., Stafford, T., Bégouën, R. & Elalouf, J.-M. (2015) Hunting the extinct steppe bison (*Bison priscus*) mitochondrial genome in the Trois-Frères paleolithic painted cave. *PLoS One*, e0128267

Martin, R. (1973) Trois nouvelles espèces de Caninae (Canidae, Carnivora) des gisements Plio-Villafranchiens d'Europe. *Documents des Laboratoires de Géologie de Lyon*, 57:87-96.

Martin, P. S. & Sharock, F. W. (1964) Pollen Analysis of Prehistoric Human Feces : a New Approach to Ethnobotany. *American Antiquity* 30:169-80

Melchior, L., Lynnerup, N., Siegismund, H. R., Kivisild, T. & Dissing, J. (2010) Genetic diversity among ancient Nordic populations. *PLoS One* 5, e11898.

Meng, C., Zhang, H. & Chen, Y. (2009) Sequencing and Analysis of Mitochondrial Genome of Chinese Grey Wolf (*Canis lupus chanco*). *Chin. J. Biochem. Mol. Biol.* 24, 1170-1176.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E. S., Knight, J. R., Irzyk, G. P., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Pringle, T. & Schuster, S. C. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456, 387-90.

Mourer-Chauviré, C. (1972) Etude de nouveaux restes de vertébrés provenant de la carrière Fournier à Châtillon-Saint-Jean (Drôme) III – Artiodactyles, chevaux et oiseaux. *Bull. Ass. fr. Et. quat.* 1972-4, 271-305

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1, 263-73.

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Paabo, S., Pritchard, J. K. & Rubin, E. M. (2006) Sequencing and analysis of Neanderthal genomic DNA. *Science* 314, 1113-8.

Noonan, J. P., Hofreiter, M., Smith, D., Priest, J. R., Rohland, N., Rabeder, G., Krause, J., Deter, J. C., Paabo, S. & Rubin, E. M. (2005) Genomic sequencing of Pleistocene cave bears. *Science* 309, 597-9.

Noro, M., Masuda, R., Dubrovo, I. A., Yoshida, M. C. & Kato, M. (1998) Molecular phylogenetic inference of the woolly mammoth *Mammuthus primigenius*, based on complete sequences of mitochondrial cytochrome b and 12S ribosomal RNA genes. *J Mol Evol* 46, 314-26.

Nowak, Z. & Olech, W. (2008) Verification of phylogenetic hypothesis concerning the evolution of genus *Bison*. *Annals of Warsaw University of Life Sciences – SGGV, Animal Science* 45, 65-72.

Olivieri, C., Ermini, L., Rizzi, E., Corti, G., Bonnal, R., Luciani, S., Marota, I., De Bellis, G. & Rollo, F. (2010) Characterization of nucleotide misincorporation patterns in the iceman's mitochondrial DNA. *PLoS One* 5, e8629.

Ollivier, M., Tresset, A., Hitte, C., Petit, C., Hughes, S., Gillet, B., Duffraisse, M., Pionnier-Capitan, M., Lagoutte, L., Arbogast, R. M., Balasescu, A., Boroneant, A., Mashkour, M., Vigne, J.-D. & Hänni, C. (2013) Evidence of coat color variation sheds new light on ancient canids. *PLoS One* 8, e75110.

Olsen, M. E., Bengtsson, C. F., Bertelsen, M. F., Willerslev, E. & Gilbert, M. T. (2011) DNA from keratinous tissue. Part II: feather. *Ann Anat* 194, 31-5.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliusen, T., Malaspina, A. S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A. M., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G. D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J. F., Weinstock, J., Gregersen, K., Røed, K. H., Eisenmann, V., Rubin, C. J., Miller, D. C., Antczak, D. F., Bertelsen, M. F., Brunak, S., Al-Rasheid, K. A., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M. T., Kjær, K., Sicheritz-Ponten, T., Jensen, L. J., Olsen, J. V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J. & Willerslev, E. (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74-8.

Oskam, C. L., Haile, J., McLay, E., Rigby, P., Allentoft, M. E., Olsen, M. E., Bengtsson, C., Miller, G. H., Schwenninger, J. L., Jacomb, C., Walter, R., Baynes, A., Dortch, J., Parker-Pearson, M., Gilbert, M. T., Holdaway, R. N., Willerslev, E. & Bunce, M. (2010) Fossil avian eggshell preserves ancient DNA. *Proc Biol Sci* 277, 1991-2000.

Pääbo, S. (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314, 644-5.

Pääbo, S., Gifford, J. A. & Wilson, A. C. (1988) Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res* 16, 9775-87.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. & Hofreiter, M. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38, 645-79.

Pang, J. F., Kluetsch, C., Zou, X. J., Zhang, A. B., Luo, L. Y., Angleby, H., Ardalan, A., Ekstrom, C., Skollermo, A., Lundeberg, J., Matsumura, S., Leitner, T., Zhang, Y. P. & Savolainen, P. (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol* 26, 2849-64.

Patou, M. (1984) Contribution à l'étude des Mammifères des couches supérieures de la grotte du Lazaret (Nice, A.-M.). Méthodes et résultats paléontologiques et paléethnographiques. Université Pierre et Marie Curie - Thèse de Doctorat, Paris 6, Muséum National d'Histoire Naturelle, Paris.

Philippe, M. (2009) Découverte de vestiges osseux humains dans le secteur orné paléolithique de la grotte des Deux-Ouvertures. *Ardèche Archeol.* 26:13-14.

Philippe, M. (2005) La grotte Chauvet et les autres « grottes à ours » de ses environs immédiats : Vallée de l'Ibie, amont des Gorges de l'Ardèche et cirque d'Estre. *Ardèche Archéologie* 22:8-19.

Philippe, M. (1997) La faune würmienne du gisement paléontologique de Siréjol à Gignac (Lot). *Bulletin de la Société Scientifique, Historique et Archéologique de la Corrèze.*

Philippe, M., Mourer-Chauviré, C. & Evin, J. (1980) Les gisements paléontologiques quaternaires des Causses de Martel et de Gramat (Corrèze et Lot) : faunes et chronologie. *Nouv. Arch. Mus. Hist. nat. Lyon*, fasc. 18, 57-65.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W. & Schuster, S. C. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311, 392-4.

Poinar, H., Kuch, M., McDonald, G., Martin, P. & Paabo, S. (2003) Nuclear gene sequences from a late pleistocene sloth coprolite. *Curr Biol* 13, 1150-2.

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G. & Paabo, S. (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281, 402-6.

Poinar, H. N., Cano, R. J. & Poinar, G. O. (1993) DNA from extinct plant. *Nature* 363-677.

Prucça, A. (2003) Caractérisation de l'impact des Loups sur des ossements d'Herbivores (Cerfs de Virginie, Orignaux, Bisons) : Etude des modifications infligées par des Loups captifs et sauvages Nord-Américains. Mémoire de DEA 2003- Université de Provence, 135p.

Ramirez, O., Gigli, E., Bover, P., Alcover, J. A., Bertranpetit, J., Castresana, J. & Lalueza-Fox, C. (2009) Paleogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine. *PLoS One* 4, e5670.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Gronnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F., Ramsey, C. B., Hansen, T. V., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Ponten, T., Villems, R., Nielsen, R., Krogh, A., Wang, J. & Willerslev, E. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757-62.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eicher, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. & Pääbo, S. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053-60.

Rogaev, E. I., Moliaka, Y. K., Malyarchuk, B. A., Kondrashov, F. A., Derenko, M. V., Chumakov, I. & Grigorenko, A. P. (2006) Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biol* 4, e73.

Rohland, N., Pollack, J. L., Nagel, D., Beauval, C., Airvaux, J., Paabo, S. & Hofreiter, M. (2005) The population history of extant and extinct hyenas. *Mol Biol Evol* 22, 2435-43.

Ronquist, F. & Huelsenbeck, J. P. (2003) MrBayes 3 : Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-4.

Rook, L. (1993) I cani dell'Eurasia dal Miocene superiore al Pleistocene medio, Università Modena, Bologna, Firenze, Thèse de doctorat, Roma "La Sapienza", 131 p.

Sadier, B., Delannoy, J. J., Benedetti, L., Bourles, D. L., Jaillet, S., Geneste, J. M.,



Lebatard, A. E. & Arnold, M. (2012) Further constraints on the Chauvet cave artwork elaboration. *Proc Natl Acad Sci U S A* 109, 8002-6.

Shapiro, B. (2008) Engineered polymerases amplify the potential of ancient DNA. *Trends Biotechnol* 26, 285-7.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S. & Cooper, A. (2004) Rise and fall of the Beringian steppe bison. *Science* 306, 1561-5.

Sheng, G. L., Soubrier, J., Liu, J. Y., Werdelin, L., Llamas, B., Thomson, V. A., Tuke, J., Wu, L. J., Hou, X. D., Chen, Q. J., Lai, X. L. & Cooper, A. (2013) Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Mol Ecol* 23, 522-33.

Sher, A. V. (1997) An early Quaternary Bison population from Untermassfeld : *Bison menneri* sp. nov. 40, 101-180.

Sillero-Zubiri, C., Hoffmann, M., MACDONALD, D. W. (2004) *Canids : Foxes, Wolves, Jackals and Dogs. Status Survey and Conservation Action Plan.* IUCN/SSC Canid Specialist Group, Gland, Switzerland and Cambridge, UK, 430 p.

Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. (2015) Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr Biol* 25, 1515-9.

Sotnikova, M. V. & Rook, L. (2010) Dispersal of the Canini (Mammalia, Canidae : Caninae) across Eurasia during the late Miocene to early Pleistocene. *Quaternary International*, 212:86-97.

Stankiewicz, B. A., Poinar, H. N., Briggs, D. E. G., Evershed, R. P. & Poinar, G. O. (1998) Chemical preservation of plants and insects in natural resins. *Proc. R. Soc. Lond. B* 265(1397):641-647.

Stiller, M., Baryshnikov, G., Bocherens, H., Grandal d'Anglade, A., Hilpert, B., Munzel, S. C., Pinhasi, R., Rabeder, G., Rosendahl, W., Trinkaus, E., Hofreiter, M. & Knapp, M. (2010) Withering away--25,000 years of genetic decline preceded cave bear extinction. *Mol Biol Evol* 27, 975-8.

Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. (2013) MEGA6 : Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30,2725-29.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpman, H-P., Loponte, D. M., Acosta, A. A., Giemsch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., Graphodatsky, A. S., Ovodov, N. D., Wahlberg, N., Freedman, A. H., Schweizer, R. M., Koepfli, K.-P., Leonard, J. A., Meyer, M., Krause, J., Pääbo, S., Green, R. E., Wayne, R. K. (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342, 871-4.

Théry-Parisot, I. & Thielbault, S. (2005) Le pin (*Pinus sylvestris*) : préférence d'un taxon ou contrainte de l'environnement ? Etude des charbons de bois de la Grotte Chauvet. *Bulletin de la société préhistorique française* 102(1):69-76.

Thomas, R. H., Schaffner, W., Wilson, A. C. & Paabo, S. (1989) DNA phylogeny of the extinct marsupial wolf. *Nature* 340, 465-7.

Thompson, J. F. & Steinmann, K. E. (2010) Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol* Chap7, Unit7.10.

Torre, D. (1967) I cani Villafranchiani della Toscana. *Palaeontographia italica*, LXIII:113-136.

Torre, D. (1974) Affinitá dentali del cane della grotta di l'Escale. *Rivista Italiana di Paleontología e Stratigrafia*, 80(1):147-156.

Tosello, G. & Fritz, C. (2005) Les dessins noirs de la Grotte Chauvet-Pont d'Arc : essai sur leur originalité dans le site et leur place dans l'art aurignien. *Bulletin de la société préhistorique française* 102(1):159-172.

Turner, A. (1984) The interpretation of variation in fossil specimens of spotted hyena (*Crocuta crocuta* Erxleben, 1777) from Sterkfontein valley sites (mammalian carnivora). *Annals of the Transvaal museum* 33:339-418.

Valdiosera, C. E., Garcia-Garitagoitia, J. L., Garcia, N., Doadrio, I., Thomas, M. G., Hanni, C., Arsuaga, J. L., Barnes, I., Hofreiter, M., Orlando, L. & Gotherstrom, A. (2008) Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proc Natl Acad Sci U S A* 105, 5123-8.

Valdiosera, C. E., Garcia, N., Anderung, C., Dalen, L., Cregut-Bonnouere, E., Kahlke, R. D., Stiller, M., Brandstrom, M., Thomas, M. G., Arsuaga, J. L., Gotherstrom, A. & Barnes, I. (2007) Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Mol Ecol* 16, 5140-8.

Valensi, P. (1994) Les grands mammifères de la grotte du Lazaret, Nice. Etude paléontologique et biostratigraphique des carnivores. *Archéozoologie des grandes faunes*.

Muséum National d'histoire Naturelle à l'Institut de Paléontologie Humaine, Quaternaire :Géologie, Paléontologie humaine, Préhistoire, Paris, Thèse de Doctorat.

Valladas, H., Clottes, J., Geneste, J.-M., Garcia, M.A., Arnold, M., Cachier, H. & Tisnérat-Laborde, N. (2001) Palaeolithic paintings. Evolution of prehistoric cave art. *Nature* 413(6855):479.

Vartanyan, S. L., Garutt, V. E. & Sher A. V. (1993) Holocene Dwarf Mammoths from Wrangel-Island in the Siberian Artic. *Nature* 362(6418):337-340.

Vasan, S., Zhang, X., Kapurniotu, A., Bernhagen, J., Teichberg, S., Basgen, J., Wagle, D., Shih, D., Terlecky, I., Bucala, R., Cerami, A., Egan, J. & Ulrich, P. (1996) An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature* 382, 275-8.

Verkaar, E. L. C., Nijman, I. J., Beeke, M., Hanekamp, E. & Lenstra, J. A. (2004) Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin ? *Mol Biol Evol*, 21:1165-70.

Veyrier, M., Beaux, E.& Combier, J. (1951) Grotte de Néron à Soyons (Ardèche). Les fouilles de 1950, leurs enseignements. *Bulletin de la société préhistorique française*, 48(1-2):70-78.

Wall, J. D. & Kim, S. K. (2007) Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet* 3, 1862-6.

Wang, X., Tedford, R.H. & Taylor, B.E. (1999) Phylogenetic systematics of the Borophaginae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History* 243:1–391.

Wang, X. (1994) Phylogenetic systematics of the Hesperocyoninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History* 221:1–207.

Ward, R. & Stringer, C. (1997) A molecular handle on the Neanderthals. *Nature* 388, 225-6.

Wayne, R. K., Leonard, J. A. & Cooper, A. (1999) Full of sound and fury : The recent history of ancient DNA. *Annual Review of Ecology and Systematics* 30:457-477.

Webb, K.M. & Allard, M.W. (2009) Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region. *J. Forensic Sci.* 54, 275-88.

Werdelin, L. & Solounias, N. (1991) The hyenidae: taxonomy systematics and evolution. *Fossils and Strata* 30:1-104.

Willerslev, E., Gilbert, M. T., Binladen, J., Ho, S. Y., Campos, P. F., Ratan, A., Tomsho, L. P., da Fonseca, R. R., Sher, A., Kuznetsova, T. V., Nowak-Kemp, M., Roth, T.

L., Miller, W. & Schuster, S. C. (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evol Biol* 9, 95.

Willerslev, E. & Cooper, A. (2005) Ancient DNA. *Proc Biol Sci* 272, 3-16.

Willerslev, E., Hansen, A. J. & Poinar, H. N. (2004a) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 19, 141-7.

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A. & Cooper, A. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791-5.

Zazula, G. D., MacKay, G., Andrews, T. D., Shapiro, B., Letts, B. & Brock, F. (2009) A late Pleistocene steppe bison (*Bison priscus*) partial carcass from Tsiigehtchic northwest territories, Canada. *Quaternary Science Review*, 28:2734-42

Zischler, H., Hoss, M., Handt, O., von Haeseler, A., van der Kuyl, A. C. & Goudsmit, J. (1995) Detecting dinosaur DNA. *Science* 268, 1192-3; author reply 1194.



## *Annexes*

RESEARCH ARTICLE

# Hunting the Extinct Steppe Bison (*Bison priscus*) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave

Marie-Claude Marsolier-Kergoat<sup>1,2</sup>, Pauline Palacio<sup>1,2</sup>, Véronique Berthouaud<sup>1</sup>, Frédéric Maksud<sup>3</sup>, Thomas Stafford<sup>4</sup>, Robert Bégouën<sup>5</sup>, Jean-Marc Elalouf<sup>1,2\*</sup>

**1** iBiTec-S/SBiGeM, CEA/Saclay 91191 Gif-sur-Yvette cedex, France, **2** CNRS-UMR7206, Eco-anthropologie et Ethnobiologie, Département Hommes, Natures et Sociétés, Musée de l'Homme, 17 place du Trocadéro, 75016 Paris, France, **3** Service Régional de l'Archéologie, 32 rue de la Dalbade, BP811 31080 Toulouse cedex 6, France, **4** AMS 14C Dating Centre, Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, Aarhus, Denmark, **5** Association Louis Bégouën, Laboratoire de Préhistoire de Pujol, 09200 Montesquieu-Avantès, France

\* [jean-marc.elalouf@cea.fr](mailto:jean-marc.elalouf@cea.fr)



OPEN ACCESS

**Citation:** Marsolier-Kergoat M-C, Palacio P, Berthouaud V, Maksud F, Stafford T, Bégouën R, et al. (2015) Hunting the Extinct Steppe Bison (*Bison priscus*) Mitochondrial Genome in the Trois-Frères Paleolithic Painted Cave. PLoS ONE 10(6): e0128267. doi:10.1371/journal.pone.0128267

**Academic Editor:** Ludovic Orlando, Natural History Museum of Denmark, University of Copenhagen, DENMARK

**Received:** December 10, 2014

**Accepted:** April 23, 2015

**Published:** June 17, 2015

**Copyright:** © 2015 Marsolier-Kergoat et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The DNA reads of this study have been deposited at EBI under accession number ERP006376. The mitochondrial genome sequence of the *Bison priscus* specimen has been registered in the GenBank database under accession number KM593920.

**Funding:** This work was supported by CEA Saclay, LOC1043ELALO, to Dr. Jean-Marc Elalouf. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Despite the abundance of fossil remains for the extinct steppe bison (*Bison priscus*), an animal that was painted and engraved in numerous European Paleolithic caves, a complete mitochondrial genome sequence has never been obtained for this species. In the present study we collected bone samples from a sector of the Trois-Frères Paleolithic cave (Ariège, France) that formerly functioned as a pitfall and was sealed before the end of the Pleistocene. Screening the DNA content of the samples collected from the ground surface revealed their contamination by *Bos* DNA. However, a 19,000-year-old rib collected on a rock apart the pathway delineated for modern visitors was devoid of such contaminants and reproducibly yielded *Bison priscus* DNA. High-throughput shotgun sequencing combined with conventional PCR analysis of the rib DNA extract enabled to reconstruct a complete mitochondrial genome sequence of 16,318 bp for the extinct steppe bison with a 10.4-fold coverage. Phylogenetic analyses robustly established the position of the *Bison priscus* mitochondrial genome as basal to the clade delineated by the genomes of the modern American *Bison bison*. The extinct steppe bison sequence, which exhibits 93 specific polymorphisms as compared to the published *Bison bison* mitochondrial genomes, provides an additional resource for the study of *Bovinae* specimens. Moreover this study of ancient DNA delineates a new research pathway for the analysis of the Magdalenian Trois-Frères cave.

## Introduction

The earliest members of the genus *Bison* appeared at the beginning of the Pleistocene in India and China. *Bison* then spread from Asia to Europe and America. Archaeological evidence shows that two bison species existed in Europe from the Middle Pleistocene on, the steppe bison and the woodland bison. The steppe bison, *Bison priscus* (Bojanus, 1827), is very

**Competing Interests:** The authors have declared that no competing interests exist.

common in Pleistocene deposits but became extinct at the end of the last Ice Age, about 10,000 years ago. It was a formidable animal with long horns and robust legs that may have stood more than 2 meters at the withers and reached a total length of more than 2.7 meters. It occupied cool, steppe-like grasslands. *Bison priscus* had a wide geographic distribution, called the great Pleistocene bison belt, which extended from England to America and from the archipelago of Nova Zembla in the Arctic Ocean in the north of Russia down to Spain, Caucasia and Mexico. It is generally reckoned that it is the steppe bison that was portrayed by paleolithic artists in painted caves like the Altamira cave in northern Spain, the Chauvet, Lascaux and Trois-Frères caves in France. The woodland bison, *Bison schoetensacki* (Freudenberg, 1910), which has left more scarce fossil remains than the steppe bison, appeared in the early Middle Pleistocene. Its body size and horns were smaller than those of the steppe bison and it was probably a forest-dweller. It became extinct at the end of the Pleistocene at the latest [1].

Only two bison species survive today: the American bison and the European bison. The American bison, in North America, includes two subspecies, the plains bison, *Bison bison bison*, and the wood bison, *Bison bison athabascae*. The European bison, *Bison bonasus*, is found in Europe and the Caucasus, where it has been reintroduced after its extinction as a wild species. The phylogenetic relationships between the extant and the extinct species are a matter of debate. A population of *Bison priscus* is believed to have first entered eastern Beringia from Asia during the middle Pleistocene, around 300 to 130 ky ago, and then moved southward into central North America approximately 130 to 75 ky ago [2]. The analysis of mitochondrial sequences of bison specimens from Siberia, China and North America (dated between present time and about 70 ky B.P. [before present]) has suggested that late Pleistocene bison from the Ural mountains to northern China are descendants of one or more dispersals from North America [2]. This study has also indicated gene flow between bison populations in Beringia and central North America between 60 and 25 ky B.P., before ice sheet formation during the Last Glacial Maximum (22 to 18 ky B.P.) hampered north-south faunal exchange [2]. The retreat of ice sheets around 14 ky B.P. created an ice-free corridor through which dispersal between Beringia and North America could occur. However, genetic exchanges between bison in Beringia and in central North America were soon after limited due to the establishment of spruce forest in Alberta and of peatland across western and north-western Canada. As a result, all modern *Bison bison* belong to a clade distinct from Beringian bison [2]. The European bison *Bison bonasus* has a complex descent: its nuclear genome is closely related to that of *Bison bison*, with which it can produce completely fertile hybrid offspring, but its mitochondrial genome is more similar to the genomes of species belonging to the *Bos* genus than to the genome of *Bison bison* [3]. These observations could be explained by lineage sorting implying an early divergence of the American and European bison mitochondrial sequences, or by a scenario according to which the nuclear DNA of an ancient Eurasian *Bos* population would have been changed by the systematic introgression of *Bison* bulls (either *Bison priscus*, *Bison bison* migrating back to Europe, or *Bison schoetensacki*). This ancient hybridization would have been facilitated by the social structure of herd species, in which most males are excluded from reproduction by the dominating bulls, and would have created a new species with a nuclear genome of bison type, but with a *Bos*-like mitochondrial DNA from the maternal ancestor. Interestingly, this scenario would explain the sudden paleontological appearance of *Bison bonasus* at the Preboreal stage (10.3 to 9 ky B.P.) [3].

In an effort to investigate the evolution and demographic history of Pleistocene bison, partial sequences of *Bison priscus* mitochondrial genome have already been obtained [2,4]. The mitochondrial control region, also called the D-loop region, has been sequenced in ~ 300 *Bison priscus* fossils collected in Siberia, China and North America [2,4]. However, these partial sequences are a few hundred bp long, representing less than 5% of the expected 17-kb



mitochondrial genome. The purpose of the present study was to fill the gap by providing a complete mitochondrial genome sequence for the extinct steppe bison.

This work presents the analysis of several bone samples that were collected from the Trois-Frères cave (Ariège, France). This cave (43° 1' N, 1° 12' E) was named after the three sons of Comte Henri Bégouën who discovered its entrance in July 1914. It contains numerous animal representations made famous by the publication of the Abbé Henri Breuil [5], including the drawing and engraving of at least 170 bison. The importance of the bison for the Magdalenian artists who decorated the cave is obvious not only from the fact that its representations outnumber those of any other species, but also from the drawing of a therianthrope figure with a bison upper part. Besides, the cave of the Trois-Frères is part of a single cave-complex with the Tuc d'Audoubert, which contains two modeled clay bison, a unique masterpiece of the Paleolithic period [6]. DNA extracted from one rib bone sample was sequenced using Illumina technology which, in combination with conventional PCR and Sanger sequencing, allowed the assembly of a complete sequence of *Bison priscus* mitochondrial genome.

## Material and Methods

### Collection of archeological samples

All necessary permits were obtained for the described study, which complied with all relevant regulations. The four samples analyzed in this study correspond to specimens SGE1, SGE2, SGE3 and SGE5 of the Trois-Frères cave (Montesquieu-Avantès, 09200, France). They were collected under the supervision of two of us (RB, FM). The samples are stored in the Museum of the Association Louis Bégouën in Montesquieu-Avantès. The permit to collect the samples and perform their analysis was provided by the Direction Régionale des Affaires Culturelles Midi-Pyrénées, Service Régional de l'Archéologie (32, rue de la Dalbade, BP811, 31080 Toulouse Cedex 6, France) where it is registered under the reference FM/MB/2013/13481.

### Contamination issues

To avoid contamination derived from previous and current analyses, initial steps (*i.e.* ancient DNA extraction, generation of the DNA library, and set-up of PCR reactions) were carried out in a dedicated laboratory located in a building where no molecular work has ever been performed on extant DNA. Access to this building was not allowed to persons who had worked on the same day in the modern DNA laboratory and rested on wearing dedicated shoes and laboratory coats. These were changed for a second pair of shoes and a single-use laboratory coat for working in the ancient DNA laboratory. Gloves were changed frequently. Negative controls included mock extracts and PCR blanks (where water was added instead of DNA extracts), which always failed to yield any amplification product. To screen further for the possibility of contaminants, we aligned the Illumina reads of the DNA library to the reference mitochondrial genome sequence of *Homo sapiens* (GenBank accession number NC\_012920.1 [7]) using BWA version 0.5.7 [8]. Out of 1 billion DNA reads, we found 11 reads with a minimal length of 30 bp that perfectly matched the mitochondrial genome sequence of *Homo sapiens*, five of which also perfectly matched the reference mitochondrial genome of *Bison bison* (GenBank accession number NC\_012346 [9]). By contrast, the same alignment parameters yielded 1,825 reads that perfectly matched the *Bison bison* reference mitochondrial genome sequence.

### DNA extraction

The bone surface was scraped to remove deposits, and 0.5 g of cortical bone powder was produced using a rotary drill. Bone powder was incubated 40 h at 42°C under constant agitation in

5 ml of extraction buffer consisting of 0.45 M EDTA, 10 mM Tris-HCl (pH 8.0), 0.1% SDS, 65 mM DTT, and 0.5 mg/ml proteinase K. After centrifugation, the supernatant was recovered, extracted once with one volume of phenol, once with a phenol-chloroform (50:50) mixture, and once with chloroform. The aqueous phase was then dialyzed and concentrated using Amicon ultra 2 ml centrifugal filter device with a cutoff of 30 kD (Millipore, Billerica, MA), the column was washed 4-times with distilled water, and the DNA extract was recovered in a volume of 100 to 120  $\mu$ l. It was purified further using a silica-membrane-based method designed for small DNA fragments (Qiagen purification kit #28004; Venlo, Netherlands). The final extract consisted of a sample of 100  $\mu$ l.

### PCR screening of the samples for bison DNA content

Screening of the bone samples was performed by PCR. All four samples were analyzed using primer pair 1 (S1 Table). Then, the most promising sample (SGE2) was further studied using primer pairs 2 to 10. PCR was performed in a 50- $\mu$ l reaction volume containing 1  $\mu$ l of mock or ancient DNA extracts, 300 pM of sense and antisense primers, 200  $\mu$ M dNTP, 2.5 mM MgCl<sub>2</sub>, 5  $\mu$ l of GeneAmp 10X PCR buffer II, and 2.5 U of AmpliTaq Gold DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Blank samples (containing water instead of bone or mock extracts) were also included in each PCR experiment. Only bone extracts yielded amplification products. After an activation step (95°C, 8.5 min), a single round of 45 PCR cycles (95°C, 15 sec; 50–60°C (according to primers *T<sub>m</sub>*), 20 sec; 70°C, 1 min) was performed in a Veriti thermal cycler (Thermo Fisher Scientific). The full reaction volume was loaded onto an 8% polyacrylamide gel stained with Sybr Green I (Thermo Fisher Scientific). PCR amplicons were eluted from the gel and inserted into pCR4-TOPO (Thermo Fisher Scientific). Each PCR amplicon was characterized by sequencing 3 to 33 cloned DNA fragments using M13 forward primer.

### Generation and sequencing of a library of DNA fragments

A library of DNA fragments suitable for high-throughput sequencing with the Illumina procedure [10] was generated from 5  $\mu$ l of the SGE2 DNA extract 2 using the Illumina DNA sample prep kit (PE-102-1001). We followed the manufacturer's recommendations (San Diego, CA, USA), except for modifications described elsewhere [11] that were introduced for the purpose of analyzing ancient DNA. Briefly, blunt-end DNA fragments were produced using T4 and Klenow DNA polymerases, and 5' ends were phosphorylated using T4 polynucleotide kinase. A deoxyadenosine was added to the 3' ends of the fragments using Klenow exo<sup>-</sup> DNA polymerase and dATP to prevent the DNA fragments from ligating to one another during the adapter ligation reaction. Ligation to Illumina adapters was then performed using DNA ligase and one tenth of the amount of adapters recommended by the manufacturer.

Amplification of the library was performed using Phusion DNA polymerase. Twelve PCR cycles were enough to produce sufficient amounts of DNA for sequencing.

DNA sequencing was performed at Genoscope (Evry, France) on Illumina HiSeq 2000 (3 sequencing lanes) and HiSeq 2500 platforms (4 sequencing lanes) using HiSeq v3 chemistry with a read length set to 101 nucleotides and analysis on the single read mode.

### Illumina sequencing data analysis

Reads were trimmed for adapter sequences, N's and low quality stretches on the 3' end, using a software based on the FASTX-Toolkit package ([http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit)) and designed by Genoscope. After this step, sequences shorter than 20 nucleotides were discarded using an in-house Python script, yielding a dataset of 1,033,460,536 DNA reads (543,346,926

reads from the HiSeq 2000 sequencing lanes, and 490,113,610 reads from the HiSeq 2500 sequencing lanes). The DNA reads of this study have been deposited at EBI under accession number ERP006376.

### Assembling the *Bison priscus* mitochondrial genome sequence

The sequence assembly was performed in three steps. First the Illumina reads ranging in size between 20 and 101 nucleotides were aligned to the mitochondrial genome of *Bison bison* and to the mitochondrial control region of *Bison priscus* isolate BS212 (GenBank accession number AY748539.1, [2]) using BWA with default options except for *-o* (gap opening) and *-n* (maximum edit distance), which were set to 0 and 2, respectively. This process yielded a set of 3,816 unique reads providing a 9.8-fold coverage of the *Bison bison* reference sequence and left 87 positions without coverage. A provisional consensus sequence was then derived from the 3,816 unique selected reads. In this study a consensus for a given position in the genome was based on at least two concordant unique sequences (either Illumina reads or consensus PCR sequences) with a minimal mapping quality of 25 for Illumina reads. The most frequent base was taken as the consensus. The few cases where more than 25% of the bases from the aligned reads differed from the consensus were manually examined: the large majority of the alternative bases could be ascribed to damage-induced G to A substitutions at the 3' end of the Illumina reads and the other cases were checked by PCR sequencing. The Illumina reads aligning to the provisional consensus sequence were retrieved using BWA with the options *-o* 0 and *-n* 0.1, which allows for a maximum edit distance that increases with read length, from one mismatch for 20–26 nucleotide-long reads to four mismatches for 88–101 nucleotide-long reads. This second set of reads contained 3,868 unique reads. Finally PCR experiments were carried out to fill the gaps (39 positions) and derive a robust sequence at positions where a single read was available (176 positions) or where no consensus was reached (5 positions). PCR primers (primer pairs 11–38 in S1 Table, annealing temperature 50 to 55°C) were designed using the provisional genome sequence. The experimental setup for amplification, purification and cloning of the PCR fragments was similar to the one described for the PCR screening procedure of the samples. We performed one to four amplifications with each primer pair and the consensus sequence deduced from an average of 6 clones for each PCR fragment was used to complete the assembly. We noticed a difference between the consensus of all PCRs and the Illumina reads for only 3 out of the 176 positions covered by a single read. The final sequence (SGE2seq) was established by combining the information derived from the Illumina reads and 81 PCR fragments. All the Illumina reads of the library were ultimately aligned to SGE2seq using BWA with the options *-o* 0 and *-n* 0.1, yielding 3,851 unique reads. We verified that the consensus sequence derived from these reads was identical to SGE2seq.

### Sequence annotation

The annotation of SGE2seq was derived from the annotation of the *Bison bison* reference mitochondrial genome. Briefly, SGE2seq and the *Bison bison* reference mitochondrial genome were aligned using the LAGAN software [12] with the annotation data of the *Bison bison* genome ([http://www.ncbi.nlm.nih.gov/nuccore/NC\\_012346.1](http://www.ncbi.nlm.nih.gov/nuccore/NC_012346.1)). LAGAN produced a list of the aligned genomic features with their coordinates in the two genomes, which was manually checked. The occurrence of the expected start and stop codons at the ends of the coding sequences and the absence of internal stop codon were verified using an in-house Python script. The SGE2seq annotated sequence has been deposited in the GenBank database (accession number KM593920).

## Analysis of ancient DNA damage

The mismatches between the 3,851 unique Illumina reads and the SGE2seq sequence were analyzed using in-house Python and R scripts available upon request. Briefly, the SAM files generated by BWA were parsed to retrieve the aligned reads and the corresponding genomic regions, which were compared for mismatches.

## Illumina reads mapping to other *Bos* and *Bison* sequences

To confirm the origin of the SGE2 sample, all the Illumina reads of the library with a minimal length of 30 bp were aligned to the reference mitochondrial genomes of *Bison bison*, *Bison bonasus* (GenBank accession number NC\_014044.1 [13]), *Bos primigenius* (GenBank accession number NC\_013996.1 [14]) and *Bos taurus* (GenBank accession number NC\_006853.1 [15]) using BWA with the options *-o* 0 (no gap opening) and *-n* 0 (perfect match).

To evaluate the coverage of the nuclear genome, the Illumina reads were aligned to the 18S rDNA sequence of *Bos taurus* (extracted from the sequence with GenBank accession number DQ222453.1) using BWA with the options *-o* 0 (no gap opening) and *-n* 1 (one mismatch). The Illumina reads were also aligned to the complete *Bos taurus* nuclear genome generated by inflating and combining the *bt\_ref\_Bos\_taurus\_UMD\_3.1.1\_chr\*.fa.gz* files along with the *bt\_alt\_Btau\_4.6.1\_chrY.fa.gz* file located at [ftp://ftp.ncbi.nlm.nih.gov/genomes/Bos\\_taurus/Assembled\\_chromosomes/seq/](ftp://ftp.ncbi.nlm.nih.gov/genomes/Bos_taurus/Assembled_chromosomes/seq/). The mapping was performed using BWA with the options *-o* 0 (no gap opening) and *-n* 0.15.

## Metagenomic analyses

A total of one million reads sampled from the seven sequencing lanes was analyzed by BLAST [16] against the GenBank *nr/nt* database with the following options: *-task* megablast *-word\_size* 19 *-max\_target\_seqs* 1 *-gapopen* 5 *-gapextend* 2. Only the hits that display an E-value lower than 0.01 were considered significant.

## Phylogenetic analyses

Phylogenetic analyses were performed using MEGA [17]. The phylogenetic relationships between the D-loop regions of SGE2seq and of several *Bison priscus* fossils and *Bison bison* specimens were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano model, as in [2]. The bootstrap consensus tree was built from 1000 replicates. Initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach. The analysis involved a total of 583 nucleotide positions of the D-loop region.

The phylogenetic relationships between the complete mitochondrial genomes of *Bison priscus* (this study), *Bison bison*, *Bison bonasus* and *Bos grunniens* available on the NCBI website (<http://www.ncbi.nlm.nih.gov/>, last accessed August 4, 2014) along with the reference mitochondrial genomes of *Bos primigenius*, *Bos taurus* and *Bubalus bubalis* (swamp buffalo, GenBank accession number NC\_006295.1) were inferred using the Maximum Likelihood, the Minimum Evolution and the Neighbor-Joining methods. The Maximum Likelihood method was based on the General Time Reversible model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). The analysis involved a total of 16,315 positions.

## AMS $^{14}\text{C}$ dating

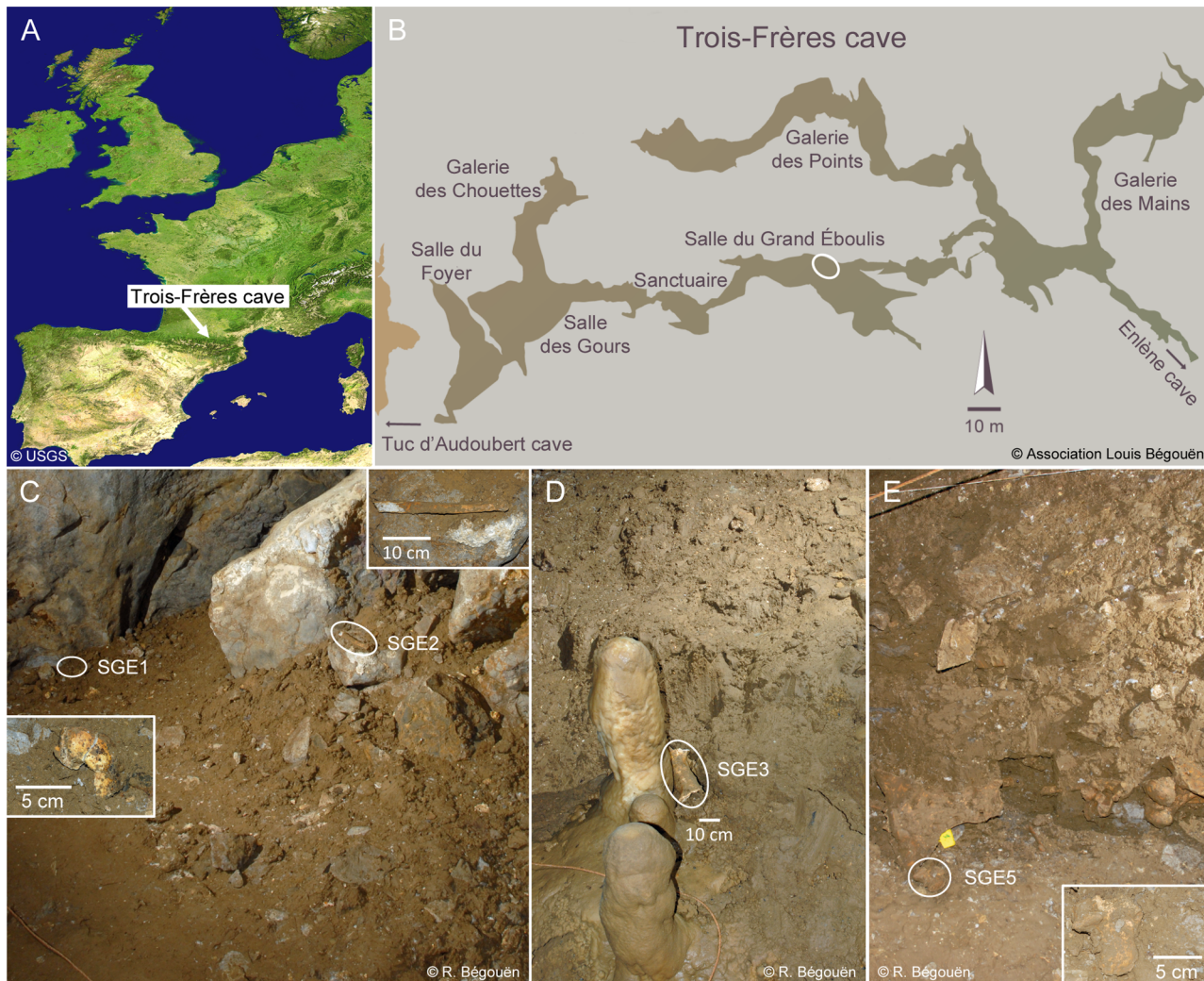
Part of the SGE2 sample was broken into 3–4 mm fragments, which were washed in deionized (DI) water with brief sonication to remove adhering sediment. The bone fragments were then decalcified in 0.2 N HCl for 6 hours at 4°C and the acid-insoluble collagen washed with DI water. The collagen was treated with 0.1 M KOH for 2 hours at 4°C, and washed to neutrality with DI water. The alkali-extracted, decalcified collagen was gelatinized in 0.05 N HCl at 90°C for 1 hour and the solution filtered through a 0.45  $\mu\text{m}$  Durapore filter membrane. The soluble collagen was passed through a 30-kDa filter, the > 30 kDa fraction retained and freeze-dried. Purified collagen was then combusted at 850°C in evacuated quartz tubes containing CuO and Ag. The  $\text{CO}_2$  was isolated cryogenically, converted into graphite using the  $\text{H}_2$ -Fe method [18] and dated at the W.M. Keck Carbon Cycle Accelerator Mass Spectrometry Laboratory, University of California-Irvine (UCIAMS). The calibrated (cal BP) date range of the sample was obtained using OxCal 4.2 software [19] and the IntCal13 calibration curve [20].

## Results and Discussion

The Trois-Frères cave consists of a series of chambers that spread over 800 meters, from the Enlène cave to the Tuc d'Audoubert cave, from which it is currently disconnected (Fig 1A and 1B). In this cave system the *Salle du Grand Éboulis* (Chamber of the Large Scree) displays unique features. It functioned as a natural trap for Pleistocene animals, and was sealed before the Holocene [21]. The bottom of the scree contains remains of the extinct cave bear (*Ursus spelaeus*), one of which was dated to 36,600–34,800 calBP [21]. In the other layers remains of the extinct steppe bison predominate [22]. They correspond to 19,400 to 17,800-year-old specimens [21].

We analyzed four bone samples that were collected in the *Salle du Grand Éboulis* (SGE) for the purpose of the present study (Fig 1C–1E). Three samples stood on the ground surface, in the vicinity of the pathway delineated since the cave discovery (SGE1, SGE3), or at the bottom of the excavation carried out in the 30's of the previous century (SGE5). The other sample (SGE2, a rib fragment) stood on a rock, three meters away from the pathway used by modern visitors.

We first analyzed the DNA content of the samples by PCR using primer pair 1 (S1 Table). The forward and reverse primers display a perfect match with the *Bos taurus* and *Bos primigenius* reference mitochondrial genome sequences and 0 to 1 mismatch with the *Bison bison* and *Bison bonasus* reference mitochondrial genomes. The mitochondrial DNA sequence flanked by these primers (63-bp) is identical between the *Bos taurus* and *Bos primigenius* reference genomes. *Bison bison* and *Bison bonasus* sequences both display 5 differences (including one transversion) with the two *Bos* genomes, and differ among each other by two transitions. All samples, but none of the negative controls (mock extracts and PCR blanks) yielded amplification products of the predicted size that were characterized through the sequencing of cloned PCR fragments (SGE1: 1 PCR, 13 clones; SGE2 extract 1: 4 PCR, 47 clones; SGE2 extract 2: 2 PCR, 39 clones; SGE3: 1PCR, 12 clones; SGE5: 1PCR, 14 clones). SGE1 and SGE3 DNA extracts yielded sequences that all displayed a perfect match with *Bos* (*Bos taurus*, *Bos primigenius*) reference mitochondrial genomes, and SGE5 returned a mixture of sequences that perfectly matched either *Bos* or *Bison bison* genomes (Fig 2). The PCR data may indicate that SGE1 and SGE3 samples correspond to *Bos* bone fragments, but the observation that SGE5 contains *Bos* as well as *Bison* DNA challenges this interpretation and rather suggests that all samples collected from the ground surface had been contaminated by modern cattle DNA brought by modern human intrusions since the cave discovery. At any rate, none of these three samples could be considered as a relevant candidate for a thorough analysis of bison DNA. The SGE2 bone sample turned

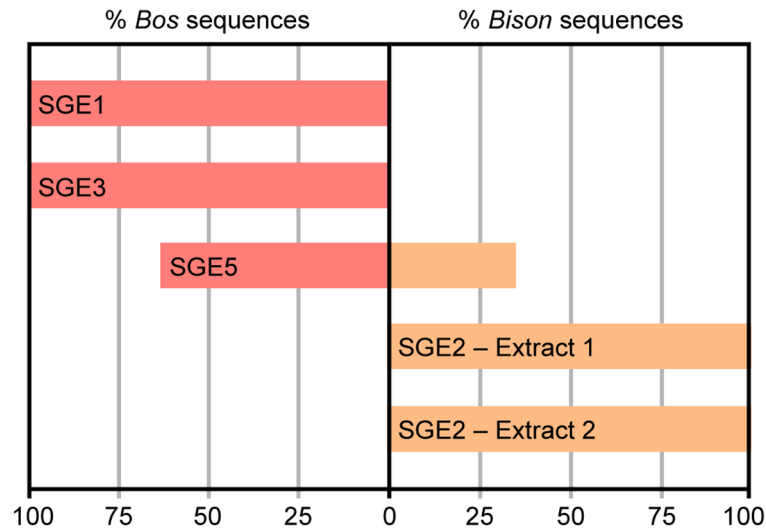


**Fig 1. Overview of the cave site and bone samples.** (A) Geographic localization of the Trois-Frères cave. (B) Topography of the cave showing the site (white ellipse) where the samples were collected in the *Salle du Grand Éboulis*. (C-E) The samples considered for DNA extraction in the present study. SGE1: ulna distal fragment; SGE2: rib; SGE3: humerus fragment; SGE5, vertebra.

doi:10.1371/journal.pone.0128267.g001

out to be much more suitable for this purpose. The two SGE2 DNA extracts yielded sequences that better aligned to *Bison bison* than to *Bos taurus* and *Bos primigenius* mitochondrial genomes (Fig 2 and S2 Table). As a whole, the majority (79%) of these sequences displayed a perfect match with the *Bison bison* reference mitochondrial genome (5 mismatches with *Bos* genomes), the other ones bearing 1–3 and 6–8 differences with the *Bison bison* and *Bos* genomes, respectively.

To further guarantee the relevance of the SGE2 bone sample to study bison DNA, we performed an additional series of PCR experiments using primers designed from *Bison* (primer pairs 2 and 7–10) or *Bos* (primer pairs 3–6) reference mitochondrial genomes. As shown in S2 Table, all these experiments yielded DNA sequences that aligned with higher BLAST scores to *Bison* than to *Bos* mitochondrial genomes. Moreover, using primers targeting the portion of the *Bison priscus* mitochondrial genome made available by previous studies [2], namely the D-loop region, we obtained DNA sequences that clearly exhibited higher alignment scores with this species than with any other one. Thus, the bulk of evidence supported the notion that



**Fig 2. PCR analysis of the samples for Bovinae sequences.** DNA extracts were analyzed for a fragment of the mitochondrial cytb gene using primers encompassing a 63-bp sequence where *Bos* (*Bos taurus*, *Bos primigenius*) and *Bison* (*Bison bison*, *Bison bonasus*) reference genomes display 5 differences. The cloned PCR products were sequenced and identified as *Bos* or *Bison* according to the number of differences with the reference genomes (see text for details).

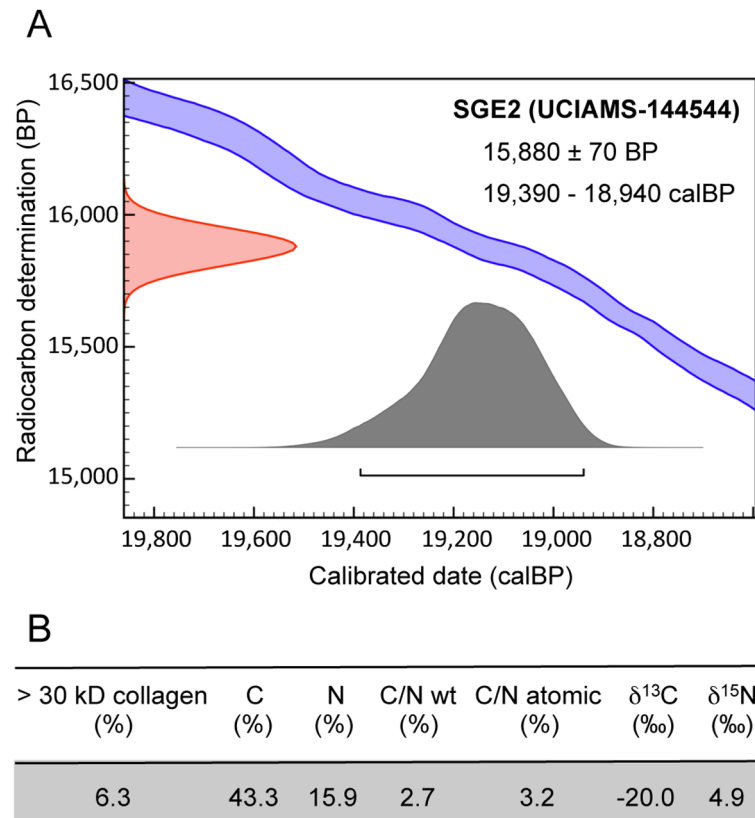
doi:10.1371/journal.pone.0128267.g002

SGE2 corresponds to a *Bison* bone sample. We therefore decided to further analyze this rib fragment.

As shown in Fig 3, SGE2 was radiocarbon dated to  $15,880 \pm 70$  B.P. (19,390–18,940 calBP). It is therefore representative of the age previously obtained for *Bison priscus* bones collected in the *Salle du Grand Éboulis* [21]. In order to obtain genomic data, we produced a library of DNA fragments from the SGE2 sample and performed Illumina sequencing.

A total of 1,033,460,536 Illumina single-pass reads, each at least 20 nucleotides in length, were produced. To confirm the origin of the SGE2 sample and further rule out contamination by modern DNA, the reads were aligned to the reference mitochondrial genomes of *Bison bison*, *Bison bonasus*, *Bos primigenius* and *Bos taurus*. The alignments were performed under stringent conditions: only the reads with a minimal length of 30 bp matching perfectly the target sequence were retained. Under these conditions, 1,825 reads mapped to the *Bison bison* mitochondrial genome. By contrast, only 324, 347 and 331 reads mapped to the mitochondrial genomes of *Bison bonasus*, *Bos primigenius* and *Bos taurus*, respectively. Moreover, the large majority of the reads mapping to the *Bison bonasus* and *Bos* genomes also mapped to the *Bison bison* genome: this was the case for 296 reads out of 324 for the *Bison bonasus* genome, and for 305 reads out of 347 and for 290 reads out of 331, respectively, for the *Bos primigenius* and *Bos taurus* genomes. These results strongly suggest that the reads mapping to the *Bison bonasus* and *Bos* genomes correspond to ancient bison DNA and map to regions that are identical in the genome of *Bison bison* and in the genomes of the other species. The alignment analysis thus corroborated the conclusion, previously drawn from the PCR results, that the SGE2 extract contains *Bison* rather than *Bos* DNA.

To evaluate the coverage of the nuclear genome by the Illumina reads, we analyzed the 18S ribosomal RNA gene, which displays approximately 200 copies in mammalian genomes [23]. Out of  $\sim 10^9$  reads, 110 aligned to the 18S rDNA sequence of *Bos taurus*. These results correspond to an average read depth of 2.2 for 200 copies and accordingly to  $\sim 0.01$  for single copy sequences. Mapping to the complete nuclear genome of *Bos taurus* yielded 741,848 unique



**Fig 3. Radiocarbon and stable isotopes data for the SGE2 bison sample.** (A) Uncalibrated (B.P.) and calibrated (calBP) age of the bone sample. The calBP data correspond to 95.4% (2  $\sigma$ ) confidence interval of the sample age using IntCal13 calibration curve. (B) Characterization of the dated material.

doi:10.1371/journal.pone.0128267.g003

reads with a mapping quality higher than 20. The sum of these reads lengths amounts to 26,117,728 bp, which represents a 0.01-fold coverage of the 2,7 Gb *Bos taurus* nuclear genome, in agreement with our previous estimate. These analyses confirmed the poor coverage of the nuclear genome, which precluded its further analysis.

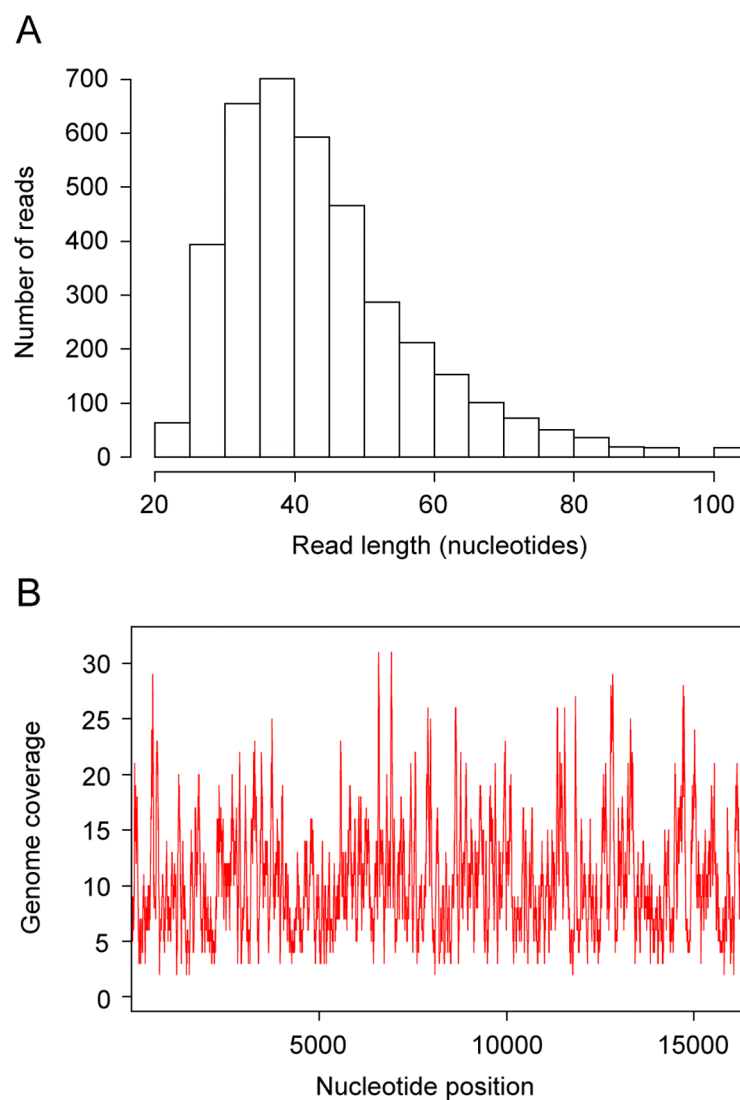
To gain some further insight into the content of the Illumina library, one million reads were analyzed by BLAST, which yielded 99,684 significant hits. The large majority of these hits (89%) corresponded to *Bacteria* while *Eukaryota* represented only 9% of the total. In the bacterial metagenome, the predominant classes were *Actinobacteria* (32% of all bacterial hits), *Beta-proteobacteria* (20%), *Gammaproteobacteria* (18%) and *Alphaproteobacteria* (16%). Similar distributions of microbial diversity have been described for ancient DNA extracts (e.g. [24]). Among the *Eukaryota*, the family most represented was the *Poaceae* (10% of all eukaryotic hits), followed by the *Bovidae* (6%) and the *Noelaerhabdaceae* (3%). The sum of the lengths of the 540 Illumina reads corresponding to the *Bovidae* hits represents a total of 26,856 nucleotides for one million Illumina reads analyzed, predicting 27,754,600 nucleotides for *Bovidae* sequences in the whole set of reads.

To analyze the *Bison* mitochondrial genome sequence, we first retrieved from the library all the reads that aligned to the mitochondrial genome of *Bison bison* (which, as indicated above, turned out to be more relevant for this purpose than *Bison bonasus*) or to the mitochondrial control region of *Bison priscus* isolate BS212 (GenBank accession number AY748539.1, [2]). This alignment was performed under less stringent conditions than the



previous ones and allowed for mismatches (see details in [Materials and Methods](#)). A provisional consensus sequence was derived from the selected reads and was used in turn to retrieve all reads aligning to it. PCR experiments were carried out to analyze regions where less than two concordant reads were available. The final sequence, termed SGE2seq, was established using 81 consensus sequences of PCR fragments to complement the information derived from the Illumina reads. All the Illumina reads were aligned one last time to SGE2seq, which yielded 3,851 unique reads.

SGE2seq consists of a circular genome, for which we obtained a 10.4-fold coverage, taking into account the 3,851 unique Illumina reads and the 81 PCR sequences. [Fig 4A](#) shows the length distribution of the 3,851 unique Illumina reads and [Fig 4B](#), the coverage obtained for each position.



**Fig 4. Characteristics of the sequences used to assemble the *Bison priscus* mitochondrial genome.** (A) Size distribution of the 3,851 unique Illumina reads. (B) Coverage for each position of the SGE2seq sequence obtained by merging the sequences of Illumina reads and PCR fragments.

doi:10.1371/journal.pone.0128267.g004

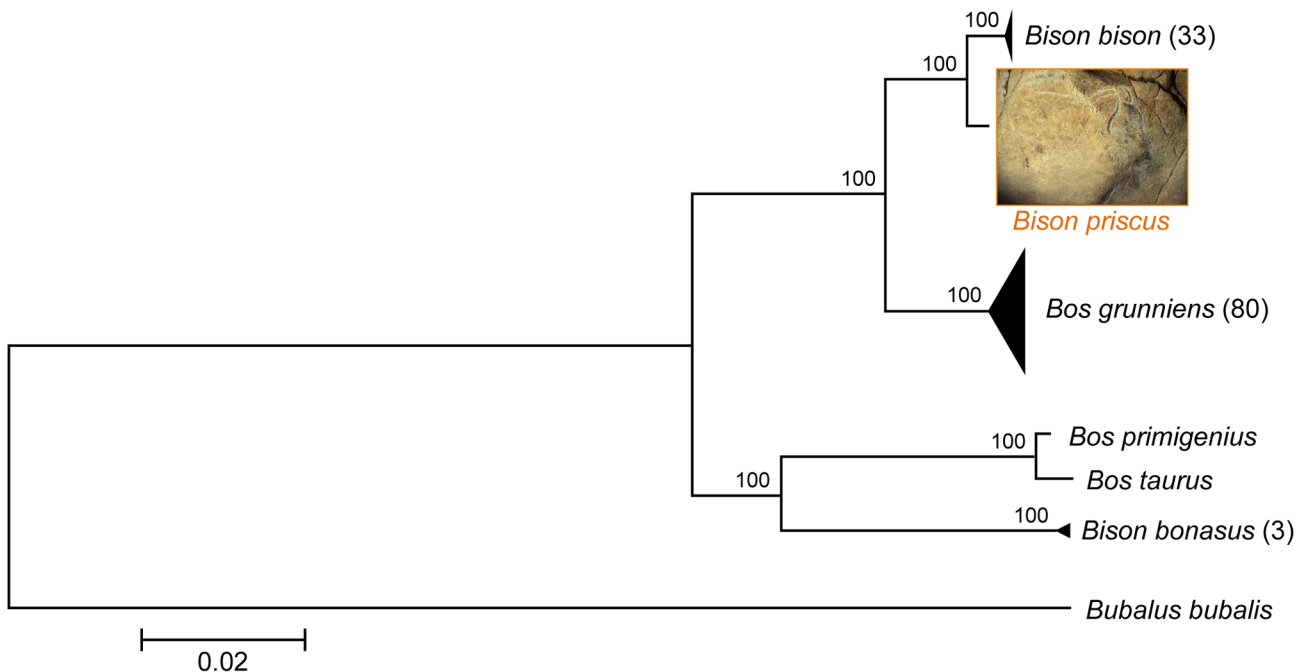
Only 1,443 mismatches out of 166,849 aligned bases are observed between the 3,851 unique Illumina reads and the consensus sequence SGE2seq, which corresponds to 99.1% identity. As shown in [S1 Fig](#), the G to A substitution rate increases at 3' ends, in agreement with what has already been reported for libraries of ancient DNA constructed with the same procedure [25]. An increased substitution rate at DNA ends is considered as a hallmark authenticating sequences generated from ancient DNA fragments which exhibit inflated cytosine deamination rates at 5'-overhangs [26], responsible for G to A transitions at the 3' end of the opposite, repaired DNA strand.

SGE2seq corresponds to a mitochondrial genome sequence of 16,318 bp. This length is similar to the lengths of *Bison bison* mitochondrial genomes, which range from 16,318 to 16,323 bp [27]. The genome consists of 13 protein-coding genes, 22 tDNA genes, 2 rDNA genes, and the D-loop region ([S3 Table](#)).

Since part of the D-loop region has been sequenced for a number of *Bison priscus* fossils [2,4], we analyzed the phylogenetic relationships between the orthologous portion of SGE2seq and of several *Bison priscus* fossils and *Bison bison* specimens. We also considered the D-loop region of several reference genomes of *Bison* and *Bos* species, including the *Bos grunniens* (yak) reference mitochondrial genome (GenBank accession number NC\_006380). The phylogenetic reconstruction, carried out using the Maximum Likelihood method, positions SGE2seq among *Bison priscus* fossils sequences, outside the strongly supported clade of modern *Bison bison* sequences ([S2 Fig](#)). This result was observed when taking into account three specimens from each of the four *Bison* groups defined in [2], and similar results were obtained when including all published D-loop sequences of *Bison priscus* and *Bison bison* (data not shown). The position of SGE2seq D-loop region in the phylogenetic tree further corroborates the assumption that SGE2seq corresponds to the mitochondrial genome of a *Bison priscus* specimen.

Several lines of evidence indicate that we obtained a reliable mitochondrial genome sequence for a Pleistocene *Bison priscus* specimen. First, the DNA material was retrieved from a 19,390 to 18,940-year-old bone sample ([Fig 3](#)). Second, considering the mitochondrial genome portion (600 bp of the D-loop) previously characterized for the extinct steppe bison [2,4], phylogenetic analysis demonstrated that the sequence assembled from our sample positions in the cluster delineated by *Bison priscus* specimens ([S2 Fig](#)). Third, we observed the damage-induced misincorporation pattern expected for ancient DNA ([S1 Fig](#)). Fourth, although of limited magnitude, the average unique read depth (10.4-fold) is large enough to overcome the error rate (< 1%, as judged by comparing the reads with the consensus sequence), and positions poorly covered were checked by PCR and sequencing of multiple cloned amplicons. Fifth, we obtained a circular genome that contains the expected number of genes (rRNAs, tRNAs, protein-coding genes) for a mammalian mitochondrial genome, with full-length coding sequences (without internal stop codons) for all 13 protein-coding genes. Thus, it is safe to conclude that we provide an authentic and complete *Bison priscus* mitochondrial genome sequence.

When single-nucleotide polymorphisms (SNPs) and insertion-deletion events are considered, there are 106–117 differences between SGE2seq and the 33 published *Bison bison* mitochondrial genomes (*Bison bison* genomes differ among themselves at 24 positions at most). SGE2seq shows 114 differences (listed in [S4 Table](#)) compared to the reference *Bison bison* mitochondrial genome, including two indels: a 1-bp insertion located in the D-loop region and a 2-bp deletion at the end of the tRNA serine gene (this shorter tRNA serine allele is also present in some *Bison bison* mitochondrial genomes [27]). The other differences comprise 105 transitions and 6 transversions. A similar, strong transitional bias for mitochondrial genomes has also been reported for *Bos* species [14]. Twenty-five differences (22%) are located in the D-loop region, which therefore exhibits a mutation rate four times higher than the rest of the genome. Finally, 73 substitutions (65%) are located in protein-coding genes, with 16, 3 and 54



**Fig 5. Maximum Likelihood phylogenetic tree of complete mitochondrial genomes of *Bison* and *Bos* species.** The numbers indicated in parentheses after the species name refer to the number of genomes examined for that species (which is equal to 1 by default). The tree with the highest log-likelihood is shown drawn to scale, with branch lengths established from the number of substitutions per site. The same topology was obtained with the Minimum Evolution (ME) and Neighbor-Joining (NJ) methods. The percentages of trees in which the associated taxa clustered together with the ML method are displayed next to the branches (these bootstrap values were determined with 1000 replicates). These percentages were also equal to 100 with the ME and the NJ methods. Photography of painted and engraved bison by Robert Bégouën.

doi:10.1371/journal.pone.0128267.g005

substitutions occurring in the first, second and third codon positions, respectively. These substitutions correspond to 11 amino acid differences. Among the 114 differences, 93 correspond to polymorphisms that are specific to *Bison priscus* and are not found in any published *Bison bison* genome (S5 Table). These 93 specific polymorphisms are mostly located in protein-coding genes (69%) and correspond to 7 amino acids differences.

A phylogenetic analysis was performed using all the available complete mitochondrial genomes of *Bison bison*, *Bison bonasus* and *Bos grunniens* along with the reference mitochondrial genomes of *Bos primigenius*, *Bos taurus* and *Bubalus bubalis* (swamp buffalo). Phylogenetic trees were constructed from this dataset with the Maximum Likelihood (ML), Minimum Evolution (ME) and Neighbor-Joining (NJ) methods. As shown in Fig 5 for the ML method, SGE2seq is basal to the modern *Bison bison* clade, and the set including SGE2seq and all *Bison bison* sequences forms a distinct and well-supported (100% bootstrap support, 1000 replicates) clade. Similar results were obtained with the ME and NJ methods, which conclusively supports the contention that SGE2seq corresponds to a *bona fide* mitochondrial sequence of *Bison priscus*.

In summary, we have presented here the analysis of a 19,000-year-old bone specimen that made it possible to obtain the first complete mitochondrial genome for the extinct steppe bison. Phylogenetic reconstruction unequivocally points to the basal position of the *Bison priscus* mitochondrial genome as compared to the clade delineated by extant *Bison bison* genomes. The sequence provided here may be a useful tool for the genomic analysis of other ancient *Bovinae* specimens. Our study also shows that the Trois-Frères cave system, already famous for Magdalenian art, evidence of human habitation and abundant animal remains, can now be considered for paleogenetic and paleogenomic analyses.

## Supporting Information

**S1 Fig. Position of the mismatches between the 3,851 unique Illumina reads and the SGE2-seq sequence.** The frequencies of the 12 possible mismatches are plotted as a function of the distance from the 5' end (A) or the 3' end (B) of reads. Since the Illumina reads are at least 20 nucleotides in length, only the ten 5' and ten 3' most positions are shown.

(TIF)

**S2 Fig. Maximum Likelihood phylogenetic tree of the D-loop sequences of *Bison* and *Bos* species.** Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The D-loop fragment of SGE2seq is boxed in red (*Bison priscus* TF for "Trois-Frères") and the D-loop sequences of the *Bison bison* clade are boxed in blue.

(TIF)

**S1 Table. PCR primers used in this study.** Primer pair 1 was used to analyze all four bone samples for *Bos* and *Bison* mitochondrial DNA. Primer pairs 2–10 were used to characterize further the SGE2 sample before initiating Illumina sequencing. Primer pairs 11–38 were used to fill in the gaps in the genome reconstructed by shotgun DNA sequencing and check the sequence of the genome regions where a single Illumina read was obtained. The position of each forward (F) and reverse (R) primer is numbered according to the SGE2seq sequence.

(DOCX)

**S2 Table. PCR analysis of the SGE2 bone sample.** Each line provides the results for a PCR sample obtained with one primer pair (column 1) and characterized through the sequencing of the indicated number of cloned fragments (column 2, number of sequences). Columns 3 and 4 indicate the size of the sequence located between the primers and its location in the mitochondrial genome, respectively. The consensus deduced from the sequence of the cloned amplicons was analyzed by BLAST against the GenBank *nr/nt* database. Results of BLAST analysis for the indicated species are displayed in columns 5–19. NA: not applicable (the corresponding genome portion of *Bison priscus* was not available in GenBank).

(DOCX)

**S3 Table. Annotation of the *Bison priscus* mitochondrial genome.** The intergenic nucleotides column indicates the number of intergenic or overlapping nucleotides ("-n" indicates overlapping regions).

(DOCX)

**S4 Table. List of the differences between the mitochondrial genome sequence of *Bison priscus* and the reference genome sequence of *Bison bison*.** The Position column indicates the position of the difference with respect to the *Bison priscus* sequence; the Strand column, the strand of the feature (irrelevant for intergenic regions); the Position codon column, the position of the polymorphism within the codon (for coding sequences); the Base priscus and the Base bison columns, the polymorphic bases for the *Bison priscus* and *Bison bison* sequences, respectively; the Sequence priscus and the Sequence bison columns, the 5-bp sequences where the polymorphism is located (the polymorphic base is in the middle); and the Type column, the type of difference between the *Bison priscus* and the *Bison bison* sequences ("t" for transition, "v" for transversion). NA, not applicable.

(DOCX)

**S5 Table. List of the polymorphisms that are specific to the genome of *Bison priscus*.** The compendium was established by comparing SGE2seq to the genome sequences available for 33 modern *Bison bison* specimens. The column names are as in [S4 Table](#). (DOCX)

## Acknowledgments

We thank A. Martel (CEA Saclay) for bioinformatics support.

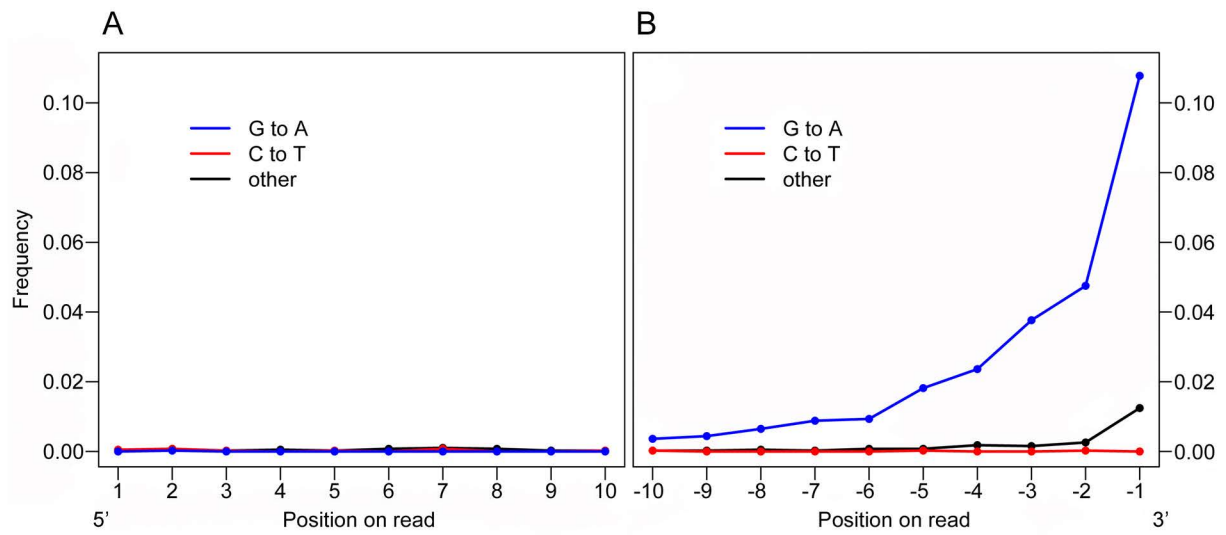
## Author Contributions

Conceived and designed the experiments: JME. Performed the experiments: PP VB TS JME. Analyzed the data: MCMK PP VB TS JME. Contributed reagents/materials/analysis tools: FM RB. Wrote the paper: MCMK JME.

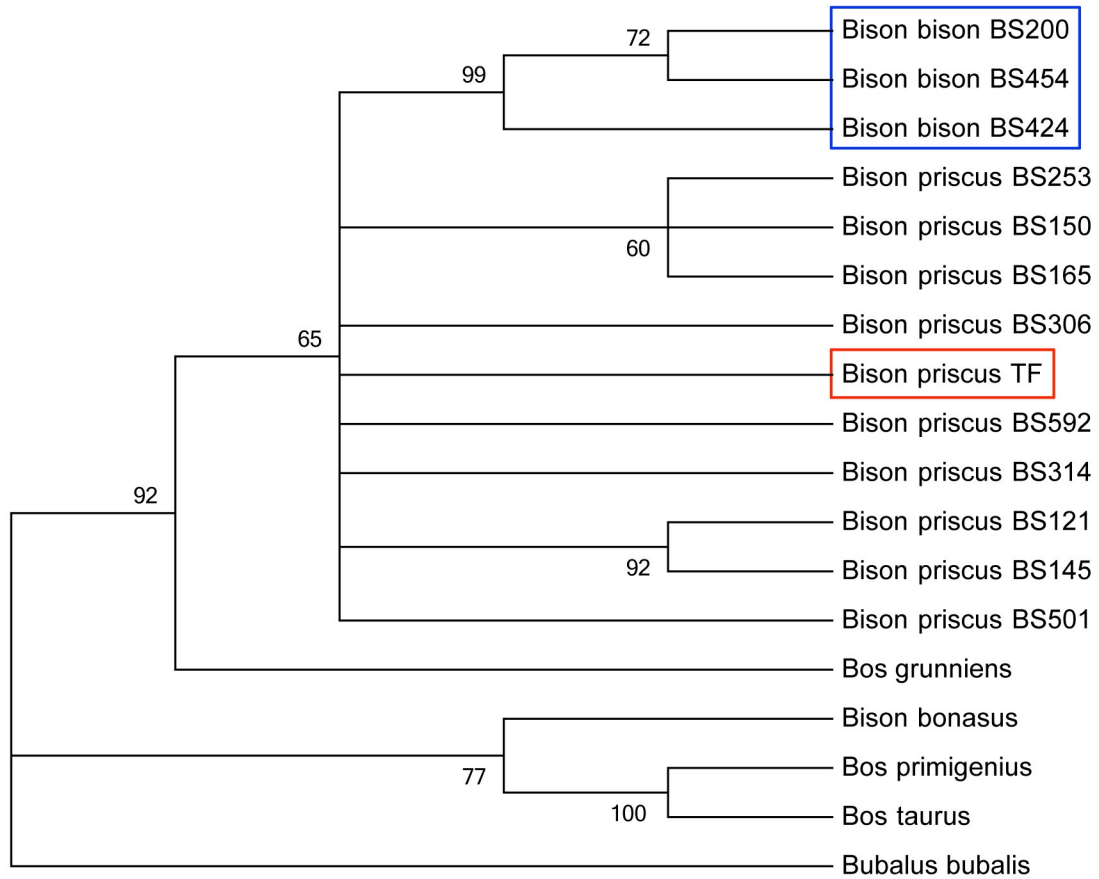
## References

1. Kurten B. Pleistocene Mammals of Europe. Chicago: AldineTransaction; 1968.
2. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, et al. Rise and fall of the Ber- ingian steppe bison. *Science*. 2004; 306: 1561–1565. doi: [10.1126/science.1101074](https://doi.org/10.1126/science.1101074) PMID: [15567864](https://pubmed.ncbi.nlm.nih.gov/15567864/)
3. Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, Lenstra JA. Maternal and paternal lineages in cross- breeding bovine species. Has wisent a hybrid origin? *Mol Biol Evol*. 2004; 21: 1165–1170. doi: [10.1093/molbev/msh064](https://doi.org/10.1093/molbev/msh064) PMID: [14739241](https://pubmed.ncbi.nlm.nih.gov/14739241/)
4. Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, Suter CM. High-resolution analysis of cytosine methylation in ancient DNA. *PLoS ONE*. 2012; 7: e30226. doi: [10.1371/journal.pone.0030226](https://doi.org/10.1371/journal.pone.0030226) PMID: [22276161](https://pubmed.ncbi.nlm.nih.gov/22276161/)
5. Bégouën H, Breuil H. Les cavernes du Volp: Trois Frères-Tuc d'Audoubert à Montesquieu-Avantès (Ariège). Paris: Arts et Métiers Graphiques; 1958.
6. Bégouën H. Les statues d'argile préhistoriques de la caverne du Tuc d'Audoubert (Ariège). *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*. 1912; 532. PMID: [17597171](https://pubmed.ncbi.nlm.nih.gov/17597171/)
7. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999; 23: 147. doi: [10.1038/13779](https://doi.org/10.1038/13779) PMID: [10508508](https://pubmed.ncbi.nlm.nih.gov/10508508/)
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754–1760. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
9. Achilli A, Olivieri A, Pellecchia M, Ubaldi C, Colli L, Al-Zahery N, et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol*. 2008; 18: R157–R158. doi: [10.1016/j.cub.2008.01.019](https://doi.org/10.1016/j.cub.2008.01.019) PMID: [18302915](https://pubmed.ncbi.nlm.nih.gov/18302915/)
10. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456: 53–59. doi: [10.1038/nature07517](https://doi.org/10.1038/nature07517) PMID: [18987734](https://pubmed.ncbi.nlm.nih.gov/18987734/)
11. Bon C, Berthonaud V, Maksud F, Labadie K, Poulain J, Artiguenave F, et al. Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci*. 2012; 279: 2825–2830. doi: [10.1098/rspb.2012.0358](https://doi.org/10.1098/rspb.2012.0358) PMID: [22456883](https://pubmed.ncbi.nlm.nih.gov/22456883/)
12. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003; 13: 721–731. doi: [10.1101/gr.926603](https://doi.org/10.1101/gr.926603) PMID: [12654723](https://pubmed.ncbi.nlm.nih.gov/12654723/)
13. Zeyland J, Wolko L, Lipiński D, Wo niak A, Nowak A, Szalata M, et al. Tracking of wisent-bison-yak mitochondrial evolution. *J Appl Genet*. 2012; 53: 317–322. doi: [10.1007/s13353-012-0090-4](https://doi.org/10.1007/s13353-012-0090-4) PMID: [22415349](https://pubmed.ncbi.nlm.nih.gov/22415349/)
14. Edwards CJ, Magee DA, Park SDE, McGettigan PA, Lohan AJ, Murphy A, et al. A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS ONE*. 2010; 5: e9255. doi: [10.1371/journal.pone.0009255](https://doi.org/10.1371/journal.pone.0009255) PMID: [20174668](https://pubmed.ncbi.nlm.nih.gov/20174668/)
15. Chung H. Phylogenetic analysis and characterization of mitochondrial DNA for Korean native cattle. *Open Journal of Genetics*. 2013; 3: 12–23. doi: [10.4236/ojgen.2013.31003](https://doi.org/10.4236/ojgen.2013.31003)

16. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10: 421. doi: [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421) PMID: [20003500](https://pubmed.ncbi.nlm.nih.gov/20003500/)
17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
18. Vogel JS, Southon JR, Nelson DE, Brown TA. Performance of Catalytically Condensed Carbon for Use in Accelerator Mass-Spectrometry. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms*. 1984; 5: 289–293.
19. Ramsey CB, Lee S. Recent and Planned Developments of the Program Oxcal. *Radiocarbon*. 2013; 55: 720–730.
20. Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, et al. Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years Cal Bp. *Radiocarbon*. 2013; 55: 1869–1887.
21. Bégouën R, Clottes J, Feruglio V, Pastoors A. *La caverne des Trois-Frères*. Paris: Somogy éditions d'art; 2014.
22. Bégouën H. Sur un squelette de bison quaternaire de la caverne des Trois-Frères à Montesquieu-Avantès. *Berichte der internationale Geschichte zur Erhaltung des Wisents*. 1936; 3: 295–300.
23. Paule MR. Class I genes. In: Paule MR, editor. *Transcription of ribosomal RNA genes by eukaryotic RNA polymerase I*. New-York, NY; 1998. pp. 1–8. PMID: [15996904](https://pubmed.ncbi.nlm.nih.gov/15996904/)
24. Sarkissian Der C, Ermini L, Jónsson H, Alekseev AN, Crubezy E, Shapiro B, et al. Shotgun microbial profiling of fossil remains. *Mol Ecol*. 2014; 23: 1780–1798. doi: [10.1111/mec.12690](https://doi.org/10.1111/mec.12690) PMID: [24612293](https://pubmed.ncbi.nlm.nih.gov/24612293/)
25. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 2013; 29: 1682–1684. doi: [10.1093/bioinformatics/btt193](https://doi.org/10.1093/bioinformatics/btt193) PMID: [23613487](https://pubmed.ncbi.nlm.nih.gov/23613487/)
26. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA*. 2007; 104: 14616–14621. doi: [10.1073/pnas.0704665104](https://doi.org/10.1073/pnas.0704665104) PMID: [17715061](https://pubmed.ncbi.nlm.nih.gov/17715061/)
27. Douglas KC, Halbert ND, Kolenda C, Childers C, Hunter DL, Derr JN. Complete mitochondrial DNA sequence analysis of *Bison bison* and bison-cattle hybrids: function and phylogeny. *Mitochondrion*. 2011; 11: 166–175. doi: [10.1016/j.mito.2010.09.005](https://doi.org/10.1016/j.mito.2010.09.005) PMID: [20870040](https://pubmed.ncbi.nlm.nih.gov/20870040/)



**S1 Fig.** Position of the mismatches between the 3,851 unique Illumina reads and the SGE2seq sequence. The frequencies of the 12 possible mismatches are plotted as a function of the distance from the 5' end (A) or the 3' end (B) of reads. Since the Illumina reads are at least 20 nucleotides in length, only the ten 5' and ten 3' most positions are shown.



**S2 Fig.** Maximum Likelihood phylogenetic tree of the D-loop sequences of *Bison* and *Bos* species. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. The D-loop fragment of SGE2seq is boxed in red (*Bison priscus* TF for "Trois-Frères") and the D-loop sequences of the *Bison bison* clade are boxed in blue.



**S1 Table.** PCR primers used in this study

Primer pair	Primer position, sequence	Amplicon (bp)
1	15259F, 5'-CCTCGGAGACCCAGATAACT-3' 15362R, 5'-TTGATCGTAAGATTGCGTATG-3'	104
2	10052F, 5'-GAGCCTCACAAACAACAACT-3' 10175R, 5'-GTACCATATTCGGTTCATTCTAGTC-3'	124
3	14711F, 5'-ACCCATATCTGCCGAGAC-3' 14812R, 5'-TCGTCCTACGTGCATATATAA-3'	102
4	14768F, 5'-GGAGCTCAATGTTTTTATCTG-3' 14867R, 5'-GGATTACTCCAATATTTTCATGTTT-3'	100
5	14818F, 5'-ATATTACGGGTCTTACACTTTTC-3' 14919R, 5'-CCTCATGGTAGGACGTATCC-3'	102
6	14936F, 5'-GGAGCAACAGTCATCACCAA-3' 15053R, 5'-AAGCGAAGAATCGGGTAAGG-3'	118
7	15357F, 5'-GATCAATCCCAACAAAC-3' 15438R, 5'-GTGTGTAGTAGGGGAATTAGA-3'	82
8	15883F, 5'-AACCAAATATTACAAAACCCACTA-3' 16012R, 5'-GGGGCATATAATTTAATGTACT-3'	130
9	16010F, 5'-CCCATGCATATAAGCAAGT-3' 16148R, 5'-TGGTAATTAAGCTCGTGATCTA-3'	139
10	16162F, 5'-CAGCAACCCGCTAGGCA-3' 16266R, 5'-GCCCTGAAGAAAGAACAGATGT-3'	105
11	156F, 5'-CATAATGGTAGGCATGCACA-3' 254R, 5'-AGTGGTAAATATTTAAGGGGGAA-3'	99
12	156F, 5'-CATAATGGTAGGCATGCACA-3' 270R, 5'-AAAATCTGTAAAAGCAGTGGT-3'	115
13	304F, 5'-CAATACTCAAATTAGCACTCCA-3' 399R, 5'-ATTTTCAGTGCCTTGCTTT-3'	96
14	1394F, 5'-AGCTTAAACAAAGCATCCAGT-3' 1483R, 5'-AGTTAAGAGGGTATTTGGGCTA-3'	90
15	1457F, 5'-AATCTAGCCCAAAATACCCTC-3' 1559R, 5'-CGCCATATTTAGATTTCTATCTCC-3'	103
16	1900F, 5'-GGTACAGCCTTTTAGAAACGGAT-3' 1998L, 5'-TTAACGCTTCTCAATTGGTGG-3'	99
17	2013F, 5'-AATTAACAGATCCCAATAACA-3' 2132R, 5'-CTTATGCAAGGAGAAAAGTT-3'	120
18	3060F, 5'-AACTTTTATACCCAGAGATTCA-3' 3177R, 5'-TTTTCGTTCTACTAATGTGAGG-3'	118
19	4024F, 5'-ATCCTTACATCAGGCATCCCA-3' 4134R, 5'-CCTATAGTTCTAGAAATAAGAGGGT-3'	111
20	4037F, 5'-GCATCCCACCACAAACAT-3' 4156R, 5'-ATTCTTAGGAGTAGGTTCAATTC-3'	120
21	4251F, 5'-CCCTTCCCGTACTAATAATCCAA-3' 4359R, 5'-CCGATTCAGACAAGTAGTCAGT-3'	109
22	4439F, 5'-TATTTTCTAACCCAATCAACAGC-3' 4539R, 5'-GGGTAAATAATTTTATTACGGTTCA-3'	101
23	5179F, 5'-CCCCTTACAAACAACAT-3' 5258R, 5'-ATACGGCTATTGTTGGTAGG-3'	80
24	5493F, 5'-ACCCCAATTAATCAGGCTTC-3' 5605R, 5'-AGTTTACACGTTGAATTGCAAAA-3'	113
25	6431F, 5'-CCTGGATTGGAATAATCTCCC-3' 6513R, 5'-GCTCAGACTATTCCTATATATCCG-3'	83
26	7264F, 5'-TTCAAGCCAACATCATAACCTC-3' 7357R, 5'-TTTCACTGTAACTTAACTTTGACA-3'	94
27	8219F, 5'-CACAACCTTCACTACAATCCAG-3' 8301R, 5'-ATTTTCGTTCAATTTGTTTCTCA-3'	83
28	8290F, 5'-ATGAACGAAAATTTATTTGCCTCT-3' 8403R, 5'-CACTAATCGATTGATGTTGGGA-3'	114
29	11068F, 5'-ATACTGAGTACAACCTGTGCAT-3' 11173R, 5'-TCAAAGGTGGAGGCCGTAT-3'	106
30	11838F, 5'-TATCGCTGCACATCTTACCT-3' 11939R, 5'-CACAATCTAATGTTTTTGTTAAACT-3'	102
31	11851F, 5'-CTTACCTTTACTACTCTTATCCCT-3' 11960R, 5'-ATGAGTTTCTATGTTAGATTCAC-3'	110
32	12077F, 5'-CGAAAAAATGGTGCAACTCC-3' 12178R, 5'-AGGCTTGTATTATAATGGGTAC-3'	102
33	12320F, 5'-CCAAACCCTTAAACTATCCCTC-3' 12393R, 5'-CGAATAGTGCTACTGGGACA-3'	74
34	12876F, 5'-ATCTTCTATTAATCCGCTTT-3' 12977R, 5'-TATCGCTGTAATAACGTGGT-3'	102
35	14087F, 5'-TCAAACACAATCTCTACCTCC-3' 14173R, 5'-TTTAGGGCATTATTAATCTGG-3'	87
36	14353F, 5'-CCCATAAATAGGTGAAGGC-3' 14457R, 5'-AGATTCCATGTAAGAATAATGA-3'	105
37	15547F, 5'-CCCATATATCATCATCGGACA-3' 15639R, 5'-AGTTTGTTTTCAATTGTGCC-3'	93
38	15922F, 5'-CCCCAAAATGCATTACCCAA-3' 16016R, 5'-GCATGGGGCATATAATTTAATGT-3'	95

Primer pair 1 was used to analyze all four bone samples for *Bos* and *Bison* mitochondrial DNA. Primer pairs 2-10 were used to characterize further the SGE2 sample before initiating Illumina sequencing. Primer pairs 11-38 were used to fill in the gaps in the genome reconstructed by shotgun DNA sequencing and check the sequence of the genome regions where a single Illumina read was obtained. The position of each forward (F) and reverse (R) primer is numbered according to the SGE2seq sequence.

**S2 Table.** PCR analysis of the SGE2 sample

Primer pair	Number of sequences	Query (bp)	Genome portion	<i>Bos taurus</i>			<i>Bos primigenius</i>			<i>Bison bonasus</i>			<i>Bison bison</i>			<i>Bison priscus</i>		
				Query cover (%)	E value	Identity (%)	Query cover (%)	E value	Identity (%)	Query cover (%)	E value	Identity (%)	Query cover (%)	E value	Identity (%)	Query cover (%)	E value	Identity (%)
1	28	63	CYTB	100%	2E-17	94%	100%	8E-16	92%	100%	4E-20	97%	100%	7E-23	100%	NA	NA	NA
1	8	63	CYTB	100%	2E-17	94%	100%	8E-16	92%	100%	4E-20	97%	100%	7E-23	100%	NA	NA	NA
1	3	63	CYTB	100%	2E-17	94%	100%	8E-16	92%	100%	4E-20	97%	100%	7E-23	100%	NA	NA	NA
1	8	63	CYTB	100%	2E-17	94%	100%	8E-16	92%	100%	4E-20	97%	100%	7E-23	100%	NA	NA	NA
1	33	63	CYTB	100%	2E-17	94%	100%	8E-16	92%	100%	4E-20	97%	100%	7E-23	100%	NA	NA	NA
1	6	63	CYTB	100%	8E-16	90%	100%	3E-14	89%	100%	2E-18	94%	100%	3E-21	97%	NA	NA	NA
2	7	77	ND3	100%	1E-27	97%	100%	5E-26	96%	100%	3E-23	94%	100%	1E-28	99%	NA	NA	NA
2	6	77	ND3	100%	5E-26	96%	100%	7E-25	95%	100%	3E-22	92%	100%	1E-27	97%	NA	NA	NA
3	8	63	CYTB	100%	4E-20	97%	100%	4E-20	97%	100%	2E-21	92%	100%	7E-23	100%	NA	NA	NA
3	6	63	CYTB	100%	4E-20	97%	100%	4E-20	97%	100%	2E-21	92%	100%	7E-23	100%	NA	NA	NA
3	8	63	CYTB	100%	4E-20	97%	100%	4E-20	97%	100%	2E-21	92%	100%	7E-23	100%	NA	NA	NA
4	7	53	CYTB	100%	3E-13	94%	100%	3E-13	94%	100%	3E-13	94%	100%	5E-16	98%	NA	NA	NA
4	8	53	CYTB	100%	3E-12	92%	100%	3E-12	92%	100%	3E-12	92%	100%	1E-17	100%	NA	NA	NA
4	8	53	CYTB	100%	3E-12	92%	100%	3E-12	92%	100%	3E-12	92%	100%	1E-17	100%	NA	NA	NA
4	8	53	CYTB	100%	3E-12	92%	100%	3E-12	92%	100%	3E-12	92%	100%	1E-17	100%	NA	NA	NA
5	15	59	CYTB	100%	2E-15	93%	100%	2E-15	93%	100%	1E-13	92%	100%	4E-19	98%	NA	NA	NA
6	6	78	CYTB	100%	2E-25	95%	100%	2E-25	95%	100%	2E-25	95%	100%	4E-28	97%	NA	NA	NA
6	4	78	CYTB	100%	2E-26	96%	100%	2E-26	96%	100%	2E-26	96%	100%	3E-29	99%	NA	NA	NA
6	6	78	CYTB	100%	2E-26	96%	100%	2E-26	96%	100%	2E-26	96%	100%	3E-29	99%	NA	NA	NA
6	8	78	CYTB	100%	2E-26	96%	100%	2E-26	96%	100%	2E-26	96%	100%	3E-29	99%	NA	NA	NA
6	7	78	CYTB	100%	2E-25	95%	100%	2E-25	95%	100%	2E-25	95%	100%	4E-28	97%	NA	NA	NA
7	16	43	CYTB	100%	7E-12	98%	100%	7E-12	98%	93%	5E-10	98%	100%	3E-14	100%	NA	NA	NA
8	16	84	D-loop	100%	2E-19	88%	82%	3E-11	85%	82%	1E-09	83%	100%	2E-31	98%	100%	2E-32	99%
9	8	98	D-loop	90%	8E-26	91%	88%	5E-22	89%	89%	4E-20	83%	100%	2E-33	94%	100%	6E-40	99%
9	8	98	D-loop	90%	2E-26	90%	88%	2E-20	87%	89%	4E-17	84%	100%	9E-32	93%	100%	7E-39	98%
10	8	65	D-loop	98%	9E-22	98%	98%	6E-18	94%	98%	6E-18	94%	100%	2E-18	94%	100%	6E-24	100%
10	8	65	D-loop	98%	9E-22	98%	98%	6E-18	94%	98%	6E-18	94%	100%	2E-18	94%	100%	6E-24	100%
10	8	65	D-loop	98%	9E-22	98%	98%	6E-18	94%	98%	6E-18	94%	100%	2E-18	94%	100%	6E-24	100%

Each line provides the results for a PCR sample obtained with one primer pair (column 1) and characterized through the sequencing of the indicated number of cloned fragments (column 2, number of sequences). Columns 3 and 4 indicate the size of the sequence located between the primers and its location in the mitochondrial genome, respectively. The consensus deduced from the sequence of the cloned amplicons was analyzed by BLAST against the GenBank *nr/nt* database. Results of BLAST analysis for the indicated species are displayed in columns 5-19. NA: not applicable (the corresponding genome portion of *Bison priscus* was not available in GenBank).

**S3 Table.** Annotation of the *Bison priscus* mitochondrial genome sequence.

Feature	Strand	Start	Stop	Length	Intergenic nucleotides
D-loop	+	1	361	361	1
tRNA-Phe	+	363	429	67	0
s-rRNA	+	430	1385	956	0
tRNA-Val	+	1386	1452	67	0
l-rRNA	+	1453	3022	1570	0
tRNA-Leu	+	3023	3097	75	2
ND1	+	3100	4055	956	0
tRNA-Ile	+	4056	4124	69	-3
tRNA-Gln	-	4122	4193	72	2
tRNA-Met	+	4196	4264	69	0
ND2	+	4265	5306	1042	0
tRNA-Trp	+	5307	5373	67	1
tRNA-Ala	-	5375	5443	69	1
tRNA-Asn	-	5445	5518	74	0
rep_origin	+	5519	5549	31	1
tRNA-Cys	-	5551	5617	67	0
tRNA-Tyr	-	5618	5685	68	1
COX1	+	5687	7231	1545	-3
tRNA-Ser	-	7229	7299	71	4
tRNA-Asp	+	7304	7372	69	1
COX2	+	7374	8057	684	3
tRNA-Lys	+	8061	8127	67	1
ATP8	+	8129	8329	201	-40
ATP6	+	8290	8970	681	-1
COX3	+	8970	9750	781	3
tRNA-Gly	+	9754	9822	69	0
ND3	+	9823	10169	347	0
tRNA-Arg	+	10170	10238	69	0
ND4L	+	10239	10535	297	-7
ND4	+	10529	11906	1378	0
tRNA-His	+	11907	11976	70	0
tRNA-Ser	+	11977	12036	60	1
tRNA-Leu	+	12038	12107	70	0
ND5	+	12108	13928	1821	-17
ND6	-	13912	14439	528	0
tRNA-Glu	-	14440	14508	69	4
CYTB	+	14513	15652	1140	4
tRNA-Thr	+	15657	15725	69	-1
tRNA-Pro	-	15725	15790	66	0
D-loop	+	15791	16318	528	0

The intergenic nucleotides column indicates the number of intergenic or overlapping nucleotides ("-n" indicates overlapping regions).

**S4 Table.** List of the differences between the mitochondrial genome of *Bison priscus* and the reference genome of *Bison bison*.

Position	Feature	Strand	Position codon	Base priscus	Base bison	Sequence priscus	Sequence bison	Amino acid priscus	Amino acid bison	Type
8	D-loop	+	NA	G	A	TGGCT	TGACT	NA	NA	t
144	D-loop	+	NA	T	C	ACTTT	ACCTT	NA	NA	t
166	D-loop	+	NA	G	A	TAGGC	TAAGC	NA	NA	t
201	D-loop	+	NA	A	-	AAATT	AA-CT	NA	NA	indel
202	D-loop	+	NA	T	C	AATTA	A-CTG	NA	NA	t
204	D-loop	+	NA	A	G	TTATA	CTGTA	NA	NA	t
461	s-rRNA	+	NA	T	C	CTTTT	CTCTT	NA	NA	t
740	s-rRNA	+	NA	T	C	ATTAA	ATCAA	NA	NA	t
807	s-rRNA	+	NA	G	A	GTGAC	GTAAC	NA	NA	t
1158	s-rRNA	+	NA	T	C	ATTAA	ATCAA	NA	NA	t
1473	l-rRNA	+	NA	C	T	TACCC	TATCC	NA	NA	t
1495	l-rRNA	+	NA	A	G	AAATA	AAGCA	NA	NA	t
1496	l-rRNA	+	NA	T	C	AATAA	AGCAA	NA	NA	t
2054	l-rRNA	+	NA	T	C	CCTAA	CCCAA	NA	NA	t
2562	l-rRNA	+	NA	G	A	TTGAC	TTAAC	NA	NA	t
2581	l-rRNA	+	NA	T	A	AATAA	AAAAA	NA	NA	v
2978	l-rRNA	+	NA	A	G	CAACA	CAGCA	NA	NA	t
3102	ND1	+	3	A	G	ATATT	ATGTT	Met	Met	t
3114	ND1	+	3	T	C	AATAT	AACAT	Asn	Asn	t
3301	ND1	+	1	A	G	CAACC	CAGCC	Thr	Ala	t
3555	ND1	+	3	G	A	TCGGT	TCAGT	Ser	Ser	t
3798	ND1	+	3	G	A	ATGAT	ATAAT	Met	Met	t
3826	ND1	+	1	C	T	TTCTA	TTTTA	Leu	Leu	t
3862	ND1	+	1	T	C	TCTAC	TCCAC	Tyr	His	t
4047	ND1	+	3	A	G	CCACA	CCGCA	Pro	Pro	t
4453	ND2	+	3	A	G	CAATC	CAGTC	Gln	Gln	t
4733	ND2	+	1	C	T	CCCTA	CCTTA	Leu	Leu	t
4754	ND2	+	1	G	A	TAGTT	TAATT	Val	Ile	t
4852	ND2	+	3	C	T	TACAA	TATAA	Tyr	Tyr	t
5107	ND2	+	3	C	T	TTCAT	TTTAT	Phe	Phe	t
5253	ND2	+	2	C	T	AGCCG	AGTCG	Ala	Val	t
5291	ND2	+	1	C	T	TACTA	TATTA	Leu	Leu	t
5374	intergene	NA	NA	T	C	TGTTA	TGCTA	NA	NA	t
5614	tRNA-Cys	-	NA	A	G	AGAGC	AGGGC	NA	NA	t
5803	COX1	+	3	T	C	GCTGA	GCCGA	Ala	Ala	t
6283	COX1	+	3	C	T	CTCCC	CTTCC	Leu	Leu	t
6301	COX1	+	3	C	T	GGCAT	GGTAT	Gly	Gly	t
6469	COX1	+	3	T	C	TATTC	TACTC	Tyr	Tyr	t
6634	COX1	+	3	C	A	ACCGG	ACAGG	Thr	Thr	v
7297.5	tRNA-Ser	-	NA	-	T	TC--A	TCTCA	NA	NA	indel
7297.5	tRNA-Ser	-	NA	-	C	C--AA	CTCAA	NA	NA	indel
7382	COX2	+	3	T	C	TATCC	TACCC	Tyr	Tyr	t
8245	ATP8	+	3	A	T	CTAAC	CTTAC	Leu	Leu	v
8475	ATP6	+	3	T	C	AATCC	AACCC	Asn	Asn	t
8656	ATP6	+	1	G	A	AAGCA	AAACA	Ala	Thr	t
8748	ATP6	+	3	T	C	CCTAT	CCCAT	Pro	Pro	t
8766	ATP6	+	3	G	A	CGGTT	CGATT	Arg	Arg	t
8818	ATP6	+	1	A	G	GGACT	GGGCT	Thr	Ala	t
8925	ATP6	+	3	T	C	TATGT	TACGT	Tyr	Tyr	t
9545	COX3	+	3	C	T	GTCTA	GTTTA	Val	Val	t
9632	COX3	+	3	T	C	CGTCA	CGCCA	Arg	Arg	t
9713	COX3	+	3	C	T	GTCTG	GTTTG	Val	Val	t
9867	ND3	+	3	T	C	ACTCT	ACCCT	Thr	Thr	t
9909	ND3	+	3	G	A	GTGTA	GTATA	Val	Val	t
10125	ND3	+	3	C	T	AGCCT	AGTCT	Ser	Ser	t
10225	tRNA-Arg	+	NA	C	T	AACTC	AATTC	NA	NA	t
10496	ND4L	+	3	C	T	GGCAC	GGTAC	Gly	Gly	t
10543	ND4	+	3	C	T	ATCAT	ATTAT	Ile	Ile	t
10692	ND4	+	2	C	T	AACAT	AATAT	Thr	Met	t

10828	ND4	+	3	T	C	ATTTT	ATCTT	Ile	Ile	t
10903	ND4	+	3	A	G	ACACT	ACGCT	Thr	Thr	t
11153	ND4	+	1	T	C	CATTA	CACTA	Leu	Leu	t
11305	ND4	+	3	C	T	TACCC	TATCC	Tyr	Tyr	t
11324	ND4	+	1	C	T	CCCTA	CCTTA	Leu	Leu	t
11468	ND4	+	1	G	A	CCGCT	CCACT	Ala	Thr	t
11551	ND4	+	3	C	T	ACCAT	ACTAT	Thr	Thr	t
11581	ND4	+	3	T	C	CTTCT	CTCCT	Leu	Leu	t
12192	ND5	+	1	C	T	AACCT	AATCC	Pro	Ser	t
12194	ND5	+	3	T	C	CCTTC	TCCTC	Pro	Ser	t
12272	ND5	+	3	C	T	ATCCA	ATCA	Ile	Ile	t
12338	ND5	+	3	C	T	TCCCT	TCTCT	Ser	Ser	t
12488	ND5	+	3	C	T	ACCAT	ACTAT	Thr	Thr	t
12542	ND5	+	3	A	G	GAAGG	GAGGG	Glu	Glu	t
12689	ND5	+	3	T	C	AATAC	AACAC	Asn	Asn	t
12701	ND5	+	3	A	T	CTACA	CTTCA	Leu	Leu	v
12744	ND5	+	1	T	C	CCTTG	CCCTG	Leu	Leu	t
12863	ND5	+	3	G	A	ATGGT	ATAGT	Met	Met	t
13040	ND5	+	3	C	T	GGCCT	GGTCT	Gly	Gly	t
13373	ND5	+	3	C	T	TACAG	TATAG	Tyr	Tyr	t
13502	ND5	+	3	G	A	GGGTA	GGATA	Gly	Gly	t
13650	ND5	+	1	C	T	ACCAC	ACTAC	His	Tyr	t
13680	ND5	+	1	C	T	CCCTG	CCTTG	Leu	Leu	t
13709	ND5	+	3	T	C	CATCG	CACCG	His	His	t
13724	ND5	+	3	C	T	TACAT	TATAT	Tyr	Tyr	t
13868	ND5	+	3	T	C	TATTT	TACTT	Tyr	Tyr	t
13898	ND5	+	3	T	C	ATTAG	ATCAG	Ile	Ile	t
14242	ND6	-	3	C	T	ACCAC	ACTAC	Val	Val	t
14350	ND6	-	3	C	T	AACCC	AATCC	Gly	Gly	t
14536	CYTB	+	3	C	T	CACCC	CATCC	His	His	t
14623	CYTB	+	3	A	G	CTAGG	CTGGG	Leu	Leu	t
14637	CYTB	+	2	T	C	AATCC	AACCC	Ile	Thr	t
14713	CYTB	+	3	A	C	GCACA	GCCCA	Ala	Ala	v
14887	CYTB	+	3	T	C	GCTAC	GCCAC	Ala	Ala	t
14989	CYTB	+	3	T	C	AATTT	AACTT	Asn	Asn	t
15200	CYTB	+	1	T	C	TCTTA	TCCTA	Leu	Leu	t
15894	D-loop	+	NA	A	G	TTACA	TGCA	NA	NA	t
15951	D-loop	+	NA	C	T	AACGT	AATAT	NA	NA	t
15952	D-loop	+	NA	G	A	ACGTA	ATATA	NA	NA	t
16034	D-loop	+	NA	G	T	TAGTT	TATCC	NA	NA	v
16035	D-loop	+	NA	T	C	AGTTC	ATCCT	NA	NA	t
16036	D-loop	+	NA	T	C	GTTCC	TCCTC	NA	NA	t
16037	D-loop	+	NA	C	T	TTCCT	CCTCT	NA	NA	t
16045	D-loop	+	NA	T	C	GATAG	GACAG	NA	NA	t
16063	D-loop	+	NA	A	G	AAATT	AAGTT	NA	NA	t
16072	D-loop	+	NA	C	T	ATCGT	ATTGT	NA	NA	t
16105	D-loop	+	NA	A	G	TGACA	TGGCA	NA	NA	t
16126	D-loop	+	NA	C	T	CACTA	CATTA	NA	NA	t
16184	D-loop	+	NA	C	T	GACTC	GATCC	NA	NA	t
16185	D-loop	+	NA	T	C	ACTCC	ATCCC	NA	NA	t
16209	D-loop	+	NA	G	A	TGGAT	TGAAC	NA	NA	t
16211	D-loop	+	NA	T	C	GATTG	AACCG	NA	NA	t
16212	D-loop	+	NA	T	C	ATTGT	ACCGT	NA	NA	t
16274	D-loop	+	NA	T	C	CATCT	CACCT	NA	NA	t
16284	D-loop	+	NA	T	C	CGTCC	CGCCC	NA	NA	t

The Position column indicates the position of the difference with respect to the *B. priscus* sequence; the Strand column, the strand of the feature (irrelevant for intergenes); the Position\_codon column, the position of the polymorphism within the codon (for coding sequences); the Base\_priscus and the Base\_bison columns, the polymorphic bases for the *B. priscus* and *B. bison* sequences, respectively; the Sequence\_priscus and the Sequence\_bison columns, the 5-bp sequences where the polymorphism is located (the polymorphic base is in the middle); and the Type column, the type of difference between the *B. priscus* and the *B. bison* sequences ("t" for transition, "v" for transversion). NA, not applicable.

**S5 Table.** List of the polymorphisms that are specific to the genome of *Bison priscus*.

Position	Feature	Strand	Position codon	Base priscus	Base bison	Sequence priscus	Sequence bison	Amino acid priscus	Amino acid bison	Type
144	D-loop	+	NA	T	C	ACTTT	ACCTT	NA	NA	t
204	D-loop	+	NA	A	G	TTATA	CTGTA,TTGTA	NA	NA	t
461	s-rRNA	+	NA	T	C	CTTTT	CTCTT	NA	NA	t
740	s-rRNA	+	NA	T	C	ATTAA	ATCAA	NA	NA	t
807	s-rRNA	+	NA	G	A	GTGAC	GTAAC	NA	NA	t
1158	s-rRNA	+	NA	T	C	ATTAA	ATCAA	NA	NA	t
1473	l-rRNA	+	NA	C	T	TACCC	TATCC	NA	NA	t
1495	l-rRNA	+	NA	A	G	AAATA	AAGCA	NA	NA	t
1496	l-rRNA	+	NA	T	C	AATAA	AGCAA	NA	NA	t
2054	l-rRNA	+	NA	T	C	CCTAA	CCCAA	NA	NA	t
2562	l-rRNA	+	NA	G	A	TTGAC	TTAAC	NA	NA	t
2581	l-rRNA	+	NA	T	A	AATAA	AAAAA	NA	NA	v
2978	l-rRNA	+	NA	A	G	CAACA	CAGCA	NA	NA	t
3102	ND1	+	3	A	G	ATATT	ATGTT	Met	Met	t
3114	ND1	+	3	T	C	AATAT	AACAT	Asn	Asn	t
3301	ND1	+	1	A	G	CAACC	CAGCC	Thr	Ala	t
3555	ND1	+	3	G	A	TCGGT	TCAGT	Ser	Ser	t
3798	ND1	+	3	G	A	ATGAT	ATAAT	Met	Met	t
3826	ND1	+	1	C	T	TTCTA	TTTTA	Leu	Leu	t
3862	ND1	+	1	T	C	TCTAC	TCCAC	Tyr	His	t
4047	ND1	+	3	A	G	CCACA	CCGCA	Pro	Pro	t
4453	ND2	+	3	A	G	CAATC	CAGTC	Gln	Gln	t
4733	ND2	+	1	C	T	CCCTA	CCTTA	Leu	Leu	t
4754	ND2	+	1	G	A	TAGTT	TAATT	Val	Ile	t
4852	ND2	+	3	C	T	TACAA	TATAA	Tyr	Tyr	t
5107	ND2	+	3	C	T	TTCAT	TTTAT	Phe	Phe	t
5253	ND2	+	2	C	T	AGCCG	AGTCG	Ala	Val	t
5291	ND2	+	1	C	T	TACTA	TATTA	Leu	Leu	t
5374	intergene	NA	NA	T	C	TGTTA	TGCTA	NA	NA	t
5614	tRNA-Cys	-	NA	A	G	AGAGC	AGGGC	NA	NA	t
6283	COX1	+	3	C	T	CTCCC	CTTCC	Leu	Leu	t
6301	COX1	+	3	C	T	GGCAT	GGTAT	Gly	Gly	t
6469	COX1	+	3	T	C	TATTC	TACTC	Tyr	Tyr	t
6634	COX1	+	3	C	A	ACCGG	ACAGG	Thr	Thr	v
7382	COX2	+	3	T	C	TATCC	TACCC	Tyr	Tyr	t
8245	ATP8	+	3	A	T	CTAAC	CTTAC	Leu	Leu	v
8475	ATP6	+	3	T	C	AATCC	AACCC	Asn	Asn	t
8748	ATP6	+	3	T	C	CCTAT	CCCAT	Pro	Pro	t
8925	ATP6	+	3	T	C	TATGT	TACGT	Tyr	Tyr	t
9545	COX3	+	3	C	T	GTCTA	GTTTA	Val	Val	t
9632	COX3	+	3	T	C	CGTCA	CGCCA	Arg	Arg	t
9713	COX3	+	3	C	T	GTCTG	GTTTG	Val	Val	t
9867	ND3	+	3	T	C	ACTCT	ACCCT	Thr	Thr	t
9909	ND3	+	3	G	A	GTGTA	GTATA	Val	Val	t
10125	ND3	+	3	C	T	AGCCT	AGTCT	Ser	Ser	t
10225	tRNA-Arg	+	NA	C	T	AACTC	AATTC	NA	NA	t
10496	ND4L	+	3	C	T	GGCAC	GGTAC	Gly	Gly	t
10692	ND4	+	2	C	T	AACAT	AATAT	Thr	Met	t
10828	ND4	+	3	T	C	ATTTT	ATCTT,ATCTC	Ile	Ile	t
10903	ND4	+	3	A	G	ACACT	ACGCT	Thr	Thr	t
11153	ND4	+	1	T	C	CATTA	CACTA	Leu	Leu	t
11305	ND4	+	3	C	T	TACCC	TATCC	Tyr	Tyr	t
11324	ND4	+	1	C	T	CCCTA	CCTTA	Leu	Leu	t
11551	ND4	+	3	C	T	ACCAT	ACTAT	Thr	Thr	t
11581	ND4	+	3	T	C	CTTCT	CTCCT	Leu	Leu	t
12192	ND5	+	1	C	T	AACCT	AATCC	Pro	Ser	t
12194	ND5	+	3	T	C	CCTTC	TCCTC	Pro	Ser	t
12272	ND5	+	3	C	T	ATCCA	ATTCA	Ile	Ile	t
12338	ND5	+	3	C	T	TCCCT	TCTCT	Ser	Ser	t
12488	ND5	+	3	C	T	ACCAT	ACTAT	Thr	Thr	t
12542	ND5	+	3	A	G	GAAGG	GAGGG	Glu	Glu	t

12701	ND5	+	3	A	T	CTACA	CTTCA	Leu	Leu	v
12744	ND5	+	1	T	C	CCTTG	CCCTG	Leu	Leu	t
12863	ND5	+	3	G	A	ATGGT	ATAGT	Met	Met	t
13040	ND5	+	3	C	T	GGCCT	GGTCT	Gly	Gly	t
13373	ND5	+	3	C	T	TACAG	TATAG	Tyr	Tyr	t
13502	ND5	+	3	G	A	GGGTA	GGATA	Gly	Gly	t
13650	ND5	+	1	C	T	ACCAC	ACTAC,ACTAT	His	Tyr	t
13680	ND5	+	1	C	T	CCCTG	CCTTG	Leu	Leu	t
13709	ND5	+	3	T	C	CATCG	CACCG	His	His	t
13724	ND5	+	3	C	T	TACAT	TATAT	Tyr	Tyr	t
13868	ND5	+	3	T	C	TATTT	TACTT	Tyr	Tyr	t
13898	ND5	+	3	T	C	ATTAG	ATCAG	Ile	Ile	t
14242	ND6	-	3	C	T	ACCAC	ACTAC	Val	Val	t
14350	ND6	-	3	C	T	AACCC	AATCC	Gly	Gly	t
14536	CYTB	+	3	C	T	CACCC	CATCC	His	His	t
14623	CYTB	+	3	A	G	CTAGG	CTGGG	Leu	Leu	t
14713	CYTB	+	3	A	C	GCACA	GCCCA	Ala	Ala	v
14989	CYTB	+	3	T	C	AATTT	AACCT	Asn	Asn	t
15200	CYTB	+	1	T	C	TCTTA	TCCTA	Leu	Leu	t
15894	D-loop	+	NA	A	G	TTACA	TTGCA,CTGCA	NA	NA	t
15951	D-loop	+	NA	C	T	AACGT	AATGT,AATAT	NA	NA	t
16034	D-loop	+	NA	G	T	TAGTT	TATCC	NA	NA	v
16035	D-loop	+	NA	T	C	AGTTC	ATCCC,ATCCT	NA	NA	t
16036	D-loop	+	NA	T	C	GTTCC	TCCTC,TCCCC	NA	NA	t
16063	D-loop	+	NA	A	G	AAATT	AAGTT	NA	NA	t
16072	D-loop	+	NA	C	T	ATCGT	ATTGT	NA	NA	t
16185	D-loop	+	NA	T	C	ACTCC	ACCCC,ATCCC	NA	NA	t
16209	D-loop	+	NA	G	A	TGGAT	TGAAC	NA	NA	t
16211	D-loop	+	NA	T	C	GATTG	AACCG	NA	NA	t
16212	D-loop	+	NA	T	C	ATTGT	ACCGT	NA	NA	t
16274	D-loop	+	NA	T	C	CATCT	CACCT	NA	NA	t
16284	D-loop	+	NA	T	C	CGTCC	CGCCC,CGCTC	NA	NA	t

The compendium was established by comparing SGE2seq to the genome sequences available for 33 modern *Bison bison* specimens. The column names are as in Table S4.



## Résumé

Longtemps utilisés en palynologie pour l'étude des paléo-environnements, les coprolithes, excréments fossilisés, sont également d'importantes sources d'information pour des espèces disparues, producteurs ou proies. Grâce à de nombreux échantillons archéologiques provenant d'une dizaine de grottes, comme la Grotte Chauvet-Pont d'Arc, deux espèces ont pu être étudiées : le loup, *Canis lupus*, et l'hyène des cavernes, *Crocota crocuta spelaea*.

À partir d'un coprolithe de canidé, daté à 34 500 ans, un génome mitochondrial complet de *Canis lupus* a pu être reconstitué. Les analyses phylogénétiques ont montré que ce spécimen se situe en dehors de la diversité des chiens et loups actuels. Puis, les analyses menées sur des gènes nucléaires ont montré que le spécimen de Chauvet ne présente pas de traces évidentes de domestication. L'étude du coprolithe met en évidence un régime carné, avec un bol alimentaire comportant des traces d'ADN d'ours des cavernes, *Ursus spelaeus*.

Dans un second temps, grâce à l'étude de nombreux coprolithes d'hyène des cavernes, une alimentation variée composée d'animaux de grande comme de petite taille a été mise en évidence pour ce carnivore. L'analyse plus fine des séquences d'ADN contenues dans l'un des échantillons a permis de reconstituer un génome mitochondrial complet pour une espèce aujourd'hui éteinte : le bison des forêts, *Bison schoetensacki*. En parallèle, grâce à l'étude d'un ossement de bison des steppes, *Bison priscus*, un génome mitochondrial complet a été obtenu pour cette espèce elle aussi éteinte. L'ajout de ces deux nouvelles séquences mitochondriales à la phylogénie des bovidés a permis d'apporter des éclaircissements à cette dernière.

## Abstract

Coprolites have long been used in palynology for paleoenvironments reconstruction. They also are an important source of information on the DNA of the producing species and its diet. Using numerous archeological samples from several caves, including the Chauvet-Pont d'Arc cave, we studied coprolites for two species: the wolf, *Canis lupus* and the cave hyena, *Crocota crocuta spelaea*.

Using a canid coprolite from the Chauvet cave, dated back to 34 500 years, we obtained a complete mitochondrial genome sequence. Phylogenetic analyses highlight a maternal lineage that positions outside the diversity of extant dogs and wolves. Then, analyzes conducted on the nuclear genes showed that the Chauvet *Canis lupus* specimen does not display obvious indication of domestication. Analysing the coprolite for other species to indicate the diet of this specimen, we detected cave bear (*Ursus spelaeus*) DNA sequences.

Second, using many cave hyena coprolites, a flexible diet consisting of large as well as small animals was demonstrated for this extinct carnivore. Focusing the analysis on a coprolite samples that contained large amounts of bovine DNA, we obtained for the first time a complete mitochondrial genome sequence for the extinct European forest bison, *Bison schoetensacki*. In parallel, a bone sample for the extinct steppe bison provided the first complete mitochondrial genome sequence for *Bison priscus*. These two genome sequences shed new lights on the phylogeny of *Bovinae*.