



HAL
open science

Diagénèse de l'ADN bactérien et analyses métagénomiques de pathologies bactériennes du passé

Olivier Gorgé

► **To cite this version:**

Olivier Gorgé. Diagénèse de l'ADN bactérien et analyses métagénomiques de pathologies bactériennes du passé. Bactériologie. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLS572 . tel-01467294

HAL Id: tel-01467294

<https://theses.hal.science/tel-01467294v1>

Submitted on 14 Feb 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.

NNT : 2016SACLS572

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N° 577
SDSV | Structure et dynamique des systèmes vivants

Spécialité de doctorat : Sciences de la Vie et de la Santé

Par

M. Olivier Gorgé

**Diagnèse de l'ADN bactérien et analyses métagénomiques
de pathologies bactériennes du passé**

Thèse présentée et soutenue à Paris, le 13/12/2016

Composition du Jury :

M. DUBOW, Michael	Professeur, Université Paris Saclay	Président
M. PEYRET, Pierre	Professeur, Université d'Auvergne	Rapporteur
M. SIMONET, Pascal	Professeur, Ecole Centrale de Lyon	Rapporteur
Mme FERNÁNDEZ-JALVO Yolanda	Directrice de recherche, Museum National des Sciences Naturelles de Madrid	Examinatrice
M. VALADE, Éric	Professeur agrégé, Institut de Recherche Biomédicale des Armées	Examineur
M. GRANGE, Thierry	Directeur de recherche, Université Paris Diderot	Examineur
M. MOREL, Yannick	Responsable pôle Sciences de l'Homme et Protection, DGA	Examineur
Mme GEIGL, Eva-Maria	Directrice de recherche, Université Paris Diderot	Directrice de thèse

Écoutant, en effet, les cris d'allégresse qui montaient de la ville, Rieux se souvenait que cette allégresse était toujours menacée. Car il savait ce que cette foule en joie ignorait, et qu'on peut lire dans les livres, que le bacille de la peste ne meurt ni ne disparaît jamais, qu'il peut rester pendant des dizaines d'années endormi dans les meubles et le linge, qu'il attend patiemment dans les chambres, les caves, les malles, les mouchoirs et les paperasses, et que, peut-être, le jour viendrait où, pour le malheur et l'enseignement des hommes, la peste réveillerait ses rats et les enverrait mourir dans une cité heureuse.

Albert Camus – La Peste

REMERCIEMENTS

Je tiens à remercier en premier lieu Pascal Simonet et Pierre Peyret pour m'avoir fait l'honneur d'être rapporteurs de cette thèse, ainsi que Yolanda Fernandez Jalvo, Michael Dubow, Eric Valade et Yannick Morel d'avoir accepté d'examiner mon travail. M. le professeur Simonet, j'ai également eu la chance d'être accueilli dans votre laboratoire et de me former et échanger sur la caractérisation des communautés microbiennes du sol. M. le professeur Peyret, j'ai eu l'opportunité de collaborer avec plusieurs personnes de votre équipe, et j'ai pu apprécier leur compétence et leur disponibilité.

Je souhaite ensuite remercier le Pr. Giuseppe Baldacci, directeur de l'Institut Jacques Monod, pour m'avoir accueilli dans son unité et m'avoir permis d'y réaliser cette thèse. A l'origine de ma « renaissance » à la science, il y a le laboratoire de Gilles Vergnaud à l'IGM d'Orsay. Ces deux années ont été très formatrices, et ce retour à la paillasse a été un point nodal qui me permet de remercier l'équipe GPMS dans cette thèse.

Nombreux sont ceux à remercier, qui ont participé, soutenu ou simplement permis cette étude, l'Ecole du Val de Grâce et ses directeurs successifs, mes responsables à DGA Maîtrise NRBC, Yannick Morel et Sébastien Plumet, et à l'IRBA, Éric Valade. Je remercie les membres des équipes de DGA Maîtrise NRBC, Pascal, Rudy, Catherine, Yann, Françoise, Sabrina, Philippe M. et Philippe L., de l'IRBA, Arnaud, Cécile, Estelle, Marine, Marine, Flora, Laetitia, Fabienne et Estelle, de l'IJM, Andy, Nathalie, Mélanie, Silvia, Giulia, Samantha et Julien. Et tous ceux qui à un moment m'ont apporté un soutien ou une aide, Gilles, Jean-Baptiste, Sandrine, Nicolas, Cyrielle, ...

Les années passées à l'Institut Jacques Monod ont été particulièrement riches, humainement et scientifiquement. Mes plus grands remerciements vont à Thierry Grange et Eva-Maria Geigl. Un « vieux » thésard comme moi pouvait inquiéter, mais Eva-Maria et Thierry m'ont écouté et accueilli. Ces années passées à l'IJM m'ont enrichi, m'ont permis d'acquérir de nouvelles compétences et m'ont donné un recul qui me sera utile pour toutes les années à venir. Je vous suis reconnaissant de la confiance que vous avez mise en moi.

Et surtout merci à ma famille, Edith et nos enfants Cyrille et Tiana. Edith, nous avons initié ce projet ensemble. Dès le début, tu m'as soutenu, tu t'es tenue à mes côtés et nous avons pu ensemble avancer, faire des choix, passer les moments difficiles et tu as toujours été là pour me pousser et m'aider à surmonter les obstacles divers qui se sont présentés. Cela a été plus long que prévu, plus compliqué, plus difficile, mais ça y est, nous y sommes, ensemble. Sans ton aide, sans ta détermination et sans ton courage, je n'y serais pas arrivé. Merci

SOMMAIRE

REMERCIEMENTS	5
SOMMAIRE	7
LISTE DES ABREVIATIONS	11
INTRODUCTION	12
SYNTHÈSE BIBLIOGRAPHIQUE	17
1. Paléogénétique et paléogénomique	17
1.1. Caractéristiques de l'ADN ancien.....	19
1.2. Mise en évidence de l'ADN ancien	20
1.3. Pathogènes anciens.....	22
1.3.1. <i>Yersinia pestis</i>	24
1.3.2. <i>Mycobacterium tuberculosis</i>	31
1.3.3. <i>Mycobacterium leprae</i>	34
1.3.4. <i>Treponema pallidum</i>	35
1.3.5. <i>Bartonella</i>	36
1.3.6. <i>Brucella melitensis</i>	37
1.3.7. <i>Tannerella forsythia</i>	38
1.3.8. <i>Variola major</i>	39
1.3.9. Grippe espagnole	40
1.3.10. Autres signes pathologiques.....	41
2. Étude de la diversité bactérienne	44
2.1. Diversité bactérienne	44
2.2. Classification et identification bactérienne	45
2.3. La culture bactérienne	46
2.4. Identification bactérienne par analyse de l'ADN extrait de l'environnement	46
2.5. Limites de la métagénomique.....	48
3. Taphonomie	49
4. Présentation du projet	52
MATÉRIELS ET MÉTHODES	55
1. Matériels	55
1.1. Echantillons animaux	55
1.1.1. Bure-Bourogne-Vellechevreux	55
1.1.2. Ostheim	56
1.1.3. Les Sablons.....	57

1.1.4.	Bernolsheim.....	57
1.1.5.	Horbourg-Wihr.....	57
1.1.6.	Muttersholtz.....	57
1.1.	Echantillons humains.....	59
1.1.1.	Artzenheim.....	59
1.1.2.	Kolbsheim.....	60
1.1.3.	Boulogne-sur-Mer.....	60
1.1.4.	Saint-Laurent-de-la-Cabrerisse.....	61
1.1.5.	Tartres des marins du Mary-Rose.....	63
2.	Méthodes.....	65
2.1.	Méthodes d'extraction, de purification et de réalisation de banques.....	65
2.1.1.	Echantillons anciens.....	65
2.1.2.	Protocole d'extraction d'ADN à partir d'os frais.....	67
2.2.	Analyses taphonomiques.....	67
2.2.1.	Protocole d'injection.....	67
2.2.2.	Protocole d'immuno-suppression.....	67
2.2.3.	Protocole expérimental.....	68
2.2.4.	Détermination de la communauté bactérienne cultivable.....	70
2.2.5.	Suivi de la dégradation de l'ADN.....	70
2.3.	Recherche de pathogènes animaux par qPCR.....	72
2.3.1.	Recherche de <i>Bacillus anthracis</i>	72
2.3.2.	Recherche de <i>Mycobacterium tuberculosis</i>	72
2.4.	Recherche de spores sur ossements.....	73
2.5.	Sélection des amorces 16S.....	76
2.6.	Bio-informatique.....	79
2.6.1.	FastQC.....	80
2.6.2.	leeHom.....	80
2.6.3.	PandaSeq.....	80
2.6.4.	Cutadapt.....	80
2.6.5.	FastX toolkit.....	81
2.6.6.	BWA.....	81
2.6.7.	Picard.....	81
2.6.8.	MetaPhlAn.....	81
2.6.9.	Samtools.....	82
2.6.10.	Krona.....	82
2.6.11.	usearch.....	82
2.6.12.	mapDamage.....	82
2.6.13.	Scripts.....	83
2.6.13.1.	fastq_trimmer_by_quality.py.....	83
2.6.13.2.	GC-boxplot.....	83

2.6.14. Pipelines bioinformatiques (annexe).....	83
TAPHONOMIE – RÉSULTATS ET DISCUSSION	85
1. Validation de la présence d'ADN bactérien dans les os.....	85
2. Validation du protocole de septicémie	85
3. Protocole taphonomie à court terme	87
3.1. Suivi de la température	88
3.2. Observations macroscopiques	89
3.3. Observations microscopiques	93
3.4. Détermination de la communauté bactérienne	98
3.5. Devenir de l'ADN murin	100
3.6. Analyse des Acides nucléiques par séquençage massif.....	105
3.6.1. Analyse du fragment HVR-5.....	107
3.6.2. Analyse du fragment HVR-345	110
3.6.3. Etude de l'activité métabolique bactérienne :.....	114
4. Protocole de taphonomie à moyen terme.....	121
4.1. Suivi de la température	121
4.2. Observations macroscopiques	121
4.3. Evolution de la population bactérienne	123
4.4. Analyse de l'ADN murin	128
4.5. Séquençage massif.....	129
4.6. Recherche de l'ADN exogène.....	137
4.7. Conclusions.....	138
ÉCHANTILLONS ANCIENS - RÉSULTATS ET DISCUSSION	140
1. Protocoles additionnels de décontamination des réactifs.....	141
1.1. Décontamination des AMORCES	141
1.2. Décontamination des réactifs par l'EMA et/ou le PMA	144
2. Échantillons animaux.....	146
2.1. Bure-Bourogne-Vellechevieux.....	146
2.2. Ostheim Birgelsgaerten. Muttersholtz et Les Sablons	149
2.3. Bernolsheim – Horbourg-Wihr	150
3. Echantillons humains.....	151
3.1. Analyse des kystes et restes associés.....	152
3.2. Analyse des Tartres des marins du Mary-Rose.....	157
CONCLUSION.....	172
Index des figures.....	179

Index des tableaux	189
ANNEXE 1.....	1
ANNEXE 2 - protocole de prélèvement.....	10
ANNEXE 3 - scripts.....	12

LISTE DES ABREVIATIONS

16S	16 Sverdberg (constante de sédimentation)
ADN	Acide DésoxyriboNucléique
ARN	Acide RiboNucléique
BLAST	Basic Local Alignment Search Tool
BHI	Bouillon Cœur-Cerveau
bwa	Burrows Wheeler Alignment tool
BSM	Boulogne Sur Mer
CDC	Centers for Diseases Control and Prevention
CEB	Centre d'Etudes du Bouchet (DEVENU DGA Maîtrise NRBC)
contig	Assemblage des séquences obtenues par séquençage haut débit
DGA Maîtrise	Direction Générale de l'Armement Maîtrise Nucléaire, Radiologique,
NRBC	Biologique et Chimique
EDTA	Ethylène Diamine Tetra Acétique
g	gramme
HTS	High Throughput Sequencing
indels	Insertions Délétions
IS	Séquence d'insertion
kb	kilobase
Mb	mégabase
MLST	Multi Locus Sequence Typing
MST	Minimum Spanning Tree
NGS	Next Generation Sequencing
NLS	N Lauryl Sarkosine
OTU	Operational Taxonomic Unit
pb	paire de bases
PCA	Analyse en Composantes Principales
PCoA	Analyse en Coordonnées Principales
PCR	Réaction de Polymérisation en Chaîne
PEG	PolyEthylène Glycol
PLS-DA	Analyse Différentielle des Moindres Carrés Partiels
pH	potentiel d'hydrogène
qPCR	Réaction de Polymérisation en Chaîne quantitative
RDP	Ribosomal Database Project
RT-q-PCR	Réaction de Polymérisation en Chaîne quantitative après rétro-transcription
scaffold	Assemblage de contigs (voir contigs)
SLC	Saint Laurent de la Cabrerisse
SNP	Single Nucleotide Polymorphism
TE	Tris EDTA
TSA	Gélose Trypticase Soja
UV	UltraViolet

INTRODUCTION

Dans son livre, « les bœufs malades de la Peste », François Vallat décrit une épizootie qui frappa l'Europe aux XVIIIe et XIXe siècle, la peste bovine. A cette époque, non seulement les traitements n'existaient pas mais la notion même de pathogène, bactérien ou viral, attendait les travaux de Pasteur et Koch. Les progrès de l'hygiène, décrits par Corbin, étaient encore à venir également. Les maladies étaient des malédictions, qui frappaient les humains et les animaux depuis des temps immémoriaux, déjà décrites dans la Bible comme les 5^{ème} et 6^{ème} plaies d'Egypte. Selon certains auteurs, en particulier Jared Diamond, ces fléaux auraient même été à l'origine de l'effondrement de civilisations entières. Les paléopathologistes s'efforcent de reconstruire l'histoire de ces maladies, de reconnaître, de concert avec les historiens, les maladies qui ont frappé les peuples, à partir de leur description historique, des signes qu'ils ont pu laisser sur les restes humains parvenus jusqu'à maintenant, et depuis une vingtaine d'années grâce à l'ADN ancien.

Les maladies sont une cause majeure de sélection naturelle pour l'espèce humaine, soit au travers de pandémies comme la peste noire ou d'affections chroniques comme la tuberculose. Ainsi, réussir à mettre en évidence des pathogènes anciens permet d'observer les adaptations des pathogènes à leur hôte, leur vitesse d'évolution et de fournir des reconstructions phylogénétiques plus assurées, en fournissant des représentants de nœuds dans les arbres phylogénétiques des espèces pathogènes étudiées.

Les premières expériences visant à mettre en évidence de l'ADN ancien remontent à 1984, quand Russ Higuchi parvint à cloner des portions d'ADN extraites des restes d'un équidé éteint depuis 150 ans, le quagga [1]. L'année suivante, Svante Paabö réussissait à amplifier de l'ADN humain à partir d'une momie de près de 2500 ans [2]. Plusieurs équipes se lancent alors dans l'aventure de l'ADN ancien, sans prêter attention aux nombreuses contraintes liées à la manipulation d'un matériel aussi rare et dégradé, mais présent en quantités phénoménales après amplification PCR. Une fois prouvé que les résultats positifs étaient liés à la présence de contaminants, la grande majorité des articles sera rétractée, le plus fameux étant celui mentionnant la mise en évidence d'ADN de dinosaure par Woodward en 1994 [3].

Après un travail méthodologique visant à fixer les bonnes pratiques pour rechercher de l'ADN ancien, il a été possible d'obtenir des résultats un peu plus fiables issus de l'amplification de l'ADN mitochondrial de restes humains de plus en plus anciens, puis d'amplifier l'ADN mitochondrial néandertalien. Cependant, les capacités de séquençage liées à la PCR et à la technique Sanger n'étaient pas adaptées pour exploiter les plus courts fragments d'ADN, censés être les plus nombreux dans un échantillon ancien, l'ADN étant hydrolysé en fragments de plus en plus courts au fil du temps. Les techniques de séquençage de seconde génération, en particulier Illumina, Helicos et SOLiD, amplifiant de courts fragments, ont permis un progrès considérable des recherches sur l'ADN ancien, avec, comme exemples qui ont le plus changé la vision de l'évolution de l'espèce humaine, le séquençage complet du génome d'un Néandertalien [4], et la découverte d'une nouvelle espèce humaine, contemporaine des Néandertaliens, les Denisoviens [5].

Les progrès ne se sont pas cantonnés aux humains, mais également à la majorité des animaux domestiqués et aux pathogènes. Pour ces derniers, de nombreuses publications discutaient de l'agent pathogène responsable de la peste noire, mais également de la peste justinienne, de la peste d'Athènes voire des affections décrites dans la Bible (Exode 9:1-12, 12:29-36 ; Nombres 16:47, 25 :9). L'étude des maladies à partir de fossiles, la paléopathologie, est un champ de recherche qui s'intéresse aux maladies anciennes en associant médecine, anthropologie, archéologie et histoire. Les premiers scientifiques ayant contribué au développement de cette discipline aux XIX^e et XX^e siècle furent le médecin et préhistorien belge Philippe-Charles Schmerling, qui s'est intéressé aux ossements d'animaux, et le médecin bactériologiste français Marc Armand Ruffer. Ce dernier a découvert un œuf de parasite dans un rein de momie, puis décrit une spondylodiscite chez une autre momie, diagnostiquant le mal de Pott, forme ostéo-articulaire de la tuberculose [6].

La paléopathologie s'est longtemps cantonnée à la description de malformations osseuses pouvant être liées à des pathologies traumatiques, à des pathologies articulaires ou squelettiques, ou encore à des maladies infectieuses. L'examen de tissus mous de momies ou de restes préservés dans le permafrost, de coprolithes ou de sédiments provenant de latrines anciennes peut aussi apporter des indications sur certaines pathologies (méga-viscères typiques de la maladie de Chagas, observation au microscope d'œufs de parasites gastro-intestinaux [7]...). Cependant, dans de nombreux cas de maladies infectieuses, les signes osseux sont inexistants, les tissus mous non

conservés et les descriptions, imprécises ou incomplètes, peuvent correspondre à différents agents pathogènes. C'est notamment le cas des grandes pestes, du latin *pestis*, « fléau » : celle d'Athènes, en 430 avant notre ère [8], maintenant considérée comme une épidémie de typhus [9], la peste justinienne, entre 541 et 767 [10], et la peste noire [11] (peste bubonique), qui fit 25 millions de victimes à travers l'Europe entre 1347 et 1352, toutes deux attribuées à la bactérie *Yersinia pestis*.

Les agents pathogènes ont évolué avec l'humanité depuis son origine. L'utilisation des approches paléogénétiques et paléogénomiques vont enrichir la palette des outils de la paléopathologie et permettre de dresser des comparaisons entre formes modernes et anciennes des pathogènes, de valider des hypothèses sur la coévolution hôte-pathogène, sur les modifications génétiques induites lors d'un changement d'hôte et sur l'adaptation progressive d'un agent pathogène à son nouvel hôte, plus particulièrement lors du passage d'une infection de l'animal vers l'être humain. Leur application rend aussi possible l'étude de la variation des allèles et de leur fréquence au cours du temps, en lien avec des épidémies ou des changements environnementaux [12].

Des lésions osseuses pathognomoniques de plusieurs maladies infectieuses ont depuis longtemps été décrites. La recherche de traces génétiques d'un agent pathogène s'est donc naturellement concentrée sur ce type d'échantillons, la probabilité de détecter l'ADN ancien du microbe y étant plus élevée. Les premières publications sur la question remontent à 1993, avec l'analyse de marqueurs spécifiques de *Mycobacterium tuberculosis* dans des restes osseux de différentes périodes portant des stigmates de tuberculose osseuse [13]. Une autre piste pour identifier les ossements susceptibles de contenir des ADN de micro-organismes pathogènes anciens est d'explorer des sépultures associées à des épidémies, comme en France avec les travaux de Didier Raoult et de ses collaborateurs qui ont mené à la détection de *Y. pestis* dans des dents extraites de fosses communes datant des XVI^e et XVIII^e siècles [14]. Enfin, l'analyse de tissus mous issus de momies ou de sépultures conservées dans le permafrost a également permis d'identifier d'autres agents pathogènes [15].

En suivant ces deux pistes, plusieurs équipes de recherche ont mis en évidence *M. tuberculosis* dans des os et des momies, puis d'autres organismes pathogènes tels que les bactéries *M. leprae* (lèpre) [16] et *Corynebacterium diphtheriae* (diphthérie) [17], ou encore le parasite *Plasmodium falciparum* (paludisme) [18, 19]. De nombreux résultats sont néanmoins toujours l'objet de polémiques du fait de lacunes méthodologiques (protocoles de décontamination, spécificité des cibles génétiques...). Un article publié en

2012 par deux chercheurs du Leibniz Institute for Zoo and Wildlife Research et de l'Université de Berlin, analysant l'ensemble des publications ayant trait à l'ADN ancien de pathogènes et au respect des bonnes pratiques permettant d'éviter la contamination, n'a ainsi retenu pour les pathogènes bactériens que deux études aux résultats solides, dont l'une mettant en évidence une absence d'ADN de *M. tuberculosis* dans les ossements présentant des lésions [20].

L'avancée suivante date de 2011, avec la publication par l'équipe de Johannes Krause du génome de la souche de *Y. pestis* responsable de la peste noire [21]. Il devient ainsi possible de comparer une souche virulente de plus de 600 ans aux souches modernes pour tenter de trouver des explications à cette pandémie qui causa la mort d'un tiers de la population européenne. La souche séquencée se révèle étonnamment moderne, avec seulement 72 polymorphismes simple-nucléotide par rapport à la souche CO92 de référence (isolée en 1992 au Colorado). Ainsi cette publication, premier exemple de séquençage d'un génome pathogène ancien, signe l'avènement de la paléopathogénomique. En 2013, un second article publié par la même équipe dévoile le génome d'une souche médiévale de *M. leprae*, agent pathogène responsable de la lèpre [22]. Une nouvelle fois, le génome apparaît comme très proche de celui des souches modernes. Ces deux publications semblent indiquer que la chute de la prévalence de ces deux agents pathogènes en Europe fut davantage liée à des facteurs sociétaux, à une adaptation de l'espèce humaine ou à des co-infections, qu'à une atténuation de la virulence.

En parallèle, les limites de détection continuent de reculer avec la publication récente d'une étude portant sur la peste justinienne, qui a frappé et durablement affaibli l'Empire byzantin du V^e au VIII^e siècle [23].

La paléovirologie est bien moins développée et les publications rares. Ceci est lié à la nature fragile des virus et à la difficulté à les mettre en évidence. Les rétrovirus ont bien été étudiés via la recherche de traces d'insertions anciennes. Mais outre le fait qu'il ne s'agisse pas de matériel génétique ancien, la mise en évidence est indirecte et ces rétrovirus, intégrés depuis longtemps, ont vu leurs séquences se modifier au cours de l'histoire. La description en 1999 du virus HTLV1 chez une momie andine a ainsi été réfutée. En revanche, en 2005, le génome de la grippe espagnole a été reconstruit à partir de prélèvements anatomo-pathologiques conservés dans la paraffine et de tissus prélevés chez un Eskimo décédé de la grippe, enterré dans le permafrost [24]. En 2012, des séquences d'ADN appartenant à l'un des virus les plus dangereux pour l'espèce

humaine, la variole, ont également été obtenues à partir de prélèvements sur des momies sibériennes [25].

Néanmoins, les phénomènes biologiques et biochimiques permettant la préservation à long et très long terme des acides nucléiques sont encore largement méconnus, et il semble de plus y avoir des différences selon l'agent pathogène considéré. Il a été possible de mettre en évidence un ADN très bien conservé et en quantité de *M. leprae* [22] et d'en séquencer le génome alors qu'il n'y a pas de publication sur *Treponema pallidum*, bactérie responsable de la syphilis, malgré de nombreuses tentatives. De même, le devenir du microbiote, et des bactéries pathogènes éventuellement présentes dans les tissus d'un individu, est mal connu.

Ce travail a cherché à mieux comprendre les phénomènes taphonomiques en se focalisant sur le devenir des bactéries présentes dans la dépouille, appartenant au microbiote endogène ou exogène, et analyser l'importance de la communauté bactérienne tellurique dans les processus de la dégradation du corps au cours des premières années après la mort. A cette fin, nous avons suivi l'évolution de carcasses de souris, enfouies ou laissées en surface, infectées ou non par des bactéries. La communauté bactérienne aérobie et anaérobie dans le sol environnant la dépouille a été explorée, l'évolution de la quantité d'ADN bactérien présent au moment de l'enfouissement analysée, tout comme l'évolution de la quantité et de la longueur de l'ADN murin tout au long de l'expérience. De plus, les populations bactériennes colonisant la dépouille ont été caractérisées par analyse métagénomique.

Prenant en compte ces données, nous avons cherché à mettre en évidence des traces de pathogènes bactériens dans des restes animaux et humains anciens en utilisant la paléogénomique et la paléogénétique. Cette recherche s'est basée essentiellement sur des techniques de biologie moléculaire, en particulier qPCR et séquençage massif.

Les difficultés inhérentes à la recherche d'ADN dans les échantillons anciens, ainsi que l'état des connaissances sur les pathogènes anciens, sont décrits dans le chapitre suivant.

SYNTHÈSE BIBLIOGRAPHIQUE

1. PALEOGENETIQUE ET PALEOGENOMIQUE

L'aventure de l'ADN ancien (aDNA) commence dans le milieu des années 80, quand successivement Higuchi [1] puis Pääbo [2] réussissent à amplifier de l'ADN présent respectivement dans des tissus de Quagga et de momie égyptienne. Ces travaux ouvraient le champ de l'étude de l'ADN ancien, qui depuis a donné lieu à 4037 publications (source PubMed au 10/09/2016), en croissance régulière (présentée en Figure 1).

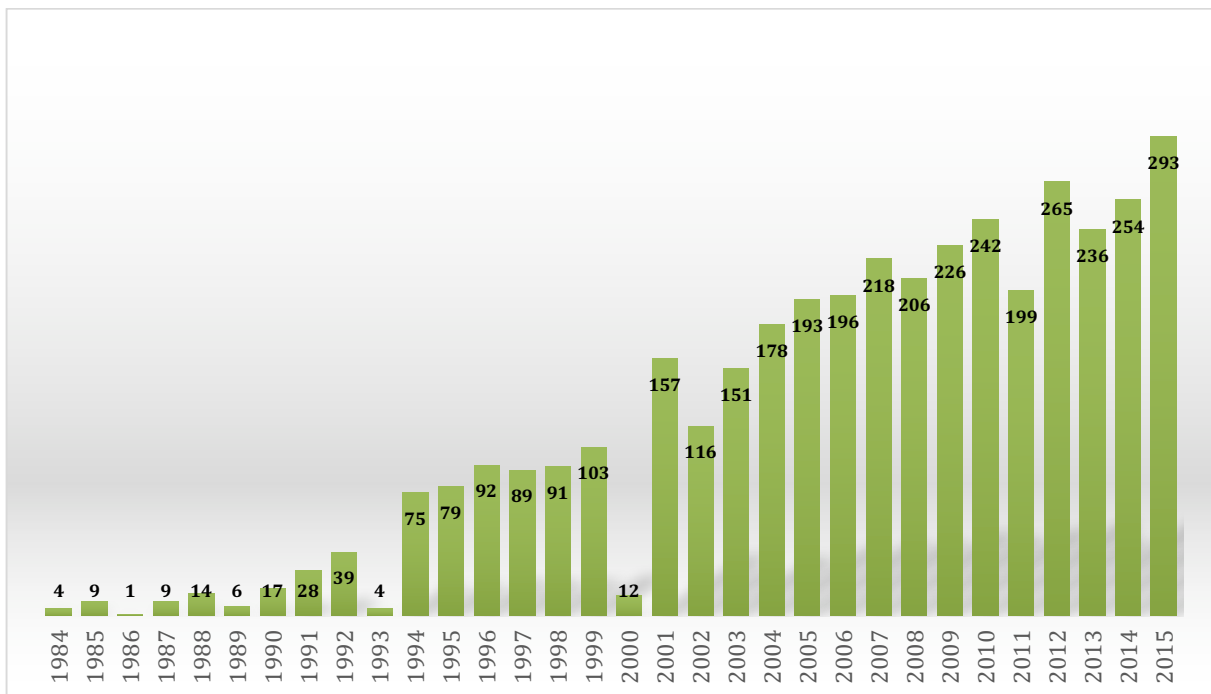


Figure 1 : nombre d'articles portant sur l'ADN ancien entre 1984 et 2015. Source Pubmed, avec une requête simple « ancient[All Fields] AND "dna"[MeSH Terms] OR adna[All Fields] ».

L'ADN conservé dans des échantillons anciens (en général des ossements) permet d'accéder à l'information génétique d'une période temporelle connue, dans un contexte paléo-écologique donné. Ce type d'informations permet de retracer l'évolution des espèces animales, espèce humaine comprise, et de connaître la biodiversité passée. Les données paléogénétiques fournissent une sorte de « photographie génétique » de moments précis de l'évolution tout en évitant les biais et les imprécisions liés à l'extrapolation dans le passé des données génétiques modernes (calibration de l'horloge moléculaire). L'avènement de la PCR (Polymerase Chain Reaction [26]) a permis l'obtention d'informations génétiques à partir d'os et de dents vieux de 100.000 ans

(pour une revue de la littérature, voir [27, 28]). La phylogénie d'animaux disparus, comme l'ours brun du Pléistocène [29], l'aigle géant de la Nouvelle Zélande [30], et le bison des steppes [31] a pu être décrite, ainsi que l'influence du réchauffement climatique sur les mammifères de l'Arctique [32]. Willerslev et collaborateurs ont réussi à reconstituer l'environnement du Groenland il y a 400.000 ans [33].

Avec l'avènement des séquenceurs de seconde génération, une grande partie de l'information génétique contenue dans un échantillon est devenue accessible. Cette capacité a entraîné un grand intérêt dans la communauté de l'ADN ancien qui s'est rapidement approprié ces nouvelles technologies. Elles ont permis depuis 2006 le séquençage des génomes du mammoth [34], du Saqqaq, un Paléo-eskimo [35], du Néandertalien [36], et d'un hominidé jusqu'alors inconnu, le Denisovien [37]. En parallèle, le séquençage de génomes modernes élargissait les possibilités de comparaison entre données modernes et anciennes.

Hormis les études réalisées sur du matériel exceptionnellement bien conservé, comme les poils ou les cheveux, une difficulté majeure de l'analyse paléogénomique d'os fossiles réside dans le fait qu'une proportion très importante (pouvant aller de 90 à 99%) des séquences obtenues provient d'organismes de l'environnement [36], principalement les bactéries qui ont contaminé les os aux différentes périodes de leur transformation (qu'on appelle dans ce cas, la diagénèse). Il est difficile de déterminer le moment où les bactéries à l'origine de ces fragments de génomes ont contaminé l'os, soit au moment de la décomposition du cadavre, soit tout au long de la période qui a précédé le prélèvement, voire à la suite du prélèvement lui-même. Par contre, dans quelques cas, il a aussi été possible de mettre en évidence par amplification par PCR la présence de traces d'ADN provenant de pathogènes qui ont contaminé l'os du vivant de l'organisme comme *Mycobacterium tuberculosis*, [38] et *Yersinia pestis* [39]. L'application des technologies NGS à ces échantillons a permis de reconstituer le génome de la souche responsable de la peste noire [21], d'une souche de lèpre [22] et d'une souche de tuberculose précolombienne [40].

L'échange de matériel génétique entre le sol et les restes d'organismes ne se produit pas seulement dans le sens d'un transfert du sol vers les restes, mais aussi des restes vers le sol. En effet, il a été possible de mettre en évidence dans le sol la présence d'ADN provenant d'animaux et de plantes qui ont disparues depuis longtemps, même en

absence de macrofossiles caractéristiques [41, 42]. Cet ADN peut aussi bien provenir de microfossiles non identifiables que d'un transfert direct à partir des cadavres ou de fèces en décomposition. Il a en effet été démontré que l'ADN peut s'adsorber sur les sédiments, ce qui pourrait favoriser sa préservation à long terme [43-45].

1.1. Caractéristiques de l'ADN ancien

L'ADN ancien est fortement modifié et dégradé par des processus biologiques, chimiques et physiques post-mortem susceptibles qui plus est d'interférer avec les techniques classiques de biologie moléculaire. Cela explique pourquoi les principales études basées sur l'ADN ancien utilisent des spécimens conservés dans le pergélisol (Paléo-eskimo, mammoth) ou dans des grottes de régions froides (Néandertaliens). La température est le seul facteur clairement identifié comme jouant un rôle majeur pour la conservation et la dégradation de l'ADN [46]. D'autres paramètres, comme l'humidité pourrait entraîner une accélération des processus de dégradation de l'ADN, en particulier ceux liés à l'attaque bactérienne et à l'hydrolyse.

Ainsi, la chaîne osidique de l'ADN va être hydrolysée, entraînant une fragmentation en morceaux de plus en plus petits. L'ADN subit également des dépurinations, entraînant la création de sites abasiques, bloquant la majorité des Taq polymérases. Cette rupture de la chaîne N-osidique fragilise le squelette phosphate-ose, ce qui est démontré par l'enrichissement relatif en guanine et en adénine en 5' des fragments d'ADN ancien [47]. Cette dépurination est un second facteur de fragmentation de l'ADN.

Dans la Figure 2 sont représentées les distributions de taille de fragments d'ADN de mammoth, amplifiés à partir de banques préparées selon deux techniques pour être séquencées sur Ion Torrent PGM. Dans les deux cas, on voit que les tailles de fragments sont inférieures à 160 pb, avec un pic de la courbe de distribution à 40 pb pour la méthode simple-brin et 50 pb pour la méthode double-brin.

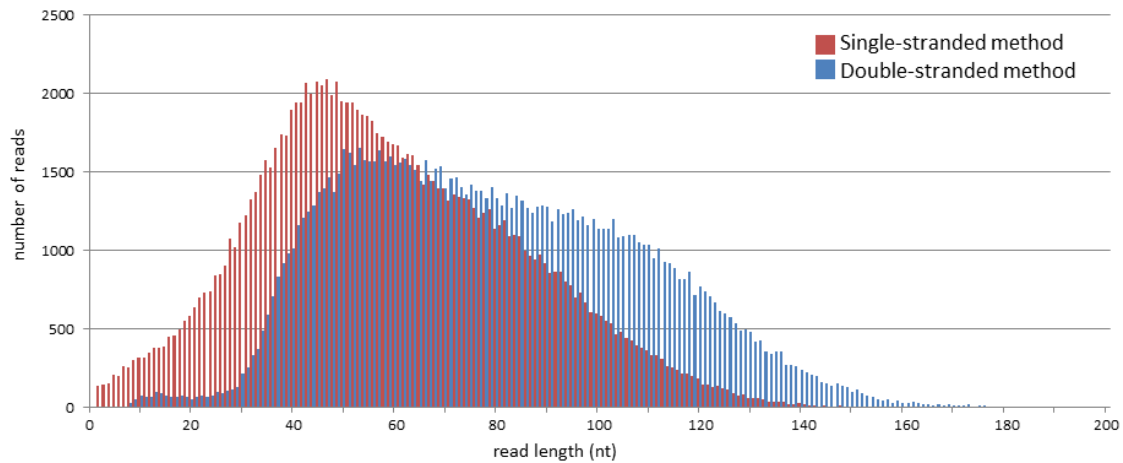


Figure 2 : distribution de la taille des fragments amplifiés à partir d'un os de mammoth. Les banques ont été préparées à partir de la technique double-brin (en bleu) et simple-brin (en rouge). D'après Bennett et al. [48].

La désamination est également rencontrée, et touche préférentiellement la cytosine, qui est alors transformée en uracil, ce qui entraîne des erreurs de séquençage.

Cette désamination se produit essentiellement aux extrémités des fragments d'ADN [47], et est utilisé comme critère d'authenticité lors de l'analyse de données [49, 50]. Ce phénomène est clairement identifié en Figure 3, avec une augmentation de fréquence de thymine à l'extrémité 5' du fragment analysé et par conséquent une augmentation d'adénine en 3'.

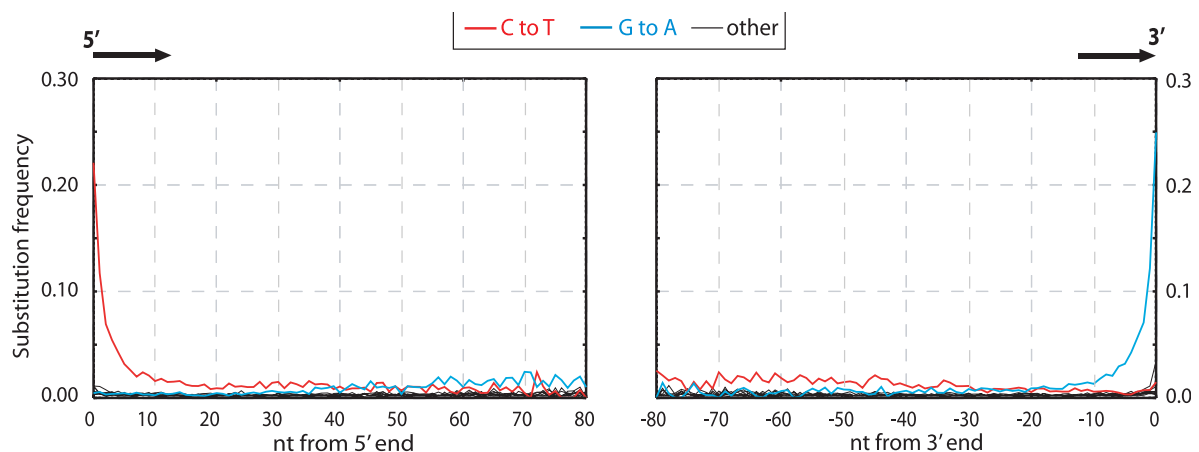


Figure 3 : distribution de la fréquence des substitutions de base dans un ADN ancien (ici Neandertal). D'après Briggs et al. [47].

1.2. Mise en évidence de l'ADN ancien

La méthode de choix, voire l'unique méthode pour mettre en évidence de l'ADN ancien a été longtemps la PCR, qui a rapidement succédé au clonage mis en œuvre par les pionniers du domaine Higuchi et Paabo.

L'ADN ancien n'est pas un substrat optimal pour la PCR du fait des modifications sur les bases (erreur de lecture), des sites abasiques ou des ruptures de la chaîne osidique sur un seul brin (« nicks ») qui bloquent la majorité des polymérase. La fragmentation en soi n'est pas un problème, la PCR fonctionnant avec des fragments de 40bp, mais à moins de rechercher des SNPs, de si petits fragments n'apportent pas d'informations de séquence, les fragments obtenus étant essentiellement constitués des amorces utilisées lors de la PCR.

La contamination par de l'ADN moderne est une plaie majeure de l'analyse paléogénétique qui a déjà failli tuer le domaine au début de son expansion et nécessité la mise en place de procédures très strictes [51]. Toutefois, le respect des procédures décrites ne garantit pas ou pas totalement l'authenticité des résultats et de nombreuses études ne satisfont pas en totalité à ces exigences strictes, ce qui fait qu'il est fréquent de voir mettre en doute des résultats publiés, parfois même dans les meilleurs journaux. Un article publié en 2012 analysant l'ensemble des publications ayant trait à l'ADN ancien de pathogènes et au respect des bonnes pratiques permettant d'éviter la contamination, n'a ainsi retenu pour les pathogènes bactériens que deux études aux résultats solides, dont l'une mettant en évidence une absence d'ADN de *M. tuberculosis* dans les ossements présentant des lésions [20].

Au sein du laboratoire d'accueil de cette thèse, des efforts très importants ont été menés pour maîtriser le problème de la contamination de la manière la plus stricte, rigoureuse et exigeante qu'il soit, afin de pouvoir traiter des échantillons issus de régions géographiques et d'époques défavorables pour la préservation de l'ADN, particulièrement l'expansion néolithique en Asie du Sud-Ouest et en Europe du Sud avec un accent particulier mis sur la domestication des bovins [52-54] ainsi que des équidés [55, 56]. La contamination peut intervenir à toutes les étapes du processus, et peut avoir des sources différentes plus ou moins gênantes selon le matériel génétique étudié. Les organismes posant le plus de problèmes sont l'humain qui est impliqué dans toutes les étapes; les animaux domestiques, et particulièrement les bœufs et les cochons, car ils constituent une source majeure dans le processus de production des réactifs de Biologie Moléculaire, comme l'albumine bovine et la gélatine qui sont utilisés de façon quasi systématique comme stabilisant des enzymes; et les bactéries environnementales qui peuvent aussi être introduites à toutes les étapes. Des procédures optimisées ont été

développées pour chaque type de réactif utilisé en quantifiant à la fois l'efficacité de la décontamination et l'effet délétère potentiel de la procédure suivie sur l'activité du réactif [57]. Un laboratoire de haut confinement a été créé et des procédures très strictes de séparation physique des différentes étapes de l'analyse et de décontamination de tous les matériels pénétrant dans le laboratoire de haut confinement ont été mises en œuvre (<http://www.ijm.fr/ijm/plates-formes/pole-paleogenomique-et-taphonomie-moleculaire/>). Des protocoles spécifiques ont été développés pour réaliser les mesures d'analyse quantitative de l'ADN contenu dans les fossiles et l'étendue de l'inhibition de la PCR conférée par l'extrait fossile [58]. Une procédure enzymatique permettant la décontamination des ADN provenant des analyses précédentes, soit produits de PCR, soit ADN plasmidique clonés a été implémentée dans cette procédure quantitative appliquée à l'ADN ancien [59]. Des procédures de prélèvement des os sur les sites ont été développées en collaboration étroite avec les archéologues de terrain. Ces procédures permettent d'assurer à la fois une contamination minimale lors du prélèvement et une bien meilleure conservation du matériel biologique ancien [54]. La combinaison de tous ces protocoles, procédures et techniques a permis à l'équipe d'analyser le matériel génétique préservé dans des os humains d'environ 400.000 ans en zone tempérée [60, 61] ou dans des os animaux de 50.000 ans provenant d'Afrique du Nord [62] et de 400.000 ans provenant de régions tempérées [63].

1.3. Pathogènes anciens

La maladie est une cause majeure de sélection naturelle, au travers d'une pandémie dévastatrice ou d'infections chroniques débilantes. L'étude des pathogènes anciens permet d'améliorer les connaissances sur les interactions hôte-pathogène, l'apparition de nouveaux pathogènes ou l'acquisition de facteurs de virulence. Les échantillons anciens permettent également de calibrer les horloges moléculaires de divers pathogènes, de reconstruire plus finement les arbres phylogénétiques des espèces étudiées et leur dispersion géographique. A titre d'exemple, nous pouvons citer *Bacillus anthracis*, agent responsable du charbon, zoonose et agent bioterroriste. Le cycle de la bactérie passe par une étape tellurique, après décomposition de l'animal décédé de l'infection (Cf. Figure 4). Il n'a pas jusqu'à présent été mis en évidence dans des échantillons anciens.

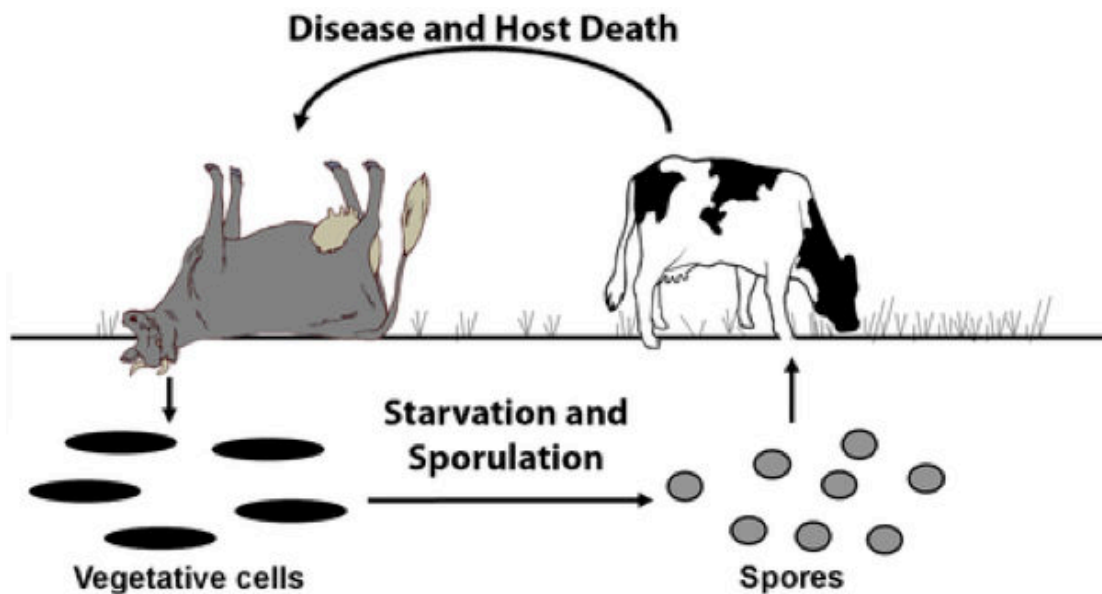


Figure 4 : cycle de vie de *Bacillus anthracis*. Les spores sont ingérées par l'animal alors qu'il broute. Les spores se transforment alors en formes végétatives qui causent une infection fulminante. Après la mort de l'hôte, la décomposition et l'action des charognards, entraîne le relargage de sang infecté dans l'environnement. A l'extérieur de l'organisme, les bactéries sporulent dans l'attente d'un nouvel hôte. Adapté de Schuch et al. [64].

En revanche, lorsqu'on reconstruit son arbre phylogénétique, on observe trois lignées, A, B et C. La lignée C est la plus restreinte, représentant moins d'un pourcent des souches connues. La lignée B est limitée à des zones géographiques telles que l'Afrique australe ou l'Europe. La lignée A est dominante à l'échelle du globe, signant son succès évolutif. De manière très intéressante, on observe à une époque qui pourrait correspondre au Néolithique une radiation massive de la lignée A, dans une fenêtre temporelle correspondant à l'expansion de l'élevage [65].

La concentration élevée d'animaux domestiqués rend plus probable la ré-infection et l'essaimage des troupeaux correspond à un phénomène d'isolation des populations bactériennes, menant à la séparation de populations bactériennes et la création de nouveaux génotypes. Les raisons du succès de la lignée A ne sont pas clairement identifiées, ce succès pourrait n'être lié qu'à l'effet de dispersion dû à l'espèce humaine lors de la progression de l'élevage, plutôt qu'à un avantage évolutif génétique. Identifier une souche de *B. anthracis* ancienne, vieille de plusieurs centaines voire quelques milliers d'années permettrait de confirmer cette hypothèse, préciser et de valider cet arbre phylogénétique, construit uniquement avec des isolats modernes.

Concernant les pathogènes humains, il a été possible de mettre en évidence dans des restes plus ou moins anciens l'ADN de ces pathogènes.

1.3.1. *Yersinia pestis*

La peste fascine les historiens et les chercheurs depuis l'épidémie de peste noire qui a décimé 1/3 de la population européenne entre 1347 et 1350. L'agent causal a longtemps été débattu, jusqu'à ce qu'il soit possible de mettre en évidence des fragments d'ADN et de protéines de *Yersinia pestis* dans des échantillons datant de la période pandémique.

Y. pestis est un petit bacille à Gram négatif, ovoïde, parfois coccobacille, immobile, non sporulé. Il appartient au genre *Yersinia*, qui compte 15 espèces dont trois pathogènes, *Y. enterocolitica*, *Y. pseudotuberculosis* et *Y. pestis*. Ces deux derniers auraient divergé entre 1.500 et 20.000 ans, tandis que la divergence avec *Y. enterocolitica* remonterait entre 41 et 186 millions d'années [66]. *Y. pestis*, l'agent responsable de la peste, a été identifié en 1894 par Alexandre Yersin, à Hong-Kong, au début de la troisième pandémie, encore présente aujourd'hui. Avec une aire de répartition mondiale, il a pour hôte habituel les rongeurs sauvages (rats, marmottes, chiens de prairie...) qui constituent le réservoir naturel. L'humain est contaminé par la piqûre de puces infestant les animaux constituant le réservoir. Le mode de transmission naturel est la piqûre de puces infectées ou à la faveur d'excoriations cutanées, de morsures, de griffures ou en manipulant des animaux infectés. La contamination est possible par inhalation : transmission par voie aérienne de gouttelettes provenant d'humains ou d'animaux de compagnie atteints de peste pulmonaire ; négligence dans la manipulation des cultures de laboratoire. La dose infectieuse est estimée entre 100 à 10.000 bactéries par voie aérienne. La transmission interhumaine par l'intermédiaire des puces de l'humain est possible également.

Historiquement, trois pandémies sont attribuées à la peste. La première est la peste justinienne, qui commença en 541. Cette pandémie est réputé avoir stoppé la reconquête par Constantinople de la partie occidentale de l'Empire romain. Elle a entraîné de forts taux de mortalité, comparables à ceux de la seconde pandémie, la peste noire. Celle-ci tua entre 17 et 28 millions d'Européens, soit entre 30 et 40% de la population totale [67]. Après avoir été signalée en Asie centrale entre 1339 et 1340, la maladie atteint les rivages de la Mer noire en 1346, dans le sillage des armées mongoles. Lors du siège de Caffa, des cadavres de pestiférés sont catapultés par les assaillants

mongols par-dessus les murailles du comptoir génois. La maladie aurait ainsi atteint l'Europe lorsque les navires génois, ayant fui la ville assiégée, firent escale à Messine, Gènes puis Marseille en 1347. De là, elle se serait répandue dans tout le continent (Figure 6) et jusqu'à l'Islande.



Figure 5 : carte de diffusion de la peste noire entre 1347 et 1351 en Europe. D'après [68]

A l'issue de cette première pandémie, des vagues épidémiques allaient traverser l'Europe jusqu'à la fin du XVIIIème Siècle. Les symptômes décrits lors de cette pandémie évoquent dans la majorité des cas la peste bubonique, décrits par exemple dans le Decameron de Bocace. Dans certains cas, les symptômes correspondent à la peste pneumonique, qui peut être contractée auprès de patients souffrant d'une forme septicémique de la peste bubonique. La troisième pandémie commença à la fin du XIXème Siècle. Après des cas sporadiques dans le Yunnan en Chine dès 1840, l'épidémie prend une tout autre envergure quand elle atteint Canton puis Hong-Kong en 1894. C'est à cette occasion qu'Alexandre Yersin, disciple de Louis Pasteur, met en évidence le bacille responsable de l'infection, bacille qui recevra ultérieurement son nom. Hong-Kong est un grand port, et les progrès de la navigation, en particulier à vapeur, vont jouer un rôle déterminant dans la dissémination de la maladie (Figure 6).

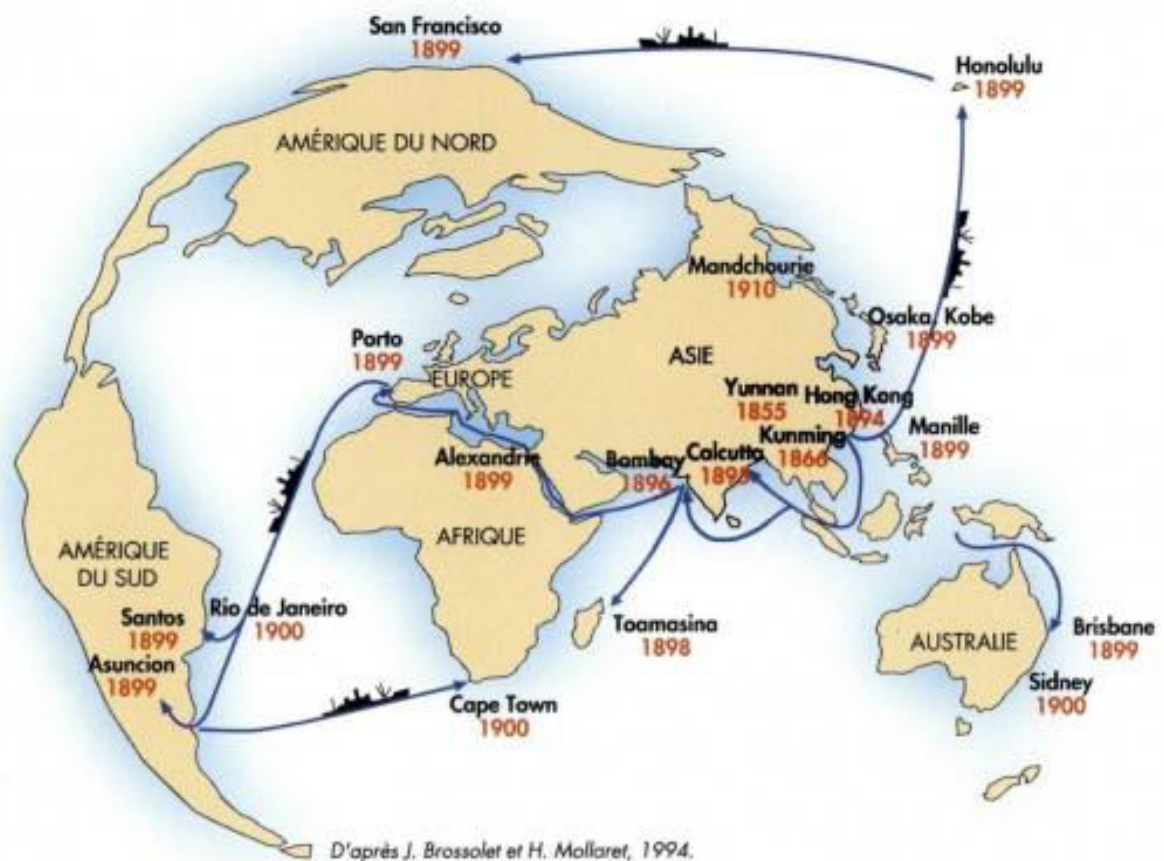


Figure 6 : dissémination de la maladie lors de la troisième pandémie

En six ans, la peste s'est répandue sur les 5 continents. Les progrès de l'hygiène et de la santé permettent de maîtriser et faire reculer cette pandémie, mais elle aurait fait entre 1898 et 1948 plus de 12 millions de morts. Encore considérée active de nos jours, des cas sont observés régulièrement à Madagascar, aux Etats-Unis, en Afrique subsaharienne et en Asie du Sud-Est. En France, la dernière épidémie remonte à 1920 à Paris. Appelée peste des chiffonniers, elle tua 34 personnes sur 96 cas déclarés, essentiellement des personnes vivant dans des conditions de grande insalubrité dans les 18^{ème} et 19^{ème} arrondissements de Paris ainsi que certaines communes limitrophes au Nord.

Y. pestis a été réparti en trois biovars décrits par Devignat [69], appelés *antiqua*, *medievalis* et *orientalis*. A partir de critères biochimiques, la perte de faculté à réduire les nitrates en nitrites entre *antiqua* et *medievalis*, puis la perte de capacité à fermenter le glycérol entre *medievalis* et *orientalis*, Devignat proposa une succession chronologique de ces trois grands biotypes, responsables chacun d'une pandémie. Cette hypothèse n'allait pas résister aux découvertes de l'ADN ancien.

Les études ont porté dans un premier temps sur l'épidémie de peste noire, de 1347 à 1351 en Europe (Figure 5). Les chercheurs se sont attachés à rechercher des traces de

Y. pestis dans des restes osseux issus de cimetières ou fosses communes connues pour avoir recueilli les cadavres de pestiférés (Figure 7). La première publication faisant état de la découverte d'ADN de *Y. pestis* est l'œuvre de Raoult *et al.* [14] en 2000. En extrayant la pulpe de dents de squelettes retrouvés dans une fosse datée de la période de la peste noire à Montpellier, l'équipe a été capable de mettre en évidence une portion de séquence (148bp) du gène *pla* spécifique de *Y. pestis*, porté par un plasmide et présent en de nombreuses copies.

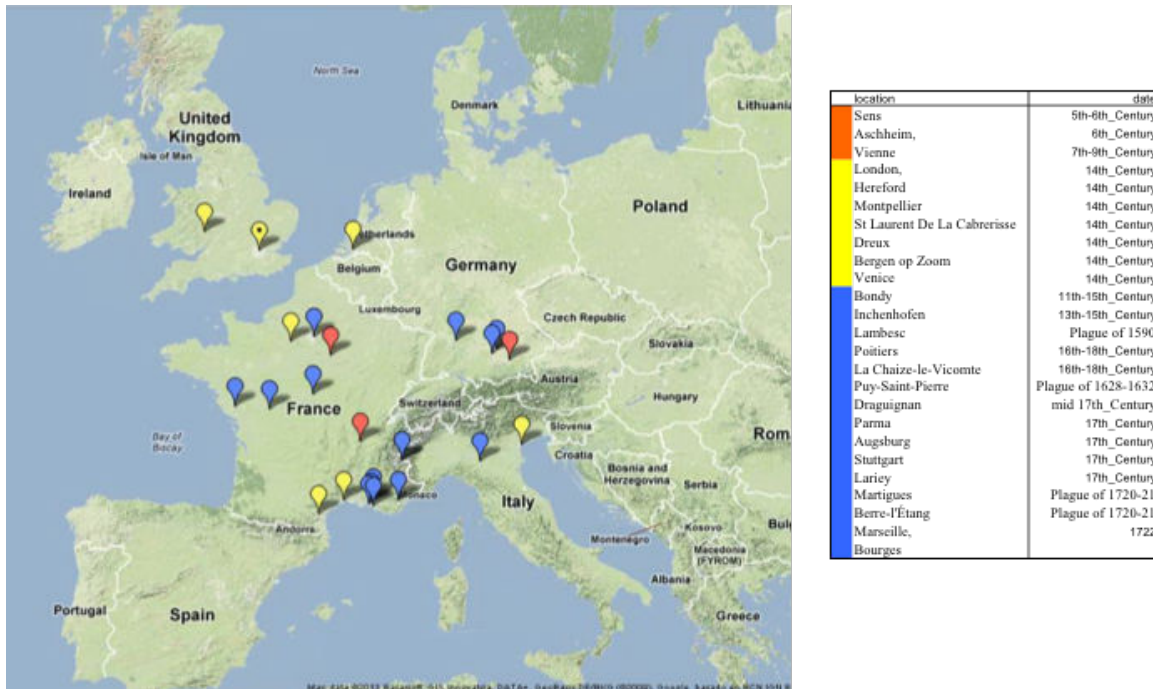


Figure 7 : localisation des sites ayant permis la mise en évidence de traces de *Y. pestis* en Europe. Les sites en orange correspondent à la peste justinienne, les sites en jaune à la peste noire, les sites en bleu à la peste noire ou ses résurgences.

Cette première publication, fort commentée essentiellement de par l'utilisation d'une approche appelée « suicide-PCR » basée sur l'utilisation unique de chaque couple d'amorces, empêchant la répétition des amplifications, a néanmoins été confirmée par des études ultérieures. En 2004, la même équipe proposait que les souches responsables de la peste noire appartenaient au biovar *orientalis*, comme la pandémie moderne, tandis que la peste justinienne était également causée par *Y. pestis*, mais sans parvenir à déterminer son biovar [70]. L'année suivante, en 2005, Wiechmann et Grupe, en utilisant les marqueurs décrits par Raoult *et al.*, mettaient en évidence de l'ADN de *Y. pestis* dans des restes osseux datant de la peste justinienne [71], confortant les découvertes de Drancourt *et al.*

En 2007, Drancourt *et al.* proposaient le biovar *orientalis* comme responsable des trois pandémies [72], en se basant sur l'amplification d'une portion de gène *gldP* porteur

d'une délétion chez *orientalis*, responsable de la perte de capacité à fermenter le glycérol. Les échantillons testés, datant de l'époque de la peste justinienne et de résurgences de la peste noire, ont, pour les positifs, montré cette délétion.

En 2010, en analysant des restes issus de trois sites en Europe et en se basant sur un panel de 76 SNPs identifiés par Achtman *et al.* en 2004 [73], Haensch *et al.* ont réussi à amplifier dans 6 échantillons sur 76 suffisamment de mutations ponctuelles (SNPs) pour situer ces échantillons sur l'arbre phylogénétique proposé par Achtman (Figure 8). Les échantillons, issus des sites anglais et français, se sont placés à la racine des branches 1 et 2 tandis que les échantillons issus du site néerlandais se plaçaient sur la branche 1, menant aux *orientalis* modernes.

Entretiens, de nouveaux génomes de *Y. pestis* modernes étaient séquencés, favorisant leur comparaison et la définition d'un panel de SNPs permettant l'analyse phylogénétique de l'espèce *Y. pestis* avec des marqueurs de très faible taille. En 2010, Morelli *et al.* ont proposé un panel de 933 SNPs sélectionné à partir de 282 isolats de *Y. pestis* [74]. L'analyse comparative de ces SNPs a permis d'établir un arbre phylogénétique couvrant la diversité connue de *Y. pestis* et proposant une origine chinoise pour *Y. pestis*, issue il y a moins de 6.500 ans de *Y. pseudotuberculosis* [73]. Ces SNPs ont été utilisés par Bos *et al.* lorsqu'ils ont séquencé les restes de personnes décédées de la peste noire à Londres. Les auteurs ont pu analyser des squelettes exhumés d'un cimetière (East Smithfield) utilisé uniquement lors de l'épidémie de peste noire. Une première analyse a permis de séquencer le plasmide pPCP1 par capture et enrichissement, puis séquençage sur plate-forme Illumina [75]. Parmi les échantillons donnant le meilleur taux de couverture pour pPCP1, Bos *et al.* [21] ont sélectionné 4 dents. L'ADN a été extrait de la pulpe de ces dents et converti en banques de séquençage. Ces banques ont été enrichies en ADN de *Y. pestis* par capture selon la technique SureSelect d'Agilent, puis séquencées sur Illumina GAI. A l'issue, les auteurs ont pu obtenir une couverture moyenne de 28,2 lectures par site pour le chromosome, et 35,2 et 31,2 respectivement pour les plasmides pCD1 et pMT1. Ce taux de couverture leur a permis de rechercher l'ensemble des SNPs décrits par Morelli *et al.* et de proposer une localisation des souches étudiées sur l'arbre phylogénétique de *Y. pestis* [74]. Il est apparu qu'ils avaient identifié deux souches, différant par 3 SNPs. Quand ces souches sont placées dans l'arbre, la première est localisée à 2 SNPs en aval du nœud séparant les trois branches de l'arbre, et la seconde est plus dérivée, à 5 SNPs de ce nœud, toutes deux sur la branche 1. Ces résultats confirment et renforcent ceux de Haensch, et grâce à

la plus grande résolutivité de leur analyse, permet un placement de ces souches dans la branche 1, à l'origine de l'ensemble des souches *orientalis* connues. Les données de Bos *et al.* ont été reprises par Cui *et al.* en 2012, qui proposent l'arbre phylogénétique le plus détaillé à l'heure actuelle (Figure 9). Cet article montre également la possibilité de séquencer intégralement un génome de pathogène à partir de restes de 650 ans.

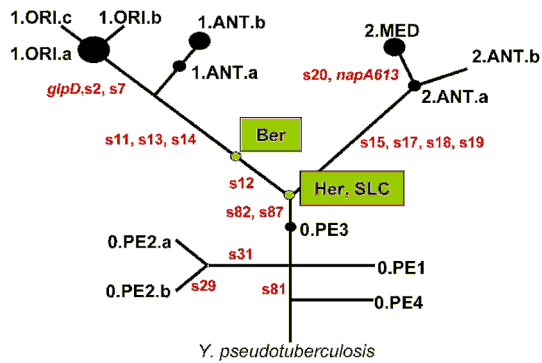


Figure 8 : représentation de la phylogénie de *Y. pestis* basée sur Achtman *et al.* [73]. L'arbre est enraciné dans *Y. pseudotuberculosis* et les souches analysées par Haensch sont représentées par des rectangles verts [76].

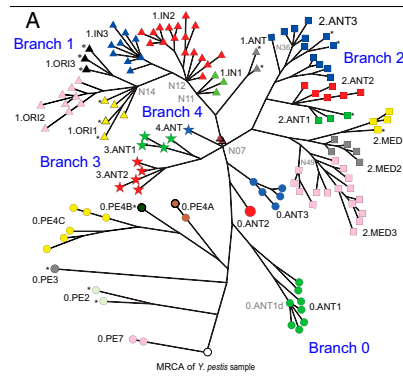


Figure 9 : représentation de la phylogénie de *Y. pestis* basée sur 133 SNPs, Les quatre lignées majeures sont représentées et l'arbre est enraciné dans *Y. pseudotuberculosis*. Les triangles orange près du nœud N07 identifient les souches séquencées par Bos *et al.* D'après Cui *et al.* [11]

En 2013, un nouvel article de Harbeck *et al.* [23] décrit l'utilisation de ces SNPs pour caractériser la souche de peste détectée par Wiechmann et Grupe [71] dans des ossements datant de la première pandémie. En mettant en évidence le statut ancien ou dérivé de 5 SNPs judicieusement choisis, les auteurs ont pu démontrer sur un échantillon que la souche présente était phylogénétiquement plus ancienne que celles responsables de la peste noire (Figure 10). Cette publication vient contredire l'article de Drancourt de 2007 [72], qui identifiait *orientalis* comme responsable des trois pandémies. Ce dernier article avait auparavant été contesté, du fait de l'absence de procédures robustes pour éviter les contaminations et la robustesse de la technique avait elle-même été discutée [77]. L'année suivante, en 2014, la même équipe a publié le génome de *Y. pestis* isolé de deux individus différents [78], obtenu par séquençage à haut débit après enrichissement par la méthode SureSelect d'Agilent. Les séquences obtenues ont permis d'obtenir 91,5% du chromosome avec un taux de couverture de 7,6, tandis que les plasmides étaient entièrement couverts. L'analyse des SNPs produit un schéma comparable à celui proposé l'année précédente, mais le grand nombre de SNPs interrogés permet une localisation plus précise de la souche présente, qui se trouve sur

une nouvelle branche entre les isolats 0.ANT1 et 0.ANT2 (Figure 10).

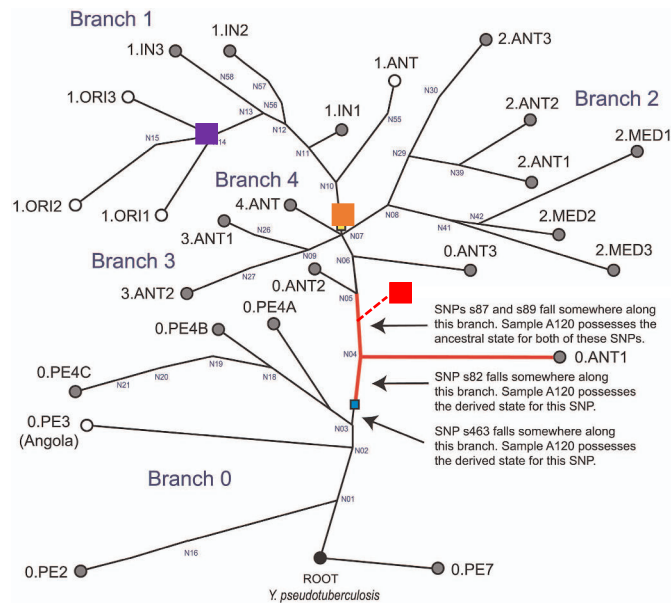


Figure 10 : représentation de la phylogénie de *Y. pestis*, basée sur Cui et al. [11]. Les quatre lignées majeures sont représentées et l'arbre est enraciné dans *Y. pseudotuberculosis*. En rouge sont localisées les souches responsables de la peste justinienne, en orange celles de la peste noire et en violet celles de la pandémie contemporaine. D'après Harbeck et al. [23].

L'ADN ancien a permis, en 15 ans, de répondre à des questions disputées depuis des dizaines d'années par les historiens sur l'agent responsable des deux premières pandémies, mais a aussi permis d'identifier les souches responsables, montrant la possibilité d'émergence de souches hautement pathogènes en plusieurs occasions, précisant ainsi la dynamique d'émergence de ce pathogène (Figure 11).

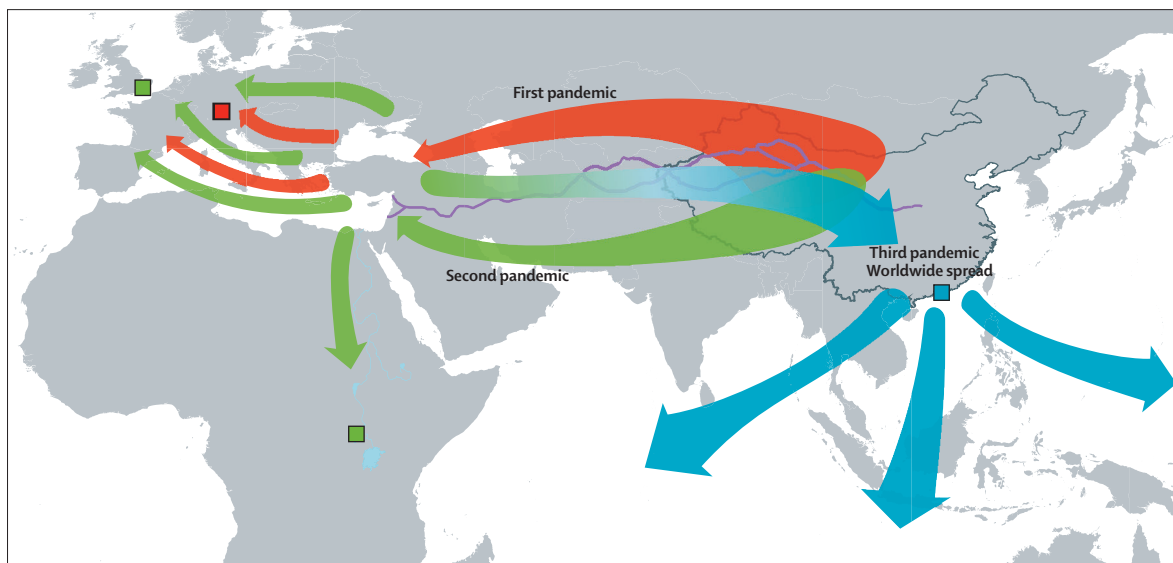


Figure 11 : scénario de dispersion des souches de peste responsables des trois pandémies. D'après Wagner et al. [78].

La lignée associée à la première pandémie aurait émergé en Asie puis atteint l'Europe *via* la route de la Soie, où elle se serait éteinte. Une seconde émergence aurait eu lieu au XIV^{ème} siècle, suivi le même itinéraire pour provoquer la peste noire. Néanmoins, cette lignée se serait établie en Europe, essaimé en Afrique et revenue vers la Chine, toujours *via* la route de la Soie. Elle aurait alors donné naissance à la lignée responsable de la troisième pandémie. Il n'a pas été isolé jusqu'à maintenant de descendants de la souche responsable de la première pandémie.

1.3.2. *Mycobacterium tuberculosis*

Mycobacterium tuberculosis est l'agent de la tuberculose (parfois appelée phtisie). Cette bactérie est aérobic, non mobile et en forme de bâtonnets. Elle est acido-alcali résistant et pousse très lentement. La tuberculose affecte un tiers de la population mondiale et est responsable du décès d'environ 1,5 million de personnes chaque année (données issues du site du CDC, www.cdc.gov/tb, dernière consultation le 21/10/2015). La tuberculose est un problème de santé publique depuis des siècles, comme en atteste cette carte de 1874 montrant la prévalence de la tuberculose aux Etats-Unis (Figure 12).

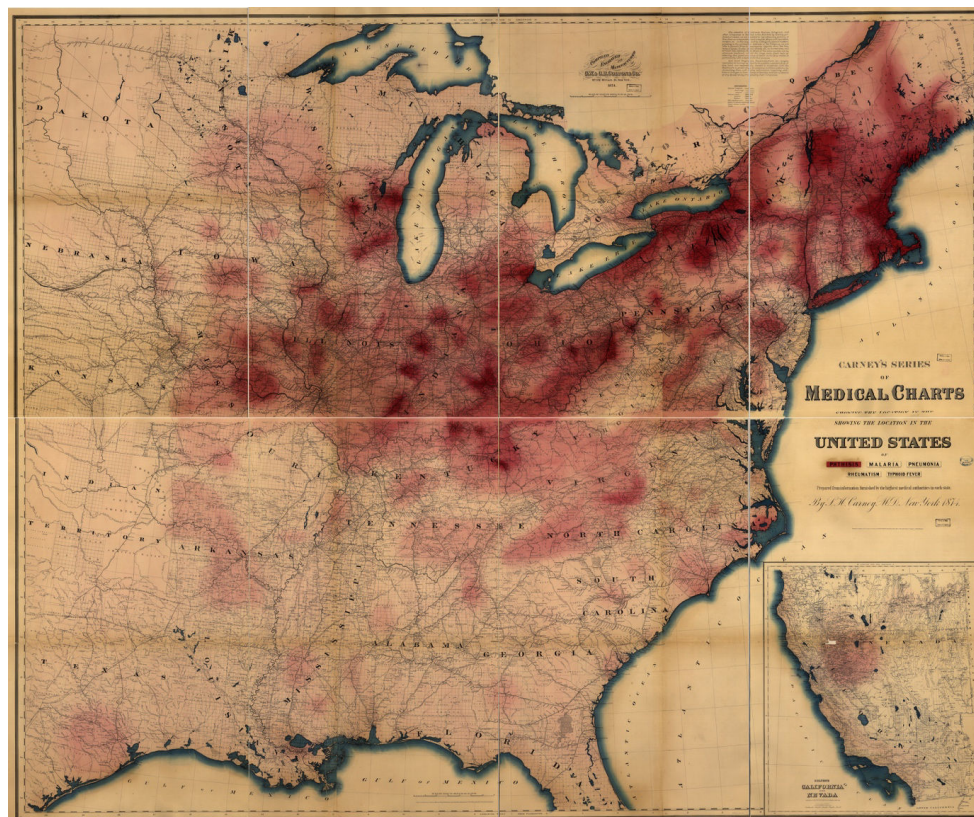


Figure 12 : prévalence de la tuberculose (phtisie) aux Etats-Unis en 1874. L'intensité de la coloration rouge reflète la morbidité [79].

En l'absence de traitement, la maladie peut entraîner la mort. Des sanatoriums ont fleuri entre le 19^{ème} et la première moitié du 20^{ème} pour aider à la guérison spontanée. L'espèce humaine est l'hôte naturel de *M. tuberculosis*, mais la bactérie peut infecter d'autres mammifères. Cette bactérie appartient au MTBC (*Mycobacterium tuberculosis* complex) qui est un groupe de mycobactéries génétiquement proches et capables de provoquer la tuberculose. Ces espèces sont *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. microti*, *M. canettii*, *M. caprae*, *M. pinnipedii*, *M. orygis* et *M. mungi* [80].

Le mode de transmission naturel repose sur les gouttelettes expectorées lorsqu'une personne infectée tousse, se mouche ou parle. La taille de ces gouttelettes leur permet de rester longtemps en suspension dans l'air. Elles contiennent entre 1 et 3 bactéries, ce qui est suffisant pour provoquer une infection chez un nouvel hôte quand elles se déposent au niveau des alvéoles pulmonaires. La bactérie est alors ingérée par un macrophage, dans lequel elle commence à se multiplier, jusqu'à former une lésion primaire par contamination d'autres macrophages. Après un certain temps, des bacilles sont transportés jusqu'aux ganglions lymphoïdes. L'infection devient alors clinique, les symptômes les plus fréquents étant une toux chronique avec éventuellement une hémoptysie, de la fièvre, des suees nocturnes et une perte de poids. La forme pulmonaire de la tuberculose est la plus fréquente, mais il existe des formes extra-pulmonaires, telles que la méningite tuberculeuse, la tuberculose uro-génitale et la tuberculose osseuse ou maladie de Pott. Cette dernière forme est d'un intérêt particulier dans les études sur l'ADN ancien car elle provoque des déformations osseuses caractéristiques (Figure 13).



Figure 13 : maladie de Pott, touchant la colonne vertébrale. D'après Evinger *et al.* [81]

La première publication mettant en évidence l'ADN de *M. tuberculosis* remonte à 1993, quand Spigelman *et al.* détectent un fragment de l'IS6110 dans des os médiévaux [13]. La séquence d'insertion IS6110 est réputée spécifique de MTBC, tout en étant présente en de multiples copies. De nombreuses publications ont permis la détection et l'analyse de l'ADN de *M. tuberculosis* dans des ossements présentant des signes de la maladie de Pott. Néanmoins, des publications récentes ont mis en doute puis invalidé l'hypothèse de la spécificité des éléments d'insertion IS6110 et IS1081 pour MTBC [82-84]. Des mycobactéries du sol comportent dans leur génome des séquences comparables. Comme elles sont susceptibles d'avoir colonisé les os étudiés, la détection des séquences d'insertion n'est plus suffisante pour confirmer la présence de *M. tuberculosis*, ce qui met en doute de nombreuses études.

Si la présence d'ADN ne peut être confirmée par l'amplification des éléments d'insertion IS6110 et IS1081, d'autres marqueurs ont été mis en évidence. En 2012, Bouwman *et al.* ont réussi à amplifier, outre les éléments d'insertion, des éléments répétés (DR) spécifiques de *M. tuberculosis*, un fragment du gène *mpt40*, réputé lui-aussi spécifique du MTBC, mais surtout des fragments ciblant des SNPs, ce qui leur a permis de proposer un génotypage de la souche étudiée, datant du 19^{ème} siècle [85]. La souche identifiée est identique à des souches encore circulantes (souche H37Rv), bien qu'absente aujourd'hui de la zone géographique dont est issu l'échantillon. En 2013, Chan *et al.* publient le génome de deux souches de *M. tuberculosis* isolées à partir d'une unique momie datant du 19^{ème} siècle [86]. L'état exceptionnel de préservation leur permet d'obtenir 8% des séquences s'alignant sur le génome de *M. tuberculosis*. Enfin, en 2014, Bos *et al.* [40] ont publié les réels premiers génomes anciens de *M. tuberculosis*, séquencés à partir d'ossements précolombiens (11^{ème} - 13^{ème} siècle). La technique employée était comparable à celle mise en œuvre pour *Y. pestis*, avec sélection d'ossements présentant des lésions, préparation de banques, enrichissement par capture de gènes spécifiques (*rpoB*, *gyrA*, *gyrB*, *katG*, and *mpt40*) et séquençage sur plateforme Illumina. Sur 68 échantillons testés, trois ont présenté suffisamment de séquences pour être sélectionnés en vue d'un séquençage génomique après capture par hybridation sur l'ensemble du génome.

La souche séquencée s'est révélée être plus proche des souches de *M. tuberculosis* adaptées aux morses que de l'ensemble des souches modernes connues. Cela remet en cause l'idée que *M. tuberculosis* serait arrivé en Amérique par le détroit de Béring avec les premiers colons venus d'Asie. Dans le même article, les auteurs proposent une date

d'apparition pour *M. tuberculosis* il y a 6.000 ans, ce qui contredit la majorité des analyses basées sur la comparaison de séquences modernes, en faveur d'une émergence en Afrique, qui abrite encore aujourd'hui la plus grande diversité de souches. *M. tuberculosis* se serait répandu dans le globe à partir du Pléistocène [87].

Néanmoins, le débat reste ouvert, d'autres publications allant dans le sens de l'hypothèse de Bos *et al.* [88, 89]. L'absence de spécificité de l'ensemble des gènes réputés spécifiques de *M. tuberculosis* a récemment été démontrée par la même équipe qui a invalidé les séquences d'insertion. De même, ces auteurs évoquent le risque de créer lors de la reconstruction d'un génome de *M. tuberculosis* à partir d'échantillons anciens une chimère intégrant des éléments de *M. tuberculosis* et d'autres mycobactéries présentes dans l'environnement, ce qui peut induire des positionnements aberrants dans l'arbre phylogénétique de *M. tuberculosis*. Une telle aberration pourrait être responsable du placement de la souche séquencée par Bos *et al.* comme étant liée aux mammifères marins.

1.3.3. *Mycobacterium leprae*

Mycobacterium leprae est l'agent de la lèpre, ou maladie de Hansen. Cette bactérie est aérobique, non mobile et en forme de batonnets. Elle est acido-alcali résistante et pousse très lentement (13 jours par génération). Sa culture est fastidieuse, car il n'existe pas de milieu de culture. La bactérie est cultivée sur des souris ou des tatous.

La maladie se manifeste par des taches de dépigmentation, des lésions cutanées avec éventuellement une perte de sensibilité, la formation de nodules, de plaques, d'épaississement de la peau et une altération de la muqueuse nasale. Les nerfs sont également atteints, ce qui explique la perte de sensibilité. A long terme, les extrémités (doigts, orteils, nez) sont détruites par des phénomènes de défense immunitaire, ce qui entraîne des modifications osseuses visibles sur les squelettes. Environ 200.000 cas de lèpre sont déclarés à l'OMS chaque année, avec une tendance à la diminution [90]. L'objectif de l'OMS est l'éradication de la maladie.

Le mode de transmission naturel repose sur les gouttelettes expectorées lorsqu'une personne infectée tousse, se mouche ou parle.

Comme pour la tuberculose, les lésions osseuses facilitent la sélection des ossements à analyser (Figure 14).

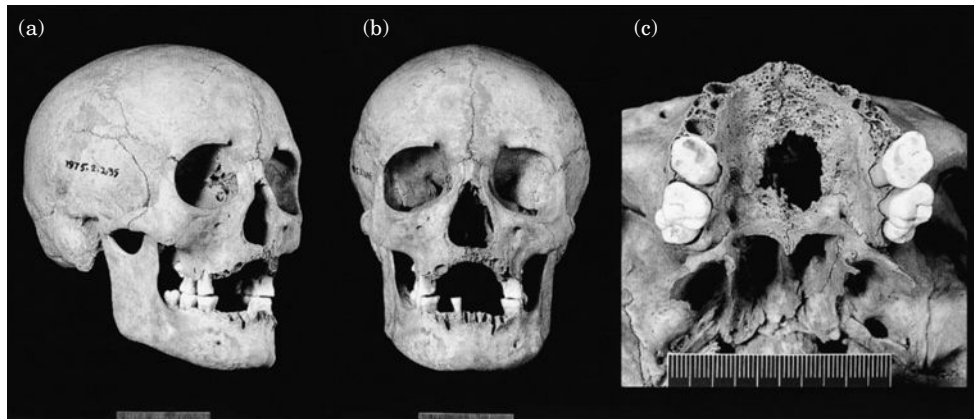


Figure 14 : lésions typiques de la lèpre au niveau de la tête : résorption des os du nez et de la partie antérieure de la mâchoire avec perte des dents antérieures. D'après Taylor *et al*, 2000 [Taylor].

Les premières détections d'ADN de *M. leprae* remontent en 1994 [91], sur des restes datant de l'an 600. Plusieurs publications ont amplifié différentes cibles chez *M. leprae* [92-95] mais la taille de certains produits de PCR (jusqu'à 500bp) semblait peu compatible avec une origine ancienne de l'ADN.

La nature des lésions osseuses laissait néanmoins peu de place au doute, et l'article de Schuenemann *et al.* en 2013 est venu apporter la confirmation qu'il était possible de trouver des fragments de grande taille pour *M. leprae* [22]. En effet, dans l'un des échantillons analysés, 40% des séquences obtenues après préparation des banques de séquençage et sans enrichissement s'alignaient avec le génome de *M. leprae*, alors que l'échantillon datait de près de 1000 ans. Les auteurs attribuent à la présence d'acides mycoliques l'exceptionnelle préservation de l'ADN. L'analyse comparative de 5 génomes anciens obtenus dans cette étude et de 11 génomes modernes a montré une très grande ressemblance, avec seulement 755 SNPs et 57 InDels entre les 16 génomes. La grande proximité entre les génomes anciens européens et les souches présentes en Amérique indiquent l'introduction par les Européens de la lèpre en Amérique, tandis que la similitude entre souches anciennes et modernes indiquent que la disparition de la lèpre en Europe à partir du 16^{ème} siècle doit plus être relié à des facteurs exogènes, tels que d'autres maladies infectieuses (*Y. pestis* et *M. tuberculosis* par exemple), l'amélioration de la santé ou une évolution de l'immunité, qu'à une perte de virulence des souches médiévales européennes.

1.3.4. *Treponema pallidum*

Treponema pallidum est un spirochète responsable de la syphilis. Cette maladie vénérienne est réputée avoir été ramenée du Nouveau Monde par les marins à partir de

la fin du XV^e siècle, et de s'être ensuite répandue en Europe, sous le nom de mal de Naples en France, et de mal français en Italie (Figure 15).



Figure 15 : à gauche, cette gravure sur bois (1496) est une des premières représentations des signes de la syphilis (Wellcome Library). A droite, cette peinture sur bois, représentant un syphilitique, est attribuée à Albrecht Dürer (1471-1528)

Certains chercheurs déclarent avoir mis en évidence des manifestations osseuses de la syphilis dans des squelettes européens précédant la découverte des Amériques [96] mais ces cas sont discutés et il n'a jamais pu être mis en évidence d'ADN de *Treponema pallidum* dans ces échantillons, pas plus que dans des échantillons plus modernes. Ce serait dû à la fragilité intrinsèque de ces bactéries et à la faible charge en bactéries lors des atteintes osseuses. Ainsi, plusieurs auteurs dont von Hunnius [97] sont pessimistes quant à la mise en évidence d'ADN tréponémal sur des échantillons anciens.

1.3.5. *Bartonella*

Les bactéries du genre *Bartonella* sont transmises par des ectoparasites (puces, tiques, poux, ...). *Bartonella quintana* est spécifique de l'espèce humaine et transmise par le pou du corps (*Pediculus humanus corporis*). Cette bactérie provoque une maladie

communément appelée fièvre des tranchées, car décrite initialement lors de la première guerre mondiale. La maladie est répandue chez les sans-abris, avec une incidence estimée à 14% [98]. Si la première description de la maladie remonte au début du 20^{ème} siècle, la bactérie infectait l'espèce humaine bien avant et l'équipe de D. Raoult a pu mettre en évidence l'ADN de cette bactérie dans des restes humains vieux de 4000 ans, à partir de pulpe dentaire [99].

1.3.6. *Brucella melitensis*

Brucella spp est un colibacille Gram négatif responsable de la brucellose (fièvre de Malte). Les principaux représentants sont *B. melitensis*, *ovis* et *abortus*, chaque espèce ayant une préférence pour une espèce mammifère. La brucellose se transmet par contact ou ingestion de produits contaminés, tissus, produits laitiers, environnement des animaux contaminés. Après une période d'incubation de 5 à 60 jours, l'infection se manifeste par un syndrome fébrile non spécifique et est susceptible d'entraîner des ostéomyélites[100], qui peuvent servir à identifier la pathologie sur des os (Figure 16).

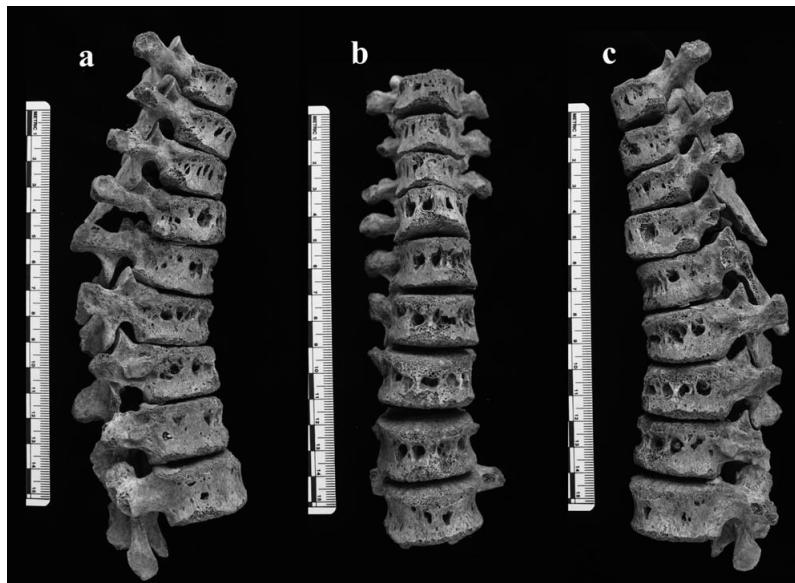


Figure 16 : vertèbres d'un squelette du 10^{ème} - 13^{ème} siècle, exhumé en Albanie (d'après Mutolo et al).

Dans la Figure 16, des lésions sur les vertèbres sont visibles, mais ne sont pas pathognomoniques de la brucellose. Dans l'article ayant étudié ces ossements, un élément d'insertion réputé spécifique du genre *Brucella*, l'IS6501, a pu être amplifié dans les extraits d'os, et la séquence s'est révélée être identique à des séquences décrites dans la littérature, confortant l'hypothèse d'une origine des lésions liées à une infection par *Brucella spp*.

Plus récemment, Kay *et al.* [101] ont été en mesure de publier le génome d'une souche de *Brucella melitensis* à partir d'un nodule calcifié humain datant du 14^{ème} siècle, isolé d'un squelette présentant des signes pathologiques, dont 32 nodules calcifiés dans la région pelvienne. Sans enrichissement ni capture, les auteurs ont pu produire des données de séquence permettant une couverture à un taux de 6,5 fois du génome de *B. melitensis*, avec 0,48 % des séquences amplifiées s'alignant sur le génome de référence *B. melitensis* 16M. De manière intéressante, par une approche phylogénétique utilisant les SNPs, les auteurs ont trouvé que la souche ancienne appartenait à un cluster encore présent, et dominant, aujourd'hui en Italie, montrant la continuité de la présence de ce cluster sur près de 700 ans.

De manière plus anecdotique, D'anastasio *et al.* décrivent des lésions pouvant évoquer une brucellose osseuse sur des vertèbres d'un *Australopithecus africanus* découvert à Sterkfontein [102]. Il semblerait néanmoins que les lésions décrites ne soient pas uniquement attribuables à une infection par *Brucella spp.*, et qu'en l'absence d'ADN, d'autres hypothèses sont à envisager.

1.3.7. Tannerella forsythia

Tannerella forsythia est l'une des bactéries fréquemment associées aux parodontites avec *Aggregatibacter actinomycetemcomitans*, *Porphyromonas gingivalis* et *Treponema denticola*. Cette pathologie consiste en une destruction de l'os de la mâchoire par des bactéries qui se multiplient dans des poches, appelées parodontales. Ces poches se forment lors de complications de gingivites, et finissent par entraîner un déchaussement des dents et une rétractation de la gencive. C'est une bactérie Gram négative, anaérobie. Il a été possible de séquencer son génome à partir de tartre datant de l'époque médiévale, en parallèle à la reconstruction du microbiome oral [103]. Les auteurs ont pu reconstruire le génome de ce pathogène avec une couverture moyenne de 5,7x, les principaux défauts de couverture étant associés à un transposon contenant des gènes putatifs de résistance à la tétracycline.

La bouche héberge un microbiome bien spécifique[104], qui peut être modifié en fonction de l'hygiène [105]. Il existe plusieurs niveaux d'atteinte dentaire, de la gingivite à la périodontite, plus ou moins grave. Dans un premier temps, un biofilm essentiellement composée initialement de bactéries Gram positif se dépose à la surface des dents, c'est la plaque dentaire. En cas d'absence de brossage, les bactéries Gram positif sont au fil du temps remplacées par des bactéries Gram négatif, puis ensuite par

des fusobactéries et des spirochètes. Le biofilm est responsable de l'irritation des gencives, la gingivite. Sans traitement, celle-ci peut évoluer en périodontite, une infection du pourtour de la dent liée au retrait de la gencive lors de la gingivite. Cette infection peut gagner l'os maxillaire et entraîner la perte de dents et le remodelage de l'os. La plaque dentaire, de manière concomitante, se minéralise et se transforme en tartre, une forme de calcaire. Trois bactéries majeures ont été identifiées comme prédominantes dans la périodontite, et regroupées sous le terme de « complexe rouge ». Ce sont *Porphyromonas gingivalis*, *Treponema denticola* et *Tannerella forsythia* (vue en 1.3.7). De nombreuses autres bactéries participent néanmoins au développement de l'infection, avec parfois des appellations de complexes orange, jaune, vert et pourpre [106]. Il a été montré que le tartre dentaire permettait une conservation sur plusieurs siècles des bactéries ayant participé à son installation [103]. Il est possible d'imaginer que d'autres bactéries, présentes dans la salive comme *Mycobacterium tuberculosis* [107] ou *Treponema pallidum* [108] puissent également être capturées dans la plaque dentaire et voir leur ADN conservé lors du processus de minéralisation.

1.3.8. Variola major

Le virus de la variole a pour hôte unique l'espèce humaine, chez qui il provoque un état fébrile, une éruption cutanée touchant tout le corps et, dans 30% des cas, la mort du malade, pour la forme majeure.

Ce virus appartient au genre orthopoxvirus. C'est un virus à ADN double-brin, encapsulé, d'assez grande taille et au génome de 200 kpb environ. Il a été éradiqué par une campagne mondiale dans les années 1980, et seuls deux laboratoires dans le monde sont encore autorisés à détenir des souches de variole.

Les stigmates de la variole ont été décrits sur plusieurs momies, en particulier Ramsès V, pharaon de la 20^{ème} dynastie, ayant régné de 1149 à 1145 avant JC. Des descriptions compatibles avec les signes cliniques de la variole peuvent être trouvées dans des écrits médicaux indiens (1500 av. JC) et chinois (1122 av. JC). Ainsi, il est estimé que la variole aurait divergé d'un clade ancestral, infectant des rongeurs en Afrique, il y a entre 16.000 et 68.000 ans. Ces données sont corroborées par l'analyse comparative de 47 séquences de ce virus conservées aux CDC d'Atlanta [109].

En 2012, Biagini *et al.* ont décrit la présence de fragments de gènes de la variole dans le poumon d'une momie sibérienne de la fin du 17^{ème} ou du début du 18^{ème} siècle [25]. Le séquençage de trois fragments de gènes a permis de confirmer l'infection par le virus de

la variole, infection confirmée selon les auteurs par le positionnement « ancien » de la souche dans l'arbre phylogénétique de la variole, construit uniquement à partir de séquences modernes (Figure 17).

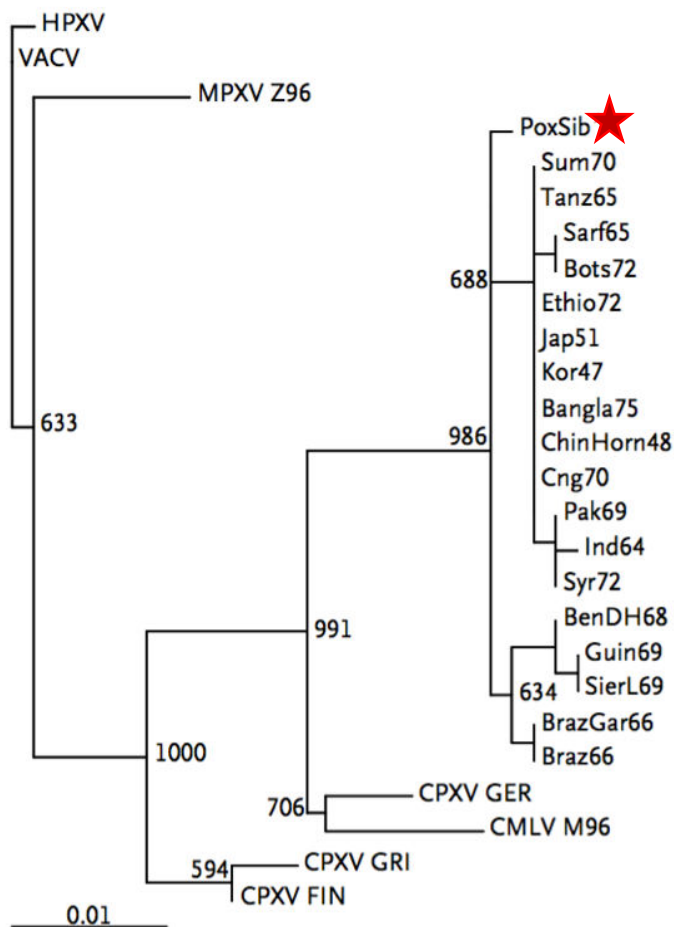


Figure 17: analyse phylogénétique du concaténat des trois gènes amplifiés dans la momie sibérienne (astérisque rouge). Ce concaténat apparaît à la racine de la diversité des souches modernes de variole. D'après [25].

Ce positionnement peut poser question, le corps ayant livré les séquences étant relativement récent comparé à l'ancienneté de l'infection de l'espèce humaine par la variole (attestée dès l'antiquité). Ce positionnement devra être affiné en augmentant le nombre de gènes séquencés.

1.3.9. Grippe espagnole

La grippe espagnole a frappé le monde entier (pandémie) de 1918 à 1919. Elle aurait fait entre 20 et 50 millions de morts, essentiellement de jeunes adultes. Originnaire de Chine, elle aurait gagné les Etats-Unis, où une mutation aurait drastiquement augmenté sa mortalité (3% au lieu de 0,1%). La pandémie gagne ensuite l'Europe et se répand dans le monde entier, infectant 1 milliard de personnes, soit la moitié de la population humaine.

Ainsi, ce virus associe une très forte contagiosité et une mortalité élevée, ce qui a longtemps intrigué les chercheurs qui ont voulu comprendre les facteurs génétiques responsables de la première grande pandémie virale de l'ère moderne.

En 2005, un article paru dans Science [110] décrit la reconstitution du génome du virus à partir d'échantillons biomédicaux conservés depuis l'épidémie et de tissus prélevés sur un cadavre enterré dans le permafrost . Ces travaux sur l'ARN ancien ont pu aboutir du fait de la bonne préservation de l'ARN dans le sol gelé. Il a pu être démontré que les exceptionnelles mortalité et contagiosité de la souche de 1918 étaient liées non à un gène particulier (neuraminidase ou hémagglutinine) mais à une combinaison de l'ensemble des 8 gènes portés par le virus [110].

1.3.10. Autres signes pathologiques

Le principal signe qui peut être également trouvé avec des restes osseux est le kyste calcifié, conséquence d'une calcification pathologique. Il s'agit de lésions liées à la précipitation de calcium en dehors des os. Le calcium forme des cristaux par association avec des ions carbonate ou phosphate. La formation de ces cristaux peut être favorisée par la présence de collagène, protéine présente en forte quantité lors de fibroses kystiques. Les kystes peuvent avoir divers aspects, selon l'abondance et la distribution des cristaux (os de seiche, coquille d'œuf, pierre, craie, ...). Le kyste peut être un mécanisme de défense de l'organisme, qui entoure un élément exogène d'une coque fibreuse. Avec le temps, cette coque fibreuse peut se calcifier, et donc être observée longtemps après la dégradation de l'ensemble des tissus mous. Un kyste peut avoir de nombreuses causes, et il est très difficile de déterminer son origine uniquement sur son aspect macroscopique. Les anatomo-pathologistes s'appuient sur l'histologie pour diagnostiquer l'origine étiologique d'un kyste. Dans le cadre de cette étude, cette approche est difficile, ne connaissant pas l'état de conservation éventuelle des tissus à l'origine de la calcification. On peut citer comme causes d'intérêt l'encapsulation de parasites (kyste hydatique) et la nécrose caséuse ou caseum (kyste lié à la tuberculose). Dans le premier cas et dans le cas humain, après éclosion dans l'intestin, les larves d'*Echinococcus* se fixent essentiellement dans un organe tel que le foie, les poumons ou les reins de l'hôte intermédiaire pour former un kyste dans lequel elles se multiplient. Comme les humains ne font pas partie du cycle de vie du parasite, il constitue une sorte d'impasse et les larves finissent par mourir. Le kyste peut alors se calcifier et il pourra être détecté par imagerie ou retrouvé dans le squelette.

La nécrose caséuse correspond à un foyer de destruction cellulaire liée à la multiplication dans les tissus, de bacilles tuberculeux. Le caseum se présente comme une substance éosinophile granuleuse ou homogène, de taille variable selon le volume de tissu atteint. Il s'agit d'une substance blanc-grisâtre, opaque, molle et onctueuse, comparable à du fromage blanc. Au fil du temps, le caseum peut se condenser, se dessécher, devenir pulvérulent, crayeux, se calcifier ou s'ossifier. Cette lésion est caractéristique de la tuberculose.

L'étude de l'ADN ancien a fait progresser nos connaissances sur le passé des espèces humaines, en montrant que plusieurs espèces ont coexisté et se sont mélangées, en laissant dans notre ADN des traces de ces croisements. Après le séquençage du génome d'un paléo-eskimo qui vivait il y a 4.000 ans en 2010 [35], le premier génome de Neandertal a été publié la même année [36], montrant la possibilité d'un flux de gènes entre Neandertal et humain moderne. L'ADN ancien a même permis de révéler l'existence d'une nouvelle espèce humaine, les Denisoviens, uniquement caractérisée par quelques dents et une phalange [37]. Sans cet outil, il n'aurait pas été possible de définir cette nouvelle espèce. Ici encore, il a été possible de mettre en évidence un flux de gènes entre Denisoviens et humain moderne, particulièrement visible dans les populations d'Asie du Sud-Est et d'Océanie.

Dans le champ de la microbiologie, l'avancée des connaissances, bien que moins visible, n'en a pas moins été majeure. Notre compréhension des origines des grandes épidémies a progressé, en premier lieu la peste. Trois pandémies de peste ont frappé l'humanité dans les temps historiques, les deux premières étant la peste justinienne et la peste noire. Dans les deux cas, la rapidité d'expansion et le taux de mortalité étaient très différents de la pandémie moderne, raison pour laquelle l'agent responsable des deux pandémies a longtemps été débattu. Il est démontré maintenant que la peste justinienne a été causée par une souche appartenant au groupe Antiqua, dont il n'est pas connu de descendant moderne. La peste noire est, elle, due à une souche du groupe orientalis, qui est encore aujourd'hui le groupe le plus répandu dans le monde, responsable donc de la seconde et de la troisième pandémie. Les analyses d'ADN ancien ont montré une bonne intégration des génomes anciens dans l'arbre phylogénétique de *Y. pestis* basé sur la diversité connue actuelle. Les études, par contre, n'ont pas pu identifier de marqueurs génétiques expliquant l'épidémiologie des deux premières pandémies par rapport à la troisième. D'autres éléments environnementaux, vectoriels ou épigénétiques sont maintenant explorés. Certains auteurs font remonter la circulation de souches de peste

en Europe à 5.000 ans, en présentant des souches dépourvues du gène *ymt*, porté par le plasmide pMt1 et impliqué dans la colonisation du tractus digestif de la puce, et donc facilitant la diffusion de la bactérie dans les populations murine et humaine [111]. Il est difficile dans ce cas d'affirmer la pathogénicité des souches identifiées, plusieurs souches modernes, placées « plus haut » dans l'arbre phylogénétique, n'étant pas pathogènes pour l'homme [112] mais seulement pour certains rongeurs. Il est donc abusif de parler de peste dans les cas décrits, aucun signe clinique ne pouvant en attester. Il est donc également possible que les souches mises en évidence dans l'article aient été apportées par des rongeurs qui auraient utilisé les corps enterrés comme source de nourriture ou les tombeaux comme terriers. L'ADN ancien a également permis d'affiner les calibrages d'horloge moléculaire pour déterminer l'ancêtre commun le plus récent (MRCA – most recent common ancestor). Le calcul de l'horloge moléculaire avec ces souches fait reculer l'âge du MRCA de 2.000 ans, pour proposer une origine il y a 5.783 ans et les souches séquencées se placent en position très basale dans l'arbre phylogénétique de *Y. pestis*. Il reste à déterminer si elles étaient réellement pathogènes pour l'homme, et capables de provoquer des épidémies, la transmission par la puce étant très déficiente.

Les mesures de MRCA effectuées ont confirmé pour *H. pylori* une très longue co-évolution avec l'espèce humaine [113]. *M. tuberculosis* était associé à l'espèce humaine pour certains auteurs depuis la sortie d'Afrique. La présence en Afrique de la grande majorité des lignages affectant l'espèce humaine, certains n'étant présents que sur ce continent, renforcent cette hypothèse, tout comme la présence dans la corne de l'Afrique d'une espèce proche, *M. canettii*, présente dans l'environnement et responsable dans de rares cas d'atteintes cliniques humaines [114]. Ainsi, l'hypothèse majeure était une origine africaine du complexe MTBC, et une expansion mondiale contemporaine des grandes migrations humaines il y a 60.000 ans [87]. Deux études ayant permis de reconstruire des génomes anciens de *M. tuberculosis* [40, 115] ont calculé un taux très proche d'horloge moléculaire, qui estime le MRCA du complexe à moins de 6.000 ans. Ces calculs supposent un taux constant de mutations pour toutes les souches du lignage, et différents taxons peuvent avoir des taux différents. 6.000 ans représentent donc l'âge minimal du complexe MTBC, mais il est fort possible que cet âge soit plus ancien.

La révolution de l'ADN ancien se poursuit, en poussant les limites des technologies et en améliorant les protocoles. En 2012, Allentoft proposait une estimation du profil de dégradation de l'ADN, avec un tableau de demi-vie de l'ADN [116]. On pouvait y lire qu'après 10.000 ans, la taille moyenne des fragments d'ADN conservés à 15°C serait de

13 bp, donc dépourvu d'information. Cette année, en 2016, Meyer *et al.* [60] ont publié le génome d'un hominidé archaïque de plus de 400.000 ans, repoussant les limites de l'ADN ancien, et identifiant des séquences nucléaires de néandertaliens archaïques du Pléistocène. Cette prouesse fait écho à celle de Orlando *et al.* [117] qui ont publié des séquences équines de 700.000 ans, provenant d'Alaska, au climat plus favorable à la préservation de l'ADN ancien. Il devient envisageable de détecter et de caractériser de l'ADN d'hominidés plus anciens, comme *Homo heidelbergensis* ou *Homo antecessor* et mieux comprendre les origines de l'espèce humaine.

Il conviendra d'être de plus en plus prudent et méticuleux, afin d'éviter de retomber dans les travers des débuts de la paléogénétique, qui ont inspiré Jurassic Park. Encore récemment, Goncharov *et al.* [118] ont annoncé avoir remis en culture et séquencé une bactérie isolée d'une carcasse de mammouth mort il y a 28.000 ans, alors qu'il semble qu'ils aient isolé une souche moderne contaminante [119]. La sensibilité extrême des techniques doit être contrebalancée par des précautions extrêmes et des contrôles stringents. De telles mesures sont nécessaires pour conserver la crédibilité des études passées et à venir.

2. ÉTUDE DE LA DIVERSITE BACTERIENNE

Nombre d'échantillons étudiés dans cette thèse sont d'origine environnementale, soit parce qu'ils ont été prélevés dans l'environnement à proximité des modèles d'expérimentation (sol au contact des souris enfouies, racines, ...) soit parce qu'ils ont séjourné très longtemps dans le sol ou les sédiments et en ont acquis certaines caractéristiques. Ainsi, lorsqu'on a cherché à caractériser la diversité bactérienne présente dans ces échantillons modernes ou anciens, les techniques d'identification par analyse de l'ADN extrait de l'environnement ont été identiques. Les protocoles d'extraction et de préparation ont été différents. La mise en culture bactérienne est d'intérêt pour les échantillons modernes, mais sauf exception, n'a pas été mise en œuvre pour des échantillons anciens.

2.1. Diversité bactérienne

Par leur biomasse élevée (4 à 6×10^{30} cellules microbiennes au niveau terrestre), leur diversité (nombre d'espèces bactériennes estimées par certains chercheurs à plus d'une dizaine de millions) [120, 121] et leurs fascinantes capacités métaboliques et d'adaptation, les microorganismes ont un rôle essentiel dans le fonctionnement et

l'évolution des cycles biogéochimiques et par conséquent celui des écosystèmes. Cependant, une minorité de microorganismes sont pathogènes et de ce fait causent des problèmes majeurs en santé végétale et animale, dont la santé humaine. L'exploration des environnements complexes demeure actuellement l'un des défis majeurs en écologie microbienne du fait de l'importante diversité des micro-organismes qu'ils hébergent. Pour une bonne caractérisation des pathogènes dans de tels environnements montrant une très forte diversité microbienne, il est nécessaire d'appliquer des méthodes de diagnostic très discriminantes. Un grand nombre de méthodes culturales, moléculaires et biochimiques ont été dès lors appliquées pour répondre à ces questions.

2.2. Classification et identification bactérienne

La taxonomie est l'ensemble des principes et théories permettant la classification des organismes. Depuis l'origine, les microbiologistes s'attachent à classer les bactéries qu'ils isolent. Avant de disposer des outils de biologie moléculaire, la classification reposait sur la coloration de Gram, la morphologie macroscopique (aspect des colonies sur milieux gélosés), la morphologie microscopique (coque, bacille, isolé, en chaînette, ...) la mobilité, la capacité à sporuler, les besoins ou contraintes de culture (aérobie, anaérobie stricte ou facultative), les besoins nutritionnels. Dans un second temps, ont été utilisées les caractéristiques biochimiques, antigéniques, pathogéniques, d'antibiorésistance et de sensibilité aux phages.

A partir de l'ensemble de ces caractéristiques, les microbiologistes identifiaient et classifiaient leurs isolats. Cependant, au sein d'une même espèce, on sait maintenant qu'il existe une variabilité dans l'expression des réponses à ces tests selon les souches. Les outils de biologie moléculaire ont alors permis de préciser la classification, de regrouper des souches au sein d'une même espèce (*Shigella spp* et *E. coli*), d'en séparer d'autres ou de réorganiser l'arbre phylogénétique procaryote. Aux études de similarité ADN-ADN a succédé la comparaison de séquences d'un gène orthologue présent dans les espèces à comparer. Le gène de l'ARN ribosomal 16S a été proposé en 1987 comme marqueur phylogénétique [122] et a été adopté par la communauté scientifique. En effet, ce gène de 1500 paires de bases environ répond à plusieurs critères essentiels pour la taxonomie [123], il est présent chez tous les procaryotes, où il remplit la même fonction. De par son rôle vital, il est soumis à une très forte pression de sélection et évolue lentement. Il peut donc être facilement séquencé car il possède des domaines hautement conservés avec peu de mutations fixées au cours de l'évolution, mais il présente

également des régions de séquences plus variables (HVR) au nombre de 9, qui sont d'autant plus différentes entre deux espèces que celles-ci sont phylogénétiquement éloignées. Ainsi, pour positionner dans un groupe phylogénétique une nouvelle espèce, on procède à l'amplification et au séquençage de ce gène puis à sa comparaison avec les séquences déjà connues disponibles dans les banques de données

2.3. La culture bactérienne

La mise en culture bactérienne par inoculation d'un échantillon issu de l'environnement dans un milieu nutritif, liquide ou solide a été utilisée très tôt après la découverte des bactéries pour permettre la croissance et la multiplication des bactéries présentes dans l'échantillon initial. Lorsqu'on utilise un milieu solide, il est possible de dénombrer le nombre de colonies ayant poussé et donc d'estimer le nombre initial de bactéries présentes dans l'échantillon. Les milieux de culture ont été formulés pour permettre la croissance de bactéries présentant un intérêt clinique, et il est connu depuis les expériences de Torsvik [124] et Whitman [125] qu'au mieux 1% de la communauté bactérienne tellurique est cultivable.

2.4. Identification bactérienne par analyse de l'ADN extrait de l'environnement

Il est estimé qu'environ un tiers de la biomasse de la planète est constitué par les microorganismes [125]. Les méthodes classiques de microbiologie basées sur la culture impliquent l'inoculation d'un échantillon environnemental sur des milieux de culture (solides ou liquides) dont la composition doit favoriser l'isolement des microorganismes d'intérêts [126]. Cependant, ces méthodes présentent certaines limites comme les difficultés à isoler certains micro-organismes et à les caractériser rapidement. En effet, on estime à moins de un pour cent la proportion d'organismes unicellulaires cultivables dans la plupart des communautés microbiennes [127]. Le sol est non seulement l'environnement avec la plus grande diversité de procaryotes, mais aussi celui avec la plus grande masse microbienne : un gramme de sol contient jusqu'à 10 milliards de micro-organismes et des milliers d'espèces différentes [128]. Aussi, le développement des techniques moléculaires permet aujourd'hui d'aborder les identifications plus simplement au travers de l'analyse des molécules d'acides

nucléiques (ADN et/ou ARN) en contournant les limites des méthodes culturales [127, 129].

Ces approches indépendantes de la culture reposent sur l'utilisation de génomes entiers ou alors de biomarqueurs capables de renseigner sur l'identité (biomarqueurs phylogénétiques) ou le rôle fonctionnel (biomarqueurs fonctionnels) d'un grand nombre de microorganismes. Les méthodes basées sur la PCR (approches amplicons également appelé barcoding) permettent de réduire la complexité des échantillons biologiques et sont largement utilisées pour la caractérisation des micro-organismes. Plus de 3,2 millions de séquences du gène de l'ADN ribosomal (ADNr) 16S (<http://rdp.cme.msu.edu/> Release 11 Update 4, consulté le 30/09/2015) sont disponibles dans les bases de données. Avec les possibilités offertes par le séquençage massif, l'ADNr 16S est (re)devenu un marqueur de choix pour effectuer la phylogénie des espèces bactériennes du fait de son ubiquité dans les génomes bactériens et du très grand nombre de séquences maintenant déterminées [130-132]; <http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>; <http://rdp.cme.msu.edu/>; <http://www.arb-silva.de/>). La diversité des écosystèmes bactériens a d'abord été caractérisée par analyse de l'ADNr 16S par des techniques d'analyses de produits de PCR par électrophorèse en gel dénaturant, ou en ciblant les polymorphismes de longueur des fragments de restriction ou par séquençage par Sanger de banques clonées [131]. L'évolution des techniques de génomique a ensuite permis d'utiliser des puces à ADN, dites PhyloChip [133], et maintenant le séquençage à haut débit qui offre lui une profondeur et une précision d'analyse inégalée par les autres techniques. Initialement, les études métagénomiques se sont appuyées sur l'utilisation et l'analyse des fragments obtenus à l'aide de la technique de pyroséquençage 454 [130] mais la technologie Illumina est devenue le standard [134-138]. Le séquençage métagénomique ciblant les régions d'ADNr 16S variables rend aujourd'hui possible la caractérisation rapide et en profondeur de la diversité du microbiome actuel ou fossile comme le montre la Figure 18.

Cependant, ces techniques basées sur la PCR restent encore incomplètes et sujettes à de nombreux biais. Outre l'extraction des acides nucléiques, la principale limite tient à l'efficacité d'amplification PCR non homogène [139-141]. Si des taxons sont suffisamment divergents au niveau de la séquence de leur gène 16S, les amorces s'hybrideront mal ou pas du tout, et la population bactérienne correspondante sera sous-évaluée voire ignorée. D'autre part, le gène ADNr 16S ne fournit pas une résolution

suffisante pour permettre dans tous les cas une discrimination au niveau de l'espèce et encore moins au niveau de la souche [142]. Les méthodes les plus ambitieuses tentent d'analyser l'intégralité de l'information génétique présente dans des échantillons métagénomiques, c'est-à-dire l'ensemble des potentialités génétiques contenu dans un échantillon environnemental [143]. Les avancées du séquençage de génome ont déjà abouti au séquençage de 4.296 génomes complets procaryotes disponibles publiquement (www.ncbi.nlm.nih.gov/genomes/lproks.cgi - données au 30/09/2015). Le développement d'outils moléculaires haut-débit ouvre ainsi de nouveaux horizons pour explorer l'extrême diversité des microorganismes et leurs étonnantes capacités métaboliques.

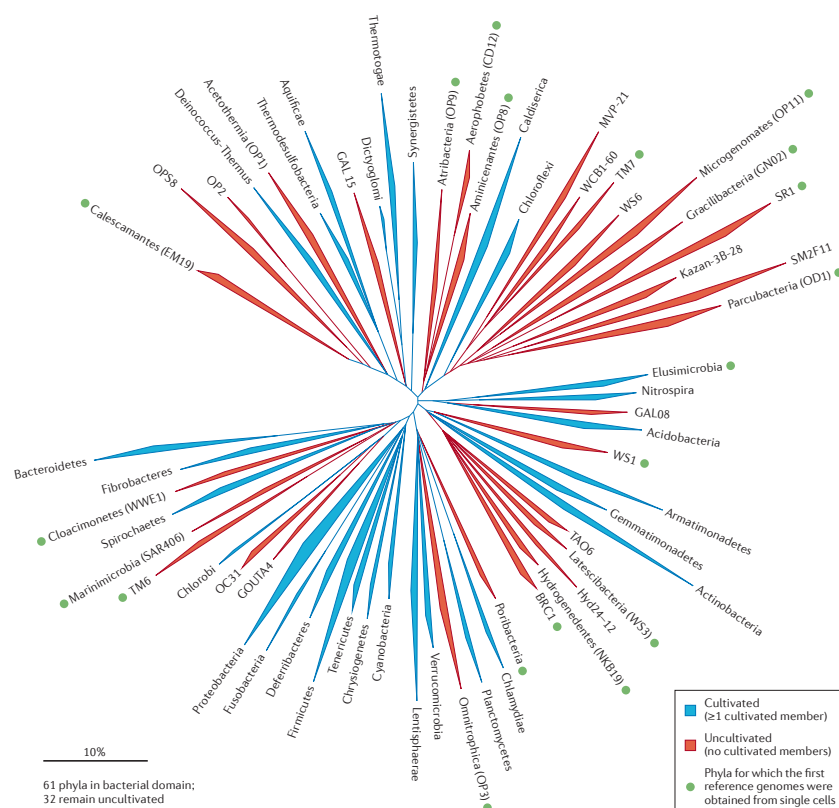


Figure 18: représentation sous forme d'arbre de la diversité microbienne connue à partir du séquençage en profondeur de l'ADNr 16S. Cette représentation illustre la grande diversité observée au sein des différents phyla bactériens et la mise en évidence de nouveaux phyla (SAR406 (Marine Group A), OP3, OP8, WS1, WS3, BRC1, CD12, EM19, EM3, NKB19 and Oct-Spa1-106) non encore isolés et cultivés. [144]

2.5. Limites de la métagénomique

La métagénomique permet l'étude globale de l'ensemble des génomes de communautés microbiennes environnementales sans nécessiter au préalable une connaissance ou une mise en culture [143, 145]. Des études métagénomiques ont été conduites au niveau de différents environnements tels que les sols, les lacs, les océans ou encore les drainages

miniers acides pour permettre d'avoir accès à la diversité phylogénétique et fonctionnelle d'organismes non cultivés [146-148]. L'exploration des environnements dans leur globalité nécessite en réalité un effort de séquençage très important, dépassant même les capacités actuelles des nouvelles approches de séquençage [149-151]. A l'heure actuelle, l'utilisation de ces nouvelles technologies reste limitée pour explorer finement les environnements complexes et coûteuse pour de nombreuses structures de recherche [152, 153]. De plus, elles ne permettent d'obtenir l'intégralité de l'information génétique que pour les micro-organismes dominants. Les pathogènes restent donc difficilement identifiables par une approche de séquençage métagénomique et cette situation devient d'autant plus critique lorsqu'il s'agit de traces d'ADN dégradé. De plus, les pathogènes anciens peuvent également être différents des pathogènes modernes. Enfin, une infime partie des échantillons présentent suffisamment d'indices pour soupçonner la présence de pathogènes, soit des lésions osseuses, soit une localisation particulière (par exemple le cimetière d'East Smithfield pour la peste [21]). Séquencer systématiquement tous les restes osseux est un non-sens, car de nombreux pathogènes ne provoquent pas de lésions osseuses, ou tuent leur hôte avant apparition de celles-ci.

3. TAPHONOMIE

La taphonomie, littéralement ce qui se passe dans la tombe (en grec *Taphos*) est la discipline qui étudie les phénomènes qui interviennent depuis la mort jusqu'à la fossilisation d'un organisme. Ce domaine de recherche a été initié par Ivan Efremov en 1940 comme étant "*the Study of the transition (in all its details) of animal remains from the biosphere into the lithosphere*".

La Figure 19 illustre le cycle taphonomique [154]. La taphonomie prend en compte les processus physiques, chimiques et biologiques, tout en intégrant les facteurs environnementaux tels que la chaleur, l'humidité ou l'enfouissement.

La mort signe le début de la dégradation de l'organisme, avec l'interruption des mécanismes d'homéostasie et de réparation des cellules, et en particulier de l'ADN endogène. Dans le même temps, le microbiote intestinal n'est plus contenu et les bactéries environnementales commencent à se répandre dans les différents compartiments de l'organisme [155].

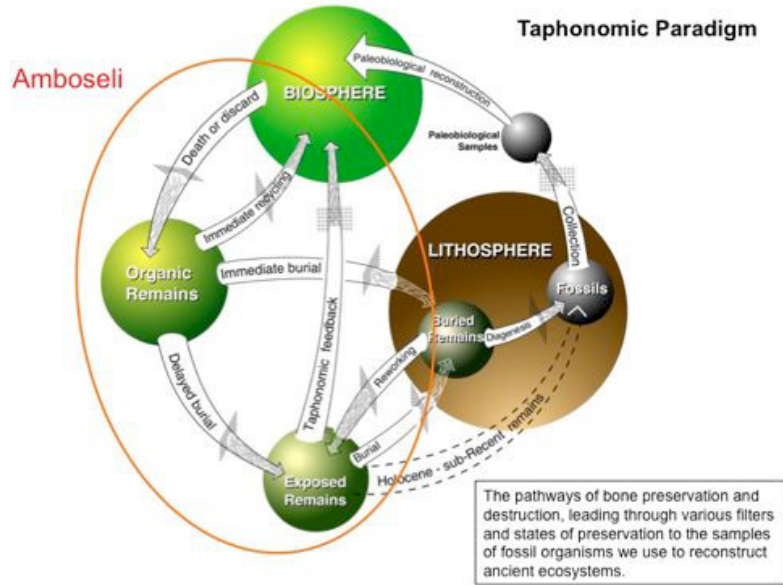


Figure 19 : représentation du cycle taphonomique, montrant les différents cheminements des organismes morts ou des rebuts organiques (fèces par exemple) selon qu'ils soient recyclés (et donc détruits) ou soient conservés et entrent dans les étapes ultérieures de la transition entre biosphère et lithosphère. Les filtres sur les flèches indiquent les processus qui modifient les informations biologiques et écologiques entre les différents stades (d'après [154]).

Cette colonisation participe à la dégradation de l'ADN endogène, les tissus mous de l'organisme servant de nutriments à ces bactéries. Par contre, il semble que l'ADN endogène présent dans les os et les dents soit relativement protégé. L'ADN d'agents pathogènes à l'origine du décès de l'organisme, provoquant une septicémie, se retrouve lui-aussi dans les os, et donc susceptible d'être également protégé.

Lors du décès d'un vertébré, des phénomènes de dégradation se succèdent et ont été classés en 5 grandes étapes [156], en l'absence d'enfouissement. Le principal facteur environnemental identifié influençant la durée de chaque étape est la température [157].

La première étape est l'autolyse [158]. C'est le résultat de l'arrêt de la circulation sanguine, qui entraîne la mort cellulaire. Cela entraîne la rupture des parois cellulaires et la libération d'enzymes dans l'espace intercellulaire. Les organes comme le pancréas libèrent de nombreuses enzymes digestives (amylases, lipases, trypsinogène), tandis que l'estomac libère de l'acide chlorhydrique. Ces molécules et enzymes débutent l'auto-digestion de ces organes et des tissus environnants. En parallèle, dès les premières minutes, des mouches nécrophiles viennent pondre leurs œufs sur le cadavre, alors que les processus d'autolyse ont à peine commencé [159].

La seconde étape, la putréfaction, est la destruction des tissus mous sous l'action de micro-organismes [160]. L'autolyse a créé un environnement interne essentiellement

anaérobie, favorable à la prolifération de bactéries issues du tractus gastro-intestinal. Ces bactéries, telles que *Bacteroidetes*, *Clostridium* ou *Streptococcus*, dégradent les protéines, glucides et lipides en acides et en gaz, responsables du gonflement progressif du corps. Ces bactéries incluent des bacilles anaérobies, des coliformes anaérobies facultatif, des microcoques et des *Proteus* [161]. Elles colonisent l'organisme par le système vasculaire. Le premier élément visuel indiquant la putréfaction est une modification de couleur de la peau, en général au niveau de la fosse iliaque droite, dans les 24 heures après le décès [162]. La peau prend une teinte verdâtre, du fait de la prolifération des bactéries présentes dans le caecum. La couleur verte est due à la dégradation de l'hémoglobine et à son interaction avec le sulfure d'hydrogène libéré par la fermentation bactérienne, produisant de la sulfhémoglobine. La peau prend ensuite un aspect marbré et présente de nombreuses vésicules voire des ampoules ou des cloques et a un toucher visqueux. L'épiderme peut alors facilement se désolidariser des couches plus profondes, tout comme les cheveux et les poils qui se détachent aisément. La couleur de la peau continue à évoluer, avec des délais selon les endroits, du verdâtre précédemment décrit à un gris-vert, puis marron-vert et s'assombrissant ensuite.

La pression engendrée par les gaz issus de l'autolyse et de la dégradation bactérienne (sulfure d'hydrogène, méthane, dioxyde de carbone, ammoniac et hydrogène) est dans un premier temps évacuée par les orifices naturels. La dégradation de l'enveloppe corporelle par le développement des larves d'insectes et la tension exercée par les gaz sous pression à l'intérieur du corps entraîne une rupture de l'abdomen avec perte des fluides de décomposition s'étant accumulés dans les cavités pleurales et abdominales [163]. Les phanères se détachent spontanément tandis que les tissus continuent à se dégrader voire se liquéfier.

La quatrième étape est la fermentation butyrique, qui commence aux environs du 20^{ème} jour [158]. Lors de la troisième étape, l'enveloppe corporelle s'est rompue, libérant les gaz et les fluides. Le corps s'aplatit, les tissus finissant d'être consommés par les bactéries et les insectes, l'eau s'évaporant également. On peut encore trouver les restes de tissus fibreux ou musculaux, comme le cœur. Au cours de cette étape est produit l'acide butyrique, qui donne au cadavre une odeur de fromage, et qui attire de nouveaux charognards. En effet, les larves et les insectes se nourrissant des tissus mous sont incapables de se nourrir des restes fibreux. Par contre, les coléoptères, ainsi que d'autres

insectes équipés de puissantes pièces buccales sont capables de découper et digérer ces tissus durs (Figure 20). Au début de cette phase, on trouve des larves de coléoptères, le plus souvent des familles Trogidae et Dermestidae [159]. Au stade adulte, ils sont capables de digérer les ligaments, de par leurs pièces buccales et leur équipement enzymatique, qui leur permet de digérer la kératine.



Trox scaber (famille des Trogidae)



Anthrenus verbasci (famille des Dermestidae)

Figure 20 : coléoptères fréquemment associés aux cadavres

La cinquième et dernière phase est appelée décomposition sèche ou squelettisation. Elle commence entre le 25^{ème} et le 50^{ème} jour *post-mortem*. Il ne reste du corps que les os, la peau et les phanères. Les os entrent dans un processus qui voit progressivement leurs composés organiques (collagène essentiellement) disparaître et laisser en place la matrice minérale (hydroxy-apatite). Les bactéries sont encore actives à cette phase, contribuant à la dégradation des macro-molécules biologiques présentes dans les os (protéines, acides nucléiques, ...). Des insectes fouisseurs, des racines de plantes peuvent également intervenir dans ce processus.

D'autres phénomènes, comme la momification naturelle ou l'adipocérification [164], peuvent remplacer les phénomènes de dégradation communs, mais ils sont plus rares et nécessitent des conditions environnementales particulières. Ces exceptions sont des extrêmes des effets de l'environnement sur la taphonomie. La température est un facteur majeur du processus mais de nombreux autres facteurs environnementaux peuvent influencer, en premier lieu l'humidité [165].

4. PRESENTATION DU PROJET

Dans le travail présenté ici, nous avons cherché à identifier des pathogènes anciens dans des échantillons humains et animaux, datant de 500 à 8000 ans. Ces échantillons ont été sélectionnés sur des caractéristiques morphologiques (déformations osseuses, kystes ...), taphonomiques (sépultures particulières, enfouissement en masse de

cadavres d'animaux) ou environnementales. Leurs ADNs ont été extraits et nous avons recherché l'ADN endogène par PCR quantitative en temps réel (qPCR) pour estimer l'état de conservation de l'ADN initial. Si une pathologie particulière était suspectée, nous avons cherché à mettre en évidence des marqueurs spécifiques de ce pathogène par qPCR. Dans le cas d'agents sporulants (*Bacillus anthracis*), une mise en culture des sédiments a été réalisée, dans des conditions de sécurité microbiologique adaptées (LSB3). En parallèle, l'ADN ribosomal 16S a été amplifié. Ce gène a la particularité d'être partagé et conservé par l'ensemble du règne bactérien, et des amorces dites universelles permettent d'amplifier des fragments polymorphes (HVR), autorisant une assignation taxonomique plus ou moins précise selon la taille du fragment amplifié. Dans le cadre de l'ADN ancien, nous nous sommes attachés à identifier un fragment hypervariable répondant à deux contraintes, une taille réduite (moins de 200 paires de base) et permettant une assignation taxonomique au niveau du genre bactérien. A l'issue de cette analyse, les échantillons les plus prometteurs ont été analysés par séquençage à haut débit (HTS) afin de rechercher la présence de fragments de génome du pathogène suspecté.

En parallèle de cette étude sur des échantillons anciens, nous avons étudié les phénomènes de dégradation de l'ADN dans l'environnement après la mort d'un organisme. Cette étude de taphonomie expérimentale a été menée sur des souris, enterrées ou laissées à la surface, et prélevées à intervalles réguliers pour analyser la vitesse de dégradation de l'ADN murin dans les ossements dans différentes conditions expérimentales, étudier les phénomènes de colonisation des tissus par les bactéries endogènes ou issues de l'environnement, et analyser la préservation d'un ADN bactérien exogène post mortem. Deux études de taphonomie expérimentale ont été menées, à court et à moyen terme. L'étude à moyen terme a eu pour objet, afin de caractériser les phénomènes naturels diagénétiques, de reproduire le vieillissement en nous appuyant sur un modèle murin infecté par deux simulants d'agents biologiques, *Bacillus atrophaeus* (Bg) et *Xanthomonas arboricola* (Xa). La première bactérie est utilisée par nombre d'équipes travaillant dans le domaine de la Biodéfense pour simuler *Bacillus anthracis*, agent responsable du Charbon, de par ses caractéristiques comparables. Néanmoins, la capacité à former des spores est susceptible de gêner notre étude en fournissant à l'ADN une forme naturelle de résistance et de protection dans le sol. *Xanthomonas arboricola*, non sporulant, a donc été sélectionné pour éviter la sporulation. Ces simulants, non pathogènes, ont été injectés dans des souris afin de

provoquer une septicémie. Les souris sacrifiées puis enterrées à quelques dizaines de centimètres de profondeur dans le sol ont été exhumées à intervalles de temps réguliers. La température du sol a été monitorée en continu.

Dans l'étude à court terme, les cadavres des souris ont été enterrés ou laissés à la surface sous protection dans un terrain d'essai. La quantité et la taille moyenne des fragments d'ADN endogène dans ces échantillons ont été analysées par qPCR, afin d'établir la cinétique de dégradation au cours du temps suivant un échantillonnage régulier et d'étudier la corrélation entre les caractéristiques de la sépulture et les caractéristiques biologiques des échantillons (surface vs enfoui). Les populations microbiennes dans les échantillons murins, ainsi que dans le contexte d'enfouissement, ont été analysées par séquençage massif en profondeur d'un fragment de l'ADNr 16S après amplification PCR. Tel que présenté plus haut, nous nous sommes concentrés sur la région V5 (28 pb), région hypervariable à même de fournir suffisamment d'informations pour classer les bactéries malgré sa taille restreinte [138]. Cibler des fragments est essentiel pour permettre l'amplification de l'ADN dégradé. Nous avons également étudié un amplicon englobant les fragments V3-V4-V5 à l'image du projet metaHit¹ (www.metahit.eu). Le couple d'amorces identifiée comme couvrant la plus grande diversité bactérienne a été employée [166]. Après PCR, les amplicons ont été préparés pour une analyse dans un séquenceur de seconde génération Illumina. Ce séquençage en profondeur a permis de révéler les populations bactériennes présentes dans l'échantillon et dans le sol, et en particulier d'analyser les transferts de populations entre ces deux compartiments. La profondeur de séquençage a permis la mise en évidence de fragments en très faible quantité, tandis que la faible taille des fragments amplifiés a été à même de compenser la dégradation de l'ADN. Nous avons analysé si les bactéries du sol envahissent l'os, si seulement un sous-ensemble de la population microbienne du sol était transféré et si la population endogène s'évadait des restes. En corrélant ces échanges avec les modifications structurales observées aux différents stades de la diagénèse osseuse, nous avons établi l'influence des modifications structurales sur les échanges bactériens et si le développement ou la mobilité de certaines populations bactériennes pouvaient être corrélées avec une modification physique et / ou avoir un impact sur la dégradation de l'ADN.

¹ Ce projet, financé par l'Union Européenne au travers du 7^{ème} Programme Cadre, avait pour objectif de d'établir des associations entre gènes du microbiote intestinal humain et la santé humaine.

1. MATÉRIELS

Comme indiqué en introduction, deux types d'échantillons ont été utilisés dans ce projet, des échantillons modernes (souris) et des échantillons anciens (tissus ou fragments minéralisés d'ancienneté variable).

Concernant les échantillons modernes, leur préparation est décrite dans le « Protocole n° 12130101 - Modèle d'infection avec *Bacillus atrophaeus* et *Xanthomonas arboricola* - Evaluation de la persistance d'ADN bactériens dans l'environnement après septicémie simulée » de DGA Maîtrise NRBC (Direction Générale de l'Armement, Maîtrise Nucléaire, Radiologique, Biologique et Chimique), présenté en annexe.

Les souris sont de type BALB/C. 68 souris ont été utilisées dans la première partie de l'étude de taphonomie expérimentale, et 20 souris dans la seconde partie, focalisée sur les phénomènes taphonomiques précoces.

Concernant les échantillons anciens, plusieurs sites ont donné des ossements ou restes potentiellement intéressants et sont décrits ici.

1.1. Echantillons animaux

1.1.1. Bure-Bourogne-Vellechevreux

Les échantillons de ces trois sites moyenâgeux (Figure 21) proviennent de bétail enterrés sans qu'il soit possible de détecter de traces de boucherie. Tout au plus les cornes et peaux de quatre individus ont été prélevées. Ainsi, les animaux n'auraient pas servi à l'alimentation humaine mais directement enterrés. C'est très rare pour la période considérée, d'autant que les famines étaient fréquentes et la viande rare. Selon l'archéozoologue, il fallait donc une motivation très importante pour ne pas consommer les bêtes, cette motivation étant pour lui liée à la maladie, en l'absence de contexte sacrificiel. A Bure, il a été trouvé 5 squelettes complets de bovidés, datés entre le 7^{ème} et le 11^{ème} siècle. Les seules traces anthropiques consistent en la mise en pièces de quatre des carcasses, sans doute pour faciliter leur manipulation. A Vellechevreux, quatre fosses, datées du 10^{ème} - 13^{ème} siècle, ont livré chacune un bovidé, à chaque fois relativement complet et montrant peu de traces d'équarrissage. Enfin, à Bourogne, deux

squelettes de bovidés datés du 9^{ème} – 13^{ème} siècle ont été retrouvés, sans traces d'équarrissage même si l'intégrité des ossements n'a pu être retrouvée. L'un d'eux présentait une pathologie osseuse.

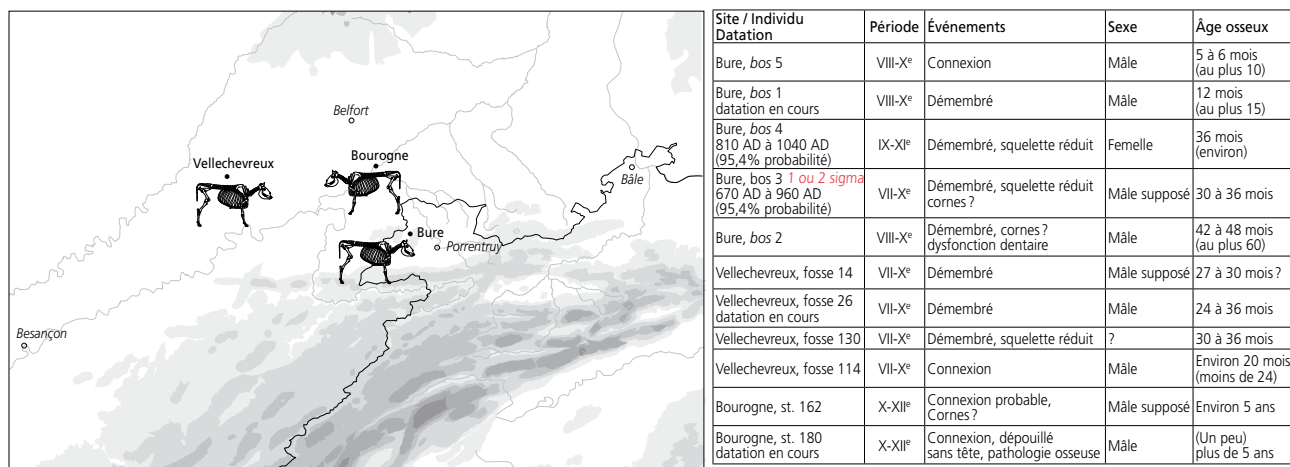


Figure 21 : localisation et composition des trois sites ayant livré des ossements bovins analysés dans ce projet. D'après Putelat [167]

1.1.2. Ostheim

Ostheim se trouve en Alsace, au Nord de Colmar. Le site altomédiéval d'Ostheim « Birgelsgaerten » (7^{ème} – 8^{ème} S. ap. J.-C.) a livré plus de 3000 restes animaux. L'étude encore en cours montre que l'abondance de gibier dans la cabane excavée 3219 et la fosse 3325, situées hors de l'enclos qui délimite l'établissement rural, tranche avec la rareté des restes de gibier constatée par ailleurs sur le site. On trouve dans ces deux structures, parmi un spectre de faune domestique des plus classiques pour l'Alsace mérovingienne, un taux élevé de faune sauvage (8% des 320 restes déterminés et 16% du poids des restes). La petite faune sauvage mise au jour (loutre, canard colvert, ichtyofaune) est nettement en relation avec l'environnement humide de « Birgelsgaerten ». On dénombre surtout trois cerfs différents, des ossements de sanglier, et, fait exceptionnel, la présence de l'aurochs (un calcanéus et probablement aussi un coxal), du bison d'Europe (un coxal, une phalange proximale) et de l'élan (diverses parties du squelette). La mise en évidence de ces trois taxons est toujours très rare en Gaule mérovingienne. Leur présence cumulée dans une même structure renforce le caractère atypique d'un corpus faunique qui caractérise le séjour d'un groupe humain aristocratique pratiquant la chasse sportive des gibiers nobles et potentiellement dangereux tels que élan, aurochs, sanglier.

1.1.3. Les Sablons

Le site « Les Sablons » en Charente est une nécropole mérovingienne avec un habitat associé. Parmi les ossements de bovidés exhumés, quatre ont été datés respectivement 600-660 ; 570-650 ; 650-720 et 570-650. Les dépôts semblent relativement contemporains. 5 des squelettes ont été en partie découpés : désarticulation des membres et de la tête et 4 d'entre eux portent des traces anthropiques : 2 des squelettes ont des traces de désarticulation sur la tête du fémur, sur l'humérus et une trace de décapitation entre l'atlas et l'axis. Les classes d'âge ne montrent pas de sélection, on a une courbe de type catastrophique.

1.1.4. Bernolsheim

Bernolsheim se trouve en Alsace, au Nord de Strasbourg. Sur ce site, le paléo-archéologue a mis au jour une fosse contenant des squelettes de chevaux, datant des 18^{ème} et 19^{ème} siècles, à l'exception d'un squelette (12080) datant du Moyen-Âge.

1.1.5. Horbourg-Wihr

Horbourg-Wihr se trouve en Alsace, immédiatement à l'Est de Colmar. Sur un site de fouille ont été retrouvés des ossements de chevaux, datant de l'Antiquité, dans une fosse, sans signe d'équarissage.

1.1.6. Muttersholtz

Muttersholtz se trouve en Alsace, à l'Est de Sélestat, entre Strasbourg et Colmar. Sur ce site, dans un ancien chenal ont été retrouvés des ossements de bovidés. De manière intéressante, les sédiments alentour ont été prélevés en même temps. Cependant, les os sont marrons, mous, et très humides, de par le prélèvement dans un chenal.

Le Tableau 1 liste l'ensemble des échantillons d'origine animale analysés.

Site	Identifiant	Os	Origine	signes de pathologie
Bourogne	0164	osselet	Bos	pas de traces d'équarrissage
	0168	osselet	Bos	
	0200	vertèbre	Bos	
Bure	0890-1	tête os long	Bos	
	0890-2	corps os long	Bos	
	0890-3	articulation	Bos	
	1272	mâchoire	Bos	
Vellechevieux	26-001	articulation	Bos	
	26-002	fragment articulation	Bos	
	26-003	osselet	Bos	
	130-010	mandibule	Bos	
	130-011	osselet	Bos	
	130-012	fracas osseux	Bos	
	114-007	mandibule	Bos	
	114-008	petits os plats feuilletés	Bos	
	114-009	fracas osseux	Bos	
Ostheim	3325-394B	os coxal	Bison	
	3325-1C	os coxal	Bison	
	3219-3	os coxal	Bison	
	3325-2D	os coxal	Bison	
Les Sablons	549-1	humérus	Bos	
	549-3	côte	Bos	
	650-2	humérus	Bos	
	709-1	fémur	Bos	
	888-2	côte G	Bos	
	888-3	fémur	Bos	
	914-1	humérus	Bos	
	917-2	côte	Bos	
	946-1	humérus	Bos	
	1555-2	ulna	Bos	
	1555-3	Côte D ?	Bos	
	1576-1	métacarpe	Bos	
	1576-3	tibia	Bos	
	1578-1	côte D	Bos	
	1578-2	côte G	Bos	
Bernolsheim	12080-1	os triangulaire	Equus	pas de traces d'équarrissage
	12080-2	dent	Equus	
	12040-1	os plat	Equus	
	12040-2	os long	Equus	
	12043.2-1	gros os	Equus	
	12043.2-2	côte	Equus	
	12043.2-3	dent	Equus	
	12041-1	os droit	Equus	
	12041-2	os courbe (prélèvement crête s)	Equus	
	12041-3	côte	Equus	
	12039.1-1	gros os	Equus	
	12039.1-2	os droit	Equus	
Horbouurg-Wihr	6013.05-01	vertèbre	Equus	
	6013.05-02	dent	Equus	

Tableau 1 : liste des échantillons analysés au cours de l'étude

1.1. Echantillons humains

1.1.1. Artzenheim

Artzenheim se trouve en Alsace, entre Colmar et le Rhin. Sur ce site ont été trouvées 46 sépultures à inhumation et 6 segments d'enclos funéraires [168]. Selon les analyses de mobilier funéraire, cet espace funéraire a été fondé à la fin du 6^{ème} Siècle et utilisé pendant toute la période mérovingienne (fin du 5^{ème} siècle, milieu du 8^{ème} siècle) voire pendant la période carolingienne (milieu du 8^{ème} siècle, 10^{ème} siècle). Parmi les sépultures, trois (tombes 81, 92 et 97) contenaient des individus porteurs de kystes (Figure 22).

ST-81 est un individu de sexe indéterminé, d'un âge inférieur à 30 ans. Le squelette est daté entre 650 et 771 après JC. La première vertèbre sacrée présente des lésions dégénératives, avec formation d'ostéophyte au niveau du plateau vertébral supérieur (signe d'arthrose vertébrale). Des caries et du tartre sont observés, et plusieurs dents étaient absentes.

ST-92 est un individu d'âge et de sexe indéterminé dont le squelette n'a pas été daté. Le squelette, hors kyste, ne présente pas de lésions ou variations anatomiques.

ST-97 est un individu de sexe indéterminé, d'un âge inférieur à 30 ans dont le squelette n'a pas été daté. Les vertèbres présentent des excroissances osseuses marginales sur le pourtour des corps vertébraux. Des microporosités sont observées sur les plateaux vertébraux, signe d'une pathologie dégénérative. Le kyste a été retrouvé dans le gril costal gauche.



ST-81

ST-92

ST-97

Figure 22 : photographies des kystes trouvés dans les tombes 81, 92 et 97 d'Artzenheim [168].

La morphologie des trois kystes est assez semblable, bien que la taille soit fort différente.

Dans un premier temps, des extraits ont été réalisés sur des os et des dents, afin d'explorer la conservation de l'ADN. Les kystes ont été analysés dans un second temps.

1.1.2. Kolbsheim

Kolbsheim se trouve à l'ouest de Strasbourg, en Alsace [169]. Sur ce site occupé depuis le néolithique ancien ont été retrouvées plusieurs tombes [170]. Dans l'une d'elles, il a été trouvé un squelette associé à quatre kystes (Figure 23). Le squelette correspond à une femme âgée entre 20 et 39 ans, ayant perdu des dents ante-mortem, daté entre 520 et 670 après JC. Il n'est pas observé d'autre lésion pathologique sur le squelette.



Figure 23 : photographies des quatre kystes retrouvés dans la tombe 514 de Kolbsheim [170]

Les quatre kystes peuvent être regroupés deux à deux, tant au niveau de la morphologie que de la taille. Leur aspect intrinsèque ne permet cependant pas de déterminer leur origine. Dans un premier temps, des extraits ont été réalisés sur des os et des dents, afin d'explorer la conservation de l'ADN. Les kystes ont été analysés dans un second temps.

1.1.3. Boulogne-sur-Mer

Les ossements ont été mis au jour lors de la fouille préventive d'une portion de terrain jouxtant la cathédrale de Boulogne-sur-Mer (Pas-de-Calais). Cette opération, menée par l'Inrap durant l'été 2012 (direction : O. Blamangin, anthropologie : S. Kacki), a conduit à la mise au jour d'une partie d'un cimetière paroissial médiéval et moderne. L'une des quelques 200 tombes fouillées a livré deux masses calcifiées, découvertes dans la région supérieure du thorax d'un sujet adulte masculin (Figure 24), par ailleurs affecté d'une maladie hyperostotique (ossification diffuse du rachis et calcification des tendons). Cet individu a été daté à 331±33BP, entre 1470 et 1650 n.e.

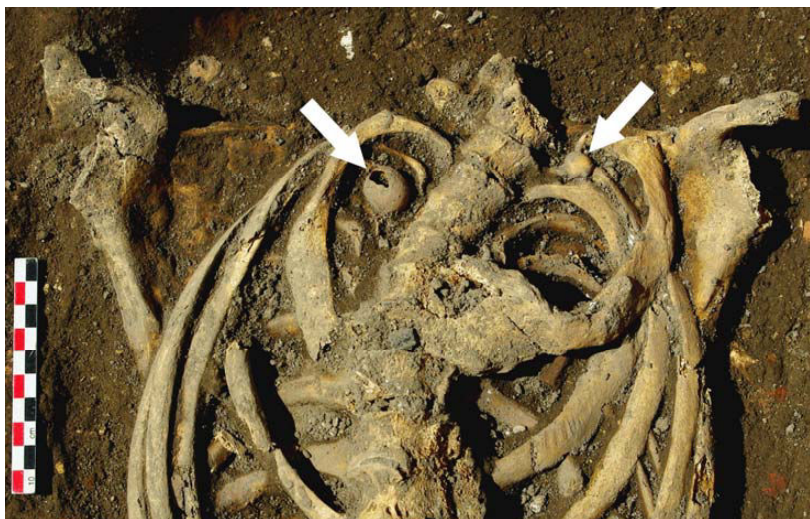


Figure 24 : calcification de Boulogne-sur-Mer *in situ* (O. Blamangin et S. Kacki).

Ces deux calcifications se présentent sous la forme de masses creuses de forme globalement sphérique, de diamètres respectifs d'environ 10 mm et 20 mm (Figure 25). Leur enveloppe, d'épaisseur millimétrique, est constituée d'un tissu calcifié de couleur brune/beige à la surface externe globalement lisse.

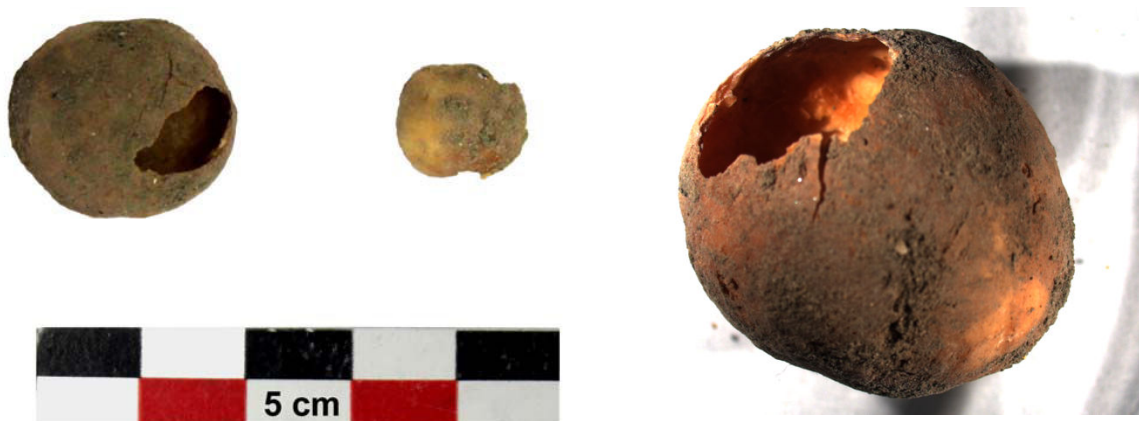


Figure 25 : Calcifications découvertes dans une tombe de Boulogne-sur-Mer (O. Blamangin et S. Kacki).

La morphologie, les dimensions et la localisation de ces éléments calcifiés constituent des éléments compatibles avec l'hypothèse de kystes hydatiques.

1.1.4. Saint-Laurent-de-la-Cabrerisse

Le site archéologique « Les Jardins de Saint- Benoît » est situé à Saint-Laurent-de-la-Cabrerisse (Aude). La fouille de ce gisement, effectuée dans un contexte d'archéologie préventive, a été réalisée aux mois d'août et septembre 2007 par la société ACTER archéologie (direction : A. Gaillard ; anthropologie : S. Kacki). L'intervention a mis au jour les vestiges d'une église de la fin du VIII^e siècle et de son cimetière attenant. La fouille partielle de ce cimetière (52 tombes fouillées sur les 150 repérées à l'issue du

décapage) a permis d'en préciser la chronologique (VIIIe-XIVe s.) et de réaliser un phasage de l'occupation funéraire, sur la base de la typologie des tombes et de plusieurs datations radiocarbone.

La calcification biologique a été découverte en association avec le squelette d'un sujet féminin adulte (S. 180), inhumé dans une structure funéraire maçonnée (réemploi d'une cuve agricole antique) attribuable à un horizon chronologique des VIIe-Xe Siècle (Figure 26). Le dégagement minutieux du squelette a permis d'identifier cette formation calcifiée dès la fouille et de préciser sa localisation dans la région thoracique, à proximité immédiate du rachis.



Figure 26 : Sépulture S. 180 du site de Saint-Laurent-de-la-Cabrèrissè (A. Gaillard et S. Kacki).

La calcification correspond à une masse creuse affectant une forme globalement ovale, surmontée d'un renflement hémisphérique. Ses dimensions sont approximativement de 27 x 20 x 17 mm. Sa partie externe est constituée d'une fine coque osseuse de couleur brune à la surface lisse, tandis que sa face interne se compose d'un tissu calcifié de couleur blanchâtre, présentant de multiples irrégularités de surface (Figure 27).

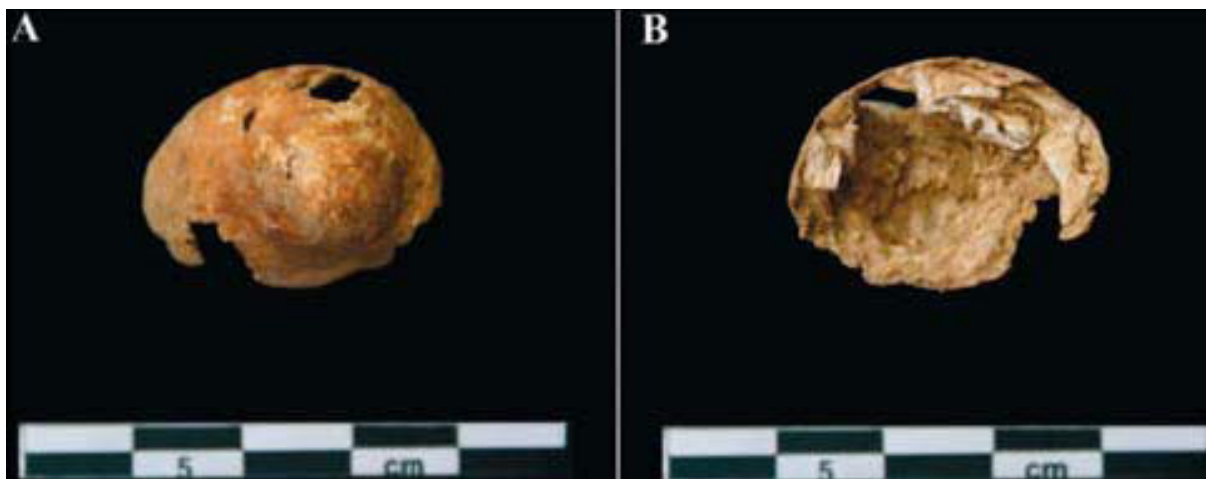


Figure 27 : Calcification biologique retrouvée au niveau du thorax de l'individu de la sépulture S. 180 (A : vue externe / B : vue interne) (A. Gaillard et S. Kacki).

La morphologie et la localisation supposée de cette calcification, ainsi que la présence de certaines lésions osseuses évocatrices d'infection (atteintes ostéolytique du plateau inférieur d'une vertèbre lombaire, réactions du périoste sur la face viscérale de plusieurs côtes), avait conduit à envisager que la calcification découverte pourrait correspondre à un ganglion tuberculeux.

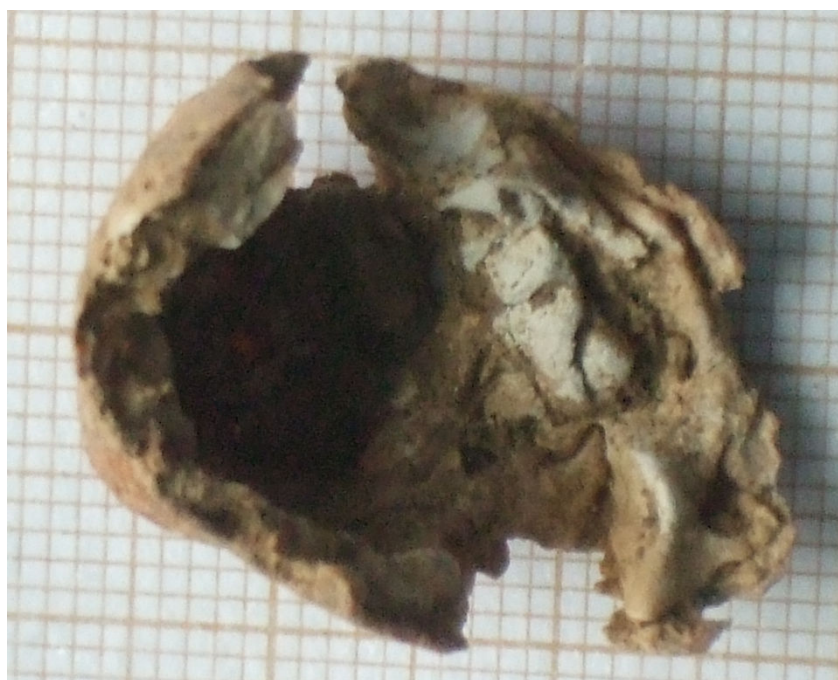


Figure 28 : photographie en gros plan de la zone calcifiée blanchâtre

La Figure 28 montre la zone blanchâtre attribuée à une calcification du caseum, au sein de la zone calcifiée.

1.1.5. Tartres des marins du Mary-Rose

Le HMS (Her/His Majesty Ship) Mary-Rose était un navire de guerre de type caraque, lancé en 1511. Elle faisait partie d'un vaste programme de développement de la marine de guerre britannique sous le règne d'Henry VIII, afin de doter le Royaume d'une flotte en rapport avec les ambitions du monarque. Une caraque est un navire aux châteaux avant et arrière surélevés par rapport au pont, afin de faciliter l'usage de canons de fort calibre. Son équipage était de 400 hommes environ lors de son naufrage. Plusieurs fois recarénée, sa dernière modification a été réalisée en 1545, année de sa disparition. Depuis 1542, la France de François 1^{er}, alliée à l'empire ottoman, combat le Saint-Empire germanique et l'Angleterre. En 1545, François 1^{er} rassemble en Normandie une flotte de plus de 400 vaisseaux afin de permettre le débarquement d'une armée de 30.000 hommes en Angleterre, et prendre à revers Henry VIII dont les armées sont stationnées à Calais et Boulogne. Un corps expéditionnaire est débarqué en Ecosse, et le 19 juillet, une bataille oppose les flottes française et anglaise à l'embouchure du Solent, bras de mer séparant l'Angleterre de l'île de Wight. Au cours de cette bataille, le Mary Rose sombra, les Français revendiquant de l'avoir coulé tandis que les Anglais affirmaient que le navire avait chaviré suite à une fausse manœuvre. Seuls 35 survivants purent être sauvés sur les 400 composant l'équipage. L'équipage était conséquent car le navire avait à son bord 185 soldats et 30 artilleurs, en plus des 200 marins. Plusieurs tentatives de renflouement furent menées dans les mois suivant le naufrage, sans succès, et l'épave sombra une seconde fois, dans l'oubli. Après quelques nouveaux essais de renflouement entre 1836 et 1840, un projet fut mis en place à partir de 1965 pour localiser et déterminer l'état de l'épave. Il fallut attendre 1971 pour retrouver précisément le Mary Rose et 1982 pour le renflouer. Depuis, l'épave est conservée dans un musée dédié à Portsmouth et de nombreuses études archéologiques ont été menées sur les outils, les armes et surtout les restes des membres d'équipage. Les restes de 179 individus ont pu être remontés à la surface, qui ont permis de reconstruire 92 squelettes presque complets. L'équipage ne comprenait que des hommes, âgés de 12 à 40 ans, ayant, selon les analyses isotopiques réalisées sur les dents, grandi dans les îles britanniques [171]. Notre laboratoire a pu collaborer avec Lynne BELL, qui a sous sa responsabilité une collection de dents des marins. Ainsi, nous avons reçu 28 échantillons de tartre, que nous avons dissous pour réaliser des recherches métagénomiques.

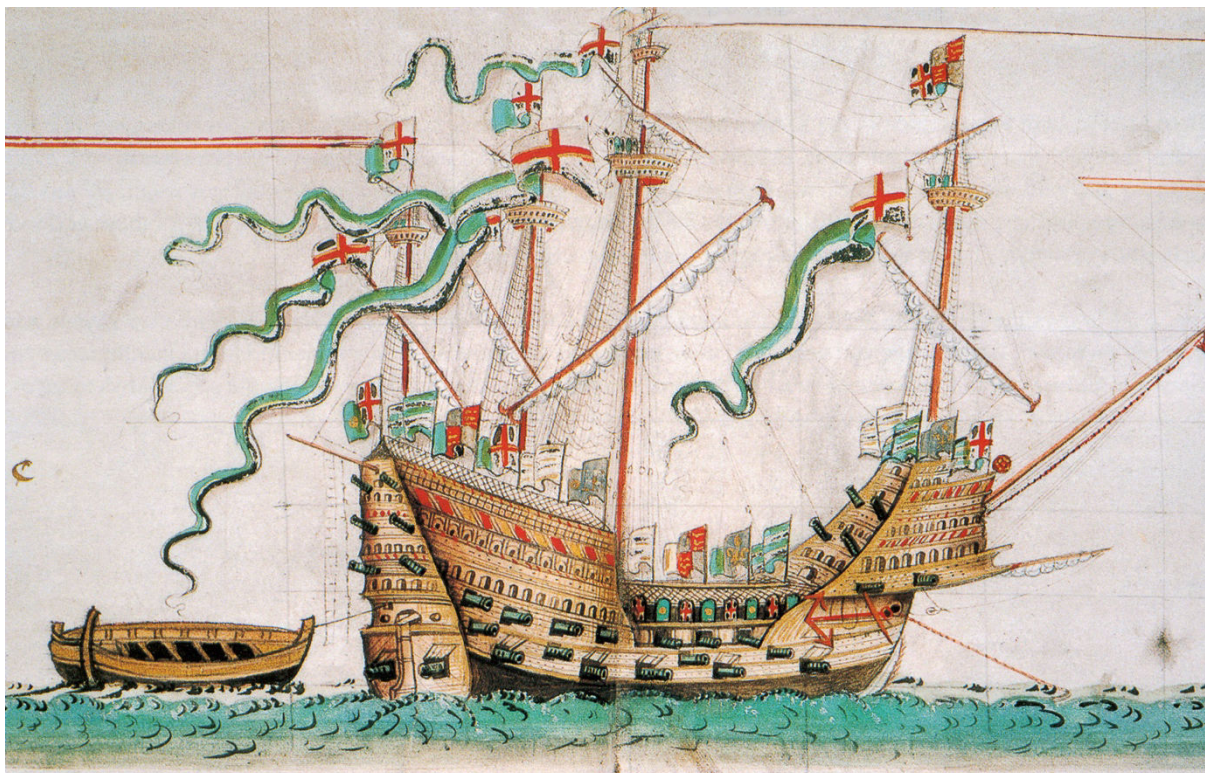


Figure 29 : représentation du Mary-Rose dans le rôle d'Anthony, registre répertoriant les navires de la marine d'Angleterre au 16^{ème} siècle. Cette image est la seule du Mary-Rose après son recarénage et montre les sabords ouverts proches de la ligne d'eau (adapté de [171]).

Ces analyses ont pour objet la recherche des pathogènes anciens et d'établir une évaluation de l'état sanitaire buccal d'un équipage de sa Royale Majesté au 16^{ème} siècle.

2. METHODES

2.1. Méthodes d'extraction, de purification et de réalisation de banques

2.1.1. Echantillons anciens

L'ensemble des méthodes mises en œuvre sur les échantillons anciens, de la préparation de l'échantillon à la réalisation de banques de séquençage, est décrit dans l'article « Analysis of Ancient DNA in Microbial Ecology » en encart [172]. Dans cet article, nous décrivons l'ensemble des techniques mises en œuvre dans notre équipe. La première étape consiste à extraire et purifier les acides nucléiques à partir d'échantillons anciens et de sédiments. Il est ensuite vérifié que les extraits réalisés n'inhibent pas la PCR, et éventuellement, il est recherché par qPCR l'ADN endogène. Les protocoles de préparation de banques de séquençage pour IonTorrent et Illumina sont décrits, avec

pour Illumina la préparation à partir d'ADN double brin et simple brin [48, 173]. Les procédures de décontamination des lieux de travail, des matériels et des réactifs sont également décrits, la contamination par de l'ADN moderne étant le phénomène le plus redouté lors d'études portant sur l'ADN ancien.

Article

Analysis of Ancient DNA in Microbial Ecology

Chapter 17

Analysis of Ancient DNA in Microbial Ecology

Olivier Gorgé, E. Andrew Bennett, Diyendo Massilani, Julien Daligault, Melanie Pruvost, Eva-Maria Geigl, and Thierry Grange

Abstract

The development of next-generation sequencing has led to a breakthrough in the analysis of ancient genomes, and the subsequent genomic analyses of the skeletal remains of ancient humans have revolutionized the knowledge of the evolution of our species, including the discovery of a new hominin, and demonstrated admixtures with more distantly related archaic populations such as Neandertals and Denisovans. Moreover, it has also yielded novel insights into the evolution of ancient pathogens. The analysis of ancient microbial genomes allows the study of their recent evolution, presently over the last several millennia. These spectacular results have been attained despite the degradation of DNA after the death of the host, which results in very short DNA molecules that become increasingly damaged, only low quantities of which remain. The low quantity of ancient DNA molecules renders their analysis difficult and prone to contamination with modern DNA molecules, in particular via contamination from the reagents used in DNA purification and downstream analysis steps. Finally, the rare ancient molecules are diluted in environmental DNA originating from the soil microorganisms that colonize bones and teeth. Thus, ancient skeletal remains can share DNA profiles with environmental samples and identifying ancient microbial genomes among the more recent, presently poorly characterized, environmental microbiome is particularly challenging. Here, we describe the methods developed and/or in use in our laboratory to produce reliable and reproducible paleogenomic results from ancient skeletal remains that can be used to identify the presence of ancient microbiota.

Key words Ancient DNA, NGS, Double-stranded library, Single-stranded library, IonTorrent, Illumina, Contamination

1 Introduction

Ancient DNA (aDNA) preserved in skeletal remains from past organisms can be a rich source of information on the evolution of species, of both the organism itself and its pathogens (for a typical ancient skeleton *see* Fig. 1). aDNA, however, is often highly degraded and the techniques for its analysis need to be optimized in order to ensure the production of authentic results. Indeed,

*Author contributed equally with all other contributors.



Fig. 1 Skeletal remains from a 7000-year-old Neolithic burial from Berry-au-Bac “le Vieux Tordoir” (Aisne, France). Excavation and photograph: CNRS, UMR 8215 “Trajectoires”



Fig. 2 Qiavac Manifold equipped with Qiaquick Spin columns and extenders to purify DNA from large-volume extracts

working with aDNA requires precautions that need to be applied even before samples enter the paleogenetic laboratory. Archaeologists should be taught the constraints of aDNA research so that they can adapt their working procedures to increase the likelihood of obtaining reliable aDNA results. For example, a common practice in archaeology is to wash bones after excavation. This washing leads to the dilution and degradation of the ancient DNA molecules as well as to contamination with environmental DNA [1, 2]. This is particularly problematic if only very low amounts of DNA are preserved in the skeletal remains. To increase the likelihood of DNA preservation in the sample, it is preferable to use freshly excavated remains for paleogenetic analyses [1]. This will be possible, however, only in a limited number of cases and does not concern previously excavated and curated remains. Moreover, DNA is heavily transformed after the death of an organism. Between several hours,

and possibly up to the first several years, after death, DNA is hydrolyzed enzymatically into small fragments leading to a median size of 50–70 bp (Guimaraes et al., unpublished) [3]. Over time, DNA bases become modified; in particular the cytosines become deaminated, which occurs preferentially close to the molecule ends [4]. The quantity of endogenous aDNA varies among samples, and even in different locations in the same skeletal remains. Although there may be other factors, temperature in particular has been characterized as playing a major role in aDNA preservation [5].

In order to study the genomes of pathogens that are associated with an animal (vertebrate) at the time of its death, one must consider all the events taking place during the diagenetic transformation of biomolecules, including DNA, following death. The body and its constituents will begin to decompose, mostly due to the action of microorganisms and insects. They not only metabolize biomolecules, but also deposit their own DNA, becoming the first contaminants of skeletal remains. Once the soft tissues and accessible organic parts of the bones have been consumed, the skeletal parts will enter a slow decay phase involving mostly chemical processes. When bones are buried, either intentionally at the time of death or simply due to natural burial of the skeleton over time, there will be a slow but regular exchange of biomolecules between bone and soil. Thus, at the time of excavation, the DNA that can be recovered may contain (1) DNA from the initial organism; (2) the DNA of the microbes, as well as parasites, that were associated with the organism during its lifetime (some of which may possibly have been the cause of its death); (3) the DNA of the organisms that have contributed to the decomposition of the body following its death; (4) and the DNA of the soil organisms that have penetrated into the bone. If at the time of excavation no special precautions are taken, and the bones are handled and washed as is routinely done, the bone can be further contaminated with fresh modern DNA, mostly of human and microbial origin, as well as from various other sources. The microbial composition of skeletal remains therefore reflects the microbial composition of the burial environment, showing that fossilizing skeletal remains resemble environmental samples. Indeed, when DNA retrieved from ancient bones is sequenced with a shotgun approach, typically only a few percent, or tenths of percent, of the sequenced DNA correspond to the initial organism, the rest being identified as “environmental DNA.” Most of this “environmental DNA” cannot be mapped to sequenced genomes, and remains as “unknown” [6]. Since older DNA is increasingly degraded until complete disappearance, one could expect that most of the environmental DNA recovered from the skeletal remains is of recent origin. The DNA decay rate depends on the environment; however, special “molecular niches” within the bone may offer more protected DNA-stabilizing microenvironments [7–9]. This can explain the exceptional preservation found in a limited number of

bones [10–13]. The more the microbiome is intimately associated with the bone and teeth matrix, the better the likelihood that it can also benefit from such “preserving molecular niches”. Thus, the DNA of pathogens that can be spread to bones and teeth with the bloodstream, like *Yersinia pestis*, *Mycobacterium leprae*, and *Mycobacterium tuberculosis*, can be retrieved from well-preserved ancient skeletal remains (e.g., [14–16]). Similarly, the DNA of ancient buccal microbiomes can be retrieved from dental plaque, which appears to offer a suitable mineralized environment for long-term DNA preservation [17]. It remains to be determined if other organisms of the microbiome can also deposit their DNA into favorable “preserving molecular niches” allowing long-term DNA preservation. Such microorganisms are the first players to colonize the body through the blood vessels and it is yet unknown whether most of the “environmental DNA” that can be retrieved from ancient bones is of recent or ancient origin. In fact, it is likely that the unique taphonomic history (the history of *postmortem* decay) of each bone, in both macro- and microniches, shows sufficient bone-to-bone diversity to allow very different outcomes in terms of DNA preservation, and of the age of the DNA that is recovered.

In order to optimize the recovery of DNA from the ancient microbiome, improved methods adapted to the preferential recovery of the most damaged molecules must be developed. Their use should prevent, as much as possible, the introduction and incorporation of modern DNA molecules into the sequencing libraries. After sequence production, bioinformatics methods adapted to analyze ancient molecules should be used. Since the analysis of ancient microbiomes is presently in its infancy, there are not yet reliable, established procedures available to ensure the recovery of authentic data. For the moment, one has to rely on the more established procedures developed to analyze ancient host DNA.

Here, we provide some guidelines designed for the analysis of ancient microbiomes. First, it is essential to use all possible means to minimize contamination with modern DNA because the minute quantities of DNA in the ancient bone and tooth extracts can readily be contaminated with traces of modern DNA from the same or other species. Indeed, the scarcer the endogenous DNA, the higher the ratio of contaminating DNA likely to be found in the extract. This requires a high-containment laboratory for extraction, purification, and library construction of aDNA. In addition, very strict protocols to avoid carryover contamination and to decontaminate reagents must be applied. Carryover contamination results from molecules produced during previous amplification or library construction steps being reintroduced into another sample. Another source of contamination is trace DNA molecules present in reagents such as DNA from domestic animals, the proteins of which are often used to stabilize enzymes,

human DNA from employees at biotech companies, or bacterial DNA from either the bacteria used for enzyme production or the bacteria introduced from the environment during the production process. In order to ensure the authenticity of the results, reagents must first be decontaminated to eliminate as much reagent-borne DNA as possible before use [18]. Here we describe the different protocols that we have developed to reduce the level of contaminating DNA found in reagents.

Second, because DNA is damaged and degraded, one must use library construction methods that allow the best possible recovery of the most damaged molecules, and, if possible, discriminate against the recovery of modern ones. We present herein experimental procedures allowing optimal recovery of short double-stranded and of highly damaged molecules, which are best recovered as single-stranded DNA [19]. In the case of amplicon sequencing (i.e., 16S rRNA gene), short regions must be targeted because of the reduced length of ancient DNA. When analyzing soils or sediments, the vast majority of DNA, and consequently 16S rRNA genes, is from modern organisms. As a consequence it is best to select for short DNA fragments, when possible, prior to amplifying targets.

Third, data produced must be analyzed using bioinformatic workflows designed to characterize ancient DNA and ancient microbiomes. We use *leeHom* [20], to quality trim and merge paired-end reads produced from short ancient DNA templates and *mapDamage 2.0* [21] to assess the authenticity of the mapped DNA. To analyze shotgun sequencing reads or 16S rRNA amplicons, we use both *MG-RAST* (presented in Chapter 4) and homemade dedicated pipelines with *leeHom* to pre-process reads, and *BWA* [22] to map them against an in-house reference sequence consisting of concatenated bacterial genomes, before taxonomic characterization.

2 Materials

Buffers are stored at room temperature while reaction mixes, primers, and most enzymes are stored at -20°C .

2.1 DNA Extraction Reagents

Prepare all solutions from autoclaved deionized water. We use household bleach (2.6 % sodium hypochlorite) and *RNase away* (Life Technologies, Carlsbad, CA, USA) as agents for decontamination and DNA removal.

1. Commercial soil extraction and purification kits are used, but the reagents are only opened in the high-containment laboratory, under controlled conditions to avoid contamination. We currently use *MoBio PowerMax Soil DNA Isolation kit* (MO

BIO, Carlsbad, CA, USA, ref. 12988), Qiagen Gel Extraction kit (Qiagen, Hilden, Germany, ref. 28704), and Qiagen PCR purification kit (ref. 28104).

2. Bone matrix disintegration and digestion buffer: 0.5 M EDTA pH 8, 0.25 M PO_4^{3-} , 0.14 M β -mercaptoethanol. 0.5 M EDTA pH 8 is prepared from EDTA powder and autoclaved water and pH is adjusted with NaOH pellets [2].
3. Buffer QG (Qiagen, ref. 19063), solubilization and binding buffer.
4. Buffer PE (Qiagen, ref. 19065), wash buffer.
5. Buffer EB (Qiagen, ref. 19086), elution buffer.

2.2 qPCR Reagents

1. MixG (homemade qPCR mix) [23]: To prepare 100 μL of 10 \times mixG, mix 19.5 μL γ -irradiated water (*see Note 1*), 6.25 μL 10 mg/mL bovine or horse serum albumin (BSA or HSA), 3 μL 10 % Lubrol-17A17 (SERVA Electrophoresis GmbH, Heidelberg, Germany), 50 μL 50 % glycerol, 1.25 μL 5 M KCl, and 20 μL 2.5 M AMPD (2-amino-2-methyl-1,3-propanediol) pH 8.3 (*see Note 2*). For volumes higher than 200 μL , aliquot 200 μL each in UV-transparent tubes (Qubit Assay tubes, Life Technologies, ref. Q32856) and treat with UV (*see Note 3*), dilute 10,000 \times SYBR-Green I (Life Technologies, ref. S-7585) 1/40 in DMSO and add 1 μL diluted SYBR-Green I per 100 μL mix. Freeze overnight at $-80\text{ }^\circ\text{C}$ (*see Note 4*).
2. BIOTEC buffer: To prepare 10 mL BIOTEC buffer, mix 200 μL 1 M Tris-HCl pH 7.5, 800 μL 25 mM MgCl_2 , 20 μL 5 M NaCl, 5 mL 50 % glycerol, 10 μL 10 % Triton x100, complete to 10 mL with γ -irradiated water. Aliquot 540 μL each in UV-transparent tubes, UV irradiate 300 s on each side on a UV cross-linker (*see Note 3*).
3. Thermolabile double-strand DNase (2 u/ μL hl-dsDNase) from ArcticZymes (Tromsø, Norway, ref. #70800). For a final activity of 0.02 u/ μL , add 1 μL 2 u/ μL of hl-dsDNase to 99 μL of BIOTEC buffer.
4. Decontaminated Taq DNA polymerase: 108 μL of 5 u/ μL Hot Start Taq polymerase in its storage buffer is supplemented with 6 μL premixed 200 mM MgCl_2 , 20 mM CaCl_2 , and 2.45 μL 50 mM DTT, then incubated with 6 μL 2 u/ μL hl-dsDNase for 30 min at $25\text{ }^\circ\text{C}$ followed by a 20-min inactivation step at $50\text{ }^\circ\text{C}$. Aliquot to desired volumes. Final activity of decontaminated Taq is 4.4 u/ μL .
5. Decontaminated dNTPs: 20 μL of dATP, dCTP, dGTP, and 40 μL of dUTP (100 mM stock solutions each) are mixed with 100 μL of γ -irradiated water. 40 μL of this dNTP mix is mixed with 35 μL γ -irradiated water, 4 μL 50 mM DTT, 20 μL 250

mM Tris pH 8, 80 μ L 50 mM MgCl₂, and 20 μ L 10 mM CaCl₂ and then incubated with 1 μ L 0.02 u/ μ L hl-dsDNase for 30 min at 25 °C followed by a 30-min inactivation at 55 °C. The final dNTP concentration is 2 mM (4 mM for dUTP).

6. 1u/ μ L codUNG (ArcticZymes, ref. #70500): codUNG is a uracil-DNA glycosylase from Atlantic cod that is completely and irreversibly inactivated by moderate heat treatment [18].

2.3 Library Preparation for PCR Products

1. Oligonucleotides
 - (a) Oligonucleotides are those proposed by Life Technologies for genomic DNA Fragment Library preparation¹ (*see Note 5*).
 - (b) Annealing buffer 10 \times
 - 25 μ L 5 M NaCl
 - 100 μ L 250 mM Tris
 - 50 μ L 250 mM MgCl₂
 - 75 μ L γ -irradiated water (*see Note 1*)
 - 40 μ M annealed adapters (*see Note 6*)
2. End-repair enzymes (3 u/ μ L T4 DNA polymerase and 10 u/ μ L T4 polynucleotide kinase), such as NEBNext end repair module (ref. E6050, New England Biolabs (NEB), Ipswich, MA, USA)
3. Commercial purification kit, based on silica columns (Qiagen or Macherey-Nagel, Düren, Germany) or SPRI magnetic beads (Ampure XP, Agencourt Technologies, Beverly, MA, USA, ref. 16388 or NucleoMag NGS clean-up and size selection, Macherey Nagel ref. 744970)
 - (a) DNA purification kit for 96 samples (96 silica column plate or SPRI magnetic beads)
 - (b) Individual DNA purification kit (silica column or magnetic beads)
4. Quick ligase, such as NEBNext Quick ligation module (NEB, ref. E6056, 2000 u/ μ L T4 DNA ligase)
5. Size selection reagents for E-gel or Caliper XT devices
6. OneTaq Hot Start 2 \times Master Mix with Standard Buffer (NEB, ref. M0484)

2.4 Double-Stranded DNA Library Preparation

Use γ -irradiated water for all solutions and buffer preparations as well as for any dilution or elution steps (unless otherwise indicated) (*see Note 1*).

¹ Appendix E, p. 56–57, Publication Part Number MAN0009847; Revision C.0 Date 29 April 2014.

1. Oligonucleotides (*see Note 7*)
 - (a) 10× Annealing buffer (*see Subheading 2.3, item 1*) UV-irradiated (*see Note 3*)
 - (b) Annealed adapters 40 μM (*see Note 6*)
2. Deaminated cytosine repair
1 u/μL USER enzyme (NEB, ref. M5505)
3. End repair
NEBNext End Repair Module (NEB, ref. E6050)
4. Purification of repaired ancient DNA extract
 - (a) MinElute Column (Qiagen ref. 28604) (*see Note 8*)
 - (b) Qiaquick Gel Extraction Kit (Qiagen ref. 28704)
 - (c) QG Buffer (Qiagen ref. 19063)
 - (d) Isopropanol
5. Blunt-end double-stranded adapter ligation
 - (a) 40 μM Double-stranded adapters
 - (b) NEBNext Quick Ligation Module (NEB ref. E6056)
6. Elongation and pre-amplification
 - (a) OneTaq Hot Start 2× Master Mix with Standard Buffer (NEB, ref. M0484)
 - (b) 10 μM Illumina amplification forward and reverse modified primers P5s/P7s (*see Note 9*)
7. Purification and size selection
 - (a) NucleoMag NGS clean-up and size selection kit (Macherey Nagel ref. 744970)
 - (b) Freshly prepared 80 % ethanol
 - (c) DNase/RNase-free water
8. qPCR determination of the optimal number of cycles for library amplification
FastStart DNA Master^{PLUS} SYBR Green I mix (Roche Applied Science, ref. 035158)
9. Final library amplification
 - (a) FastStart Taq DNA Polymerase, dNTPack (Roche Applied Science, ref. 04738)
 - (b) DNase/RNase-free water
10. Library characterization and purification
Agilent Bioanalyzer high sensitivity DNA kit (Agilent ref. 5067-4626)

2.5 Single-Stranded DNA Library Preparation

Use γ -irradiated water for all solutions and buffer preparations as well as for any dilution or elution steps (unless indicated otherwise) (*see Note 1*).

1. Oligos (*see Note 10*)
 - (a) 10 \times Annealing buffer UV-irradiated (*see Note 3*)
 - (b) 40 μ M Annealed CL53/CL73 adapters (*see Note 6*)
2. DNA preparation
 - (a) 100 u/ μ L Circligase II ssDNA ligase with 10 \times Circligase buffer and 50 mM MnCl₂ solution (Epicentre, Chicago, IL, USA, ref. CL902)
 - (b) 10 u/ μ L Endonuclease VIII (NEB, ref. M0299)
 - (c) 1 u/ μ L codUNG from ArticZymes (optional)
 - (d) 1 u/ μ L FastAP (Thermo Scientific, Waltham, MA, USA, ref. EF065), a thermosensitive alkaline phosphatase
3. First adapter ligation
 - (a) 50 % PEG 4000 (Sigma-Aldrich, St. Louis, MO, USA, ref. 95904)
 - (b) Dynabeads MyOne Streptavidin C1 (Life Technologies, ref. 6500)
 - (c) Bead binding buffer: 1 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 0.05 % Tween-20, 0.5 % SDS. Prepare buffer just before use and discard immediately. Buffer has no shelf life after adding SDS.
 - (d) Wash buffer A: 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 0.05 % Tween-20, 0.5 % SDS. Can be stored at room temperature for a month.
 - (e) Wash buffer B: 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 0.05 % Tween-20. Can be stored at room temperature for a year.
 - (f) Stringency wash buffer: 0.1 \times SSC, 0.1 % SDS. Can be stored at room temperature for a month.
 - (g) 10 \times ThermoPol Buffer (NEB, ref. B9004)
 - (h) Bst 2.0 DNA Polymerase (NEB ref. M0537)
 - (i) 10 \times Tango buffer (Thermo Scientific, Waltham, MA, USA, ref. BY5)
 - (j) 1 % Tween-20 (Sigma-Aldrich, St. Louis, MO, USA, ref. P2287)
 - (k) T4 DNA polymerase (Thermo Scientific, ref. EP006)
 - (l) Stop solution: 0.5 M EDTA pH 8.0, 2 % Tween-20

4. Second adapter ligation
 - (a) T4 DNA ligase (Thermo Scientific, ref. EL001)
 - (b) EBT: 10 mM Tris-HCl pH 8.0, 0.05 % Tween-20
5. Library amplification
 - (a) qPCR master mix such as LightCycler FastStart DNA Master SYBR Green I (Roche Applied Science)
 - (b) Either MinElute PCR purification kit (Qiagen), AMPure XP (Agencourt Technologies), or NucleoMag NGS Clean-up and Size Select kit (Macherey-Nagel)

2.6 Equipment (other than common devices for molecular biology laboratories)

1. qPCR-capable thermocycler such as LightCycler 2.0 (Roche Applied Sciences)
2. UV-crosslinker such as Spectrolinker XL 1500 UV-crosslinker (Spectronics Corp., Westbury, NY, USA)
3. Multi-purpose rotating tool such as Dremel 9100 Fortiflex Heavy Duty Flex Shaft Tool (Robert Bosch GmbH, Stuttgart, Germany) with diamond cutting wheel (ref. 545) and high-speed cutting or drilling bits (e.g., ref. 194)
4. Freezer-mill such as Spex Certiprep 6770 Freezer/Mill® (SPEX, Metuchen, NJ, USA)
5. Electrophoresis system for DNA sizing and purification, such as E-Gel SizeSelect (Life Technologies), Caliper Labchip XT (Perkin-Elmer, Waltham, MA, USA) or Pippin Prep (Sage Science, Beverly, MA, USA)
6. Lab-on-chip electrophoresis system, such as Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA)
7. Fluorimeter for DNA quantification, such as Qubit 2.0 (Life Technologies)
8. Optional: Robotic platform, such as a TECAN EVO 100 (Tecan, Maennedorf, Switzerland), to facilitate high-throughput sample treatment, but all steps can be performed manually
9. Heating/cooling block with mixing capability, such as an Eppendorf Thermomixer Comfort (Eppendorf, Hamburg, Germany) (*see Note 11*)
10. Ice-water bath
11. Magnetic rack for 1.5 mL tubes

3 Methods (See Note 12)

3.1 DNA Extraction and Purification

1. DNA extraction from soil
To extract and purify DNA from soil (*see Note 13*), we found PowerMax Soil DNA Isolation kit from MoBio useful for a wide range of samples (*see Note 14*).

Typically, 10 g of soil are processed following the guidelines recommended by the manufacturer.

After extraction and purification, DNA is dissolved in 5 mL resuspension buffer. It is subsequently concentrated and re-purified with a modified Qiagen Gel Extraction protocol as follows:

- (a) Set an incubator (e.g., heating block) or water bath to 50 °C.
- (b) Heat an aliquot of EB to 50 °C (150 µL per sample to be extracted).
- (c) Add 30 mL of QG and 20 mL of isopropanol to the DNA solution.
- (d) Pass through columns mounted on a manifold device (QIAvac 24 Plus, Qiagen, ref. 19413) equipped with extension tubes (Qiagen ref. 19587) (*see* Fig. 2).
- (e) Once all the samples have passed through the columns (*see* **Note 15**), remove extenders to wash them with autoclaved water and install them in place again.
- (f) Wash columns with 2 mL PE.
- (g) Break vacuum and stop vacuum pump.
- (h) Remove columns from the manifold and transfer them in clean 2 mL collection tubes.
- (i) Centrifuge at $9300 \times g$ for 2 min to dry the columns.
- (j) Transfer columns to new clean 1.5 mL microcentrifuge tubes.
- (k) Carefully dispense 100 µL of preheated (50 °C) elution buffer (EB or molecular grade water) and incubate for 1 min at room temperature.
- (l) Centrifuge at $9300 \times g$ for 1 min, discard columns, and store eluate (*see* **Note 16**).

2. DNA extraction from bone

Ancient bones and teeth can be considered as environmental samples (*see* **Note 17**), since they are a matrix in which bacteria, fungi, and other organisms reside. The dedicated procedures developed to access the endogenous DNA also extract the degraded bacterial (as well as fungal and eukaryotic) DNA present in the ancient bone (*see* **Note 18**).

The cleaning and powdering steps of the skeletal remains are performed in a high-containment laboratory [19] (*see* **Note 19**). The surface of the remains is removed in a UV-irradiated protective hood. Bone is then drilled or ground to fine powder in a freezer mill (Spex Certiprep 6750). Further processing of the bone powder is performed as described [19]. Blank extractions are carried out for each extraction series.

- (a) Decontaminate the hood and all tools with pure household bleach (use RNase away for decontaminating metal tools).
- (b) Cover the base of the working station with a sheet of aluminum foil.
- (c) Remove the outer layer of the sample with a scalpel. Use a new scalpel for each sample, and discard used scalpels.
- (d) Set up and turn on a vacuum cleaner during cutting to prevent the dispersal of bone powder in the working station.
- (e) Sample preparation in a freezer mill:
 - Use a Dremel multitool equipped with a diamond cutting wheel to cut pieces (up to 200 mg), adjusting the speed to the density/mineralization of the bone. Decontaminate the cutting wheel using the medium flame of a Bunsen burner at around 500 °C, but not above, to avoid damaging the diamond wheel.
 - Weigh the bone fragment(s) and pulverize it in a freezer mill in liquid nitrogen for 1 min (10 impacts per second).
 - After transferring the powder to a clean tube, clean freezer mill tubes. Metal parts: remove remaining bone powder in a bath of RNase away, rinse in a water bath, and dry under UV light. Plastic tubes: brush with water to remove remaining bone powder, rinse in a bleach bath, then with water and let them dry overnight on clean aluminum foil. Do not expose to UV light.
- (f) Sample preparation using a drill
 - Depending on the density of the bone, its shape, and the location of the sampling area, various drill bits can be chosen according to the user's needs. Assemble and decontaminate the drill bit using the medium flame of a Bunsen burner (ca. 500 °C). Drill at the lowest speed possible giving efficient bone powder production.
 - Transfer the bone powder in a pre-weighed tube and weigh it.
- (g) Cleaning
 - Between preparations of each sample, change the aluminum foil, clean the working station with bleach, and change gloves. Flame the drill bit or the cutting wheel using the medium flame of a Bunsen burner (ca. 500 °C).
 - After completion of the preparation series, decontaminate the Dremel tool with 70 % ethanol and RNase away (not bleach), and flame the drill bit or the cutting wheel using the medium flame of a Bunsen burner (ca. 500 °C). Clean and decontaminate with bleach the working station and the vacuum cleaner, place a UV

lightsource inside the working station so that it irradiates the surface of the bench at close proximity for at least 3 h. Clean the area around and wash floor with 10 % household bleach.

- (h) Once powdered, the bone is mixed with digestion buffer (1 mL per 100 mg of bone powder) and incubated at 37 °C in an orbital shaker (10 RPM) for 24/48 h or upon complete dissolution of the bone powder.
- (i) After disintegration of the bone matrix, DNA is purified on silica columns.
 - Centrifuge the suspension for 10 min at 13,000 × *g*.
 - Prepare manifold with QiaAmp spin columns and extenders as in Subheading 3.1, step 1.
 - Transfer supernatant to 15 mL Falcon tubes and store pellets (−20 °C) for possible future re-extraction.
 - Follow the modified Qiagen Gel Extraction protocol described in Subheading 3.1, step 1.

3.2 DNA Amplification

1. To test whether the DNA extracts from the ancient bone samples inhibit the PCR (*see Note 20*), serial dilutions of the extracts spiked with a known quantity of positive internal control are amplified [24]. Subsequently, the Ct (crossing point at threshold) is analyzed for each serial dilution of the same sample. We dilute the extract two- and fourfold at the highest, considering that further dilution of ancient DNA extracts potentially containing only few molecules could cause the loss of targets.
2. For DNA amplification (*see Note 21*), we systematically use a home-made decontaminated qPCR mix (mixG). Endogenous DNA detection relies on the amplification of DNA fragments that are informative enough to discriminate between animals or, for bacteria, between phyla, classes, orders, genera, or even species depending on the targeted genes. To assess bacterial diversity, the 16S rRNA gene is a powerful tool, thanks to its ubiquity and structure, which make it an optimal marker for the characterization of environmental samples (*see Note 22*).

Amplification is performed with a LightCycler 2.0 (Roche Applied Sciences). Prepare a PCR mix using 1 μL of DNA extract, 1 μL of mixG, 1 mM MgCl₂, 1 μM each of primer, 0.01 u of codUNG, 200 μM dNTPs (with dUTP in place of dTTP), 0.1 u of Taq, and complete to 10 μL with γ-irradiated water.

Hybridization times and temperatures are primer dependent and elongation times are defined according to user's needs.

3. With qPCR, amplification is reported by fluorescence emission of SYBR Green I intercalated in double-strand DNA. Non-specific products, which are sometimes synthesized during qPCR and are mainly primer-dimers, also lead to fluorescence

emission. A first control is the use of several non-template controls (NTCs) to monitor the cycle number required to amplify such dimers and to determine the T_m s of the various possible dimers. If similar C_t s and T_m s are observed for both samples and NTCs, the amplification of dimers can be suspected. Primer-dimers, however, may sometimes have a T_m close to that of the desired product. Gel visualization is then needed to discriminate between primer-dimers, other non-specific products, and the desired product. Direct sequencing of products is a mandatory step to authenticate ancient DNA PCR results. When amplifying a target gene present only in the species of interest, such direct sequencing can be done by Sanger sequencing (i.e., Eurofins Genomics, Ebersberg, Germany). When amplifying genes like the 16S rRNA, the PCR product is a mix of thousands of different sequences best analyzed through high-throughput sequencing (HTS). The IonTorrent platform is well suited for this.

3.3 Library Preparation for PCR Products

1. For each sample, mix 20 μ L of PCR product (diluted if necessary—*see Note 23*), 5 μ L of 10 \times enzyme buffer, 0.1 μ L of end repair enzyme, and 24.9 μ L of γ -irradiated water. Gently pipet the total volume up and down 1–2 times to mix and incubate for 30 min at 25 °C (*see Note 24*).
2. Purify end-repaired products using a 96-well plate system (for high-throughput) or in tubes by magnetic beads or silica columns according to the manufacturer's instructions (*see Note 25*). The final elution volume is usually 50 μ L. *Samples may be frozen at -20 °C at this point.*
3. In a 96-well plate, add 1 μ L of a different barcoded adapter mix (A+P1) combination to each well to be used. Distribute a premix composed of 6 μ L of 5 \times ligation buffer, 1 μ L of Quick ligase, and 2 μ L of γ -irradiated water to each well and add 20 μ L of each purified end-repaired sample (*see Note 26*). Gently pipet up and down 1–2 times to mix and incubate for 30 min at 16 °C.
4. Immediately after the ligation, add binding buffer to each well according to the manufacturer's recommendation (60 μ L of NT buffer, Macherey-Nagel), mix well, and pool all samples before loading on a silica column. The binding, washing, and drying steps are performed as recommended. Elute in a volume between 30 and 50 μ L. *Samples may be frozen at -20 °C at this point.*
5. Use an electrophoresis system for DNA sizing and purification to size-select your library and eliminate adapter dimers or multimers of amplicons. We use the E-Gel SizeSelect or the Caliper Labchip XT depending on the size range of amplicon size

- (*see Note 27*). To determine the size range to select, add the length of the two adapters (85–87 bp depending on the barcode length, *see Note 5*) to your minimal and maximal amplicon size.
6. Nick repair and amplification are made sequentially with the same reaction mix (*see Note 28*). Add 8 μL of size-selected sample, 1 μL of 10 μM Primer A, 1 μL of 10 μM Primer P1, and 10 μL of 2 \times OneTaq Hot Start Master Mix with buffer. Gently pipet up and down 1–2 times to mix, incubate for 20 min at 68 $^{\circ}\text{C}$ (nick repair step), and then amplify the library using the following program: initial denaturation at 94 $^{\circ}\text{C}$ for 5 min (94 $^{\circ}\text{C}$ for 15 s, 60 $^{\circ}\text{C}$ for 15 s, 68 $^{\circ}\text{C}$ for 40 s) for six cycles, final elongation at 68 $^{\circ}\text{C}$ for 5 min (*see Note 29*).
 7. Purify the amplified libraries with a silica column or SPRI magnetic beads. Final elution volume is usually 30 μL . *Samples may be frozen at -20°C at this point.*
 8. Qualitative analysis on a Bioanalyzer is recommended to check the final library product (*see Note 30*). Products saved at intermediate steps (sizing and amplification) can also be run on the same chip and compared. Libraries are quantified using a Qubit 2.0 to determine the concentration and adapt it to the emulsion PCR for IonTorrent PGM sequencing. A qPCR is also recommended to compare the new library to a known reference to ensure that the Qubit measurement corresponds to samples ligated with the two adapters.
 9. Follow the manufacturer's recommendations to prepare the chip for sequencing. Depending on the heterogeneity of the PCR product(s) and the number of samples analyzed, either 314, 316, or 318 chips (V2) can be used.

3.4 Double-Stranded DNA Library Preparation

1. Mix in a 1.5 mL tube 1–500 ng of DNA extract, 3 μL of 10 \times NEBNext end repair buffer, 1.5 μL of 1 u/ μL USER enzyme, and complete to 28.5 μL with γ -irradiated water. Mix by pipetting and incubate for 1 h at 37 $^{\circ}\text{C}$ in a heating block.
2. After the cytosine deamination repair step, add 1.5 μL of NEBNext End Repair Enzyme Mix directly to the tube. Mix well by pipetting gently times and incubate for 30 min at 20 $^{\circ}\text{C}$.
3. The extract is purified with Qiagen MinElute kit, following the manufacturer's protocol, except that elution is done twice with 17 μL of γ -irradiated water (preheated at 50 $^{\circ}\text{C}$).
4. Add 10 μL of 5 \times Quick ligation reaction buffer, 1 μL of the 40 μM annealed P50X adapter, 1 μL of the 40 μM annealed P7XX adapter (*see Note 31*), 2 μL of Quick DNA ligase, and 6 μL of γ -irradiated water to the sample. Gently pipet 1–2 times to mix and incubate for 30 min at 20 $^{\circ}\text{C}$.

5. Add to the sample 3 μL of each of the 10 μM Illumina amplification forward and reverse primers P5s and P7s and 50 μL of OneTaq 2x Master Mix. Transfer the 100 μL reaction solution into a 0.2 mL PCR tube, and run in a thermocycler with the following program: 15 min OneTaq elongation step of the ligated products at 68 $^{\circ}\text{C}$, followed by six cycles involving denaturation for 20 s at 95 $^{\circ}\text{C}$, annealing for 35 s at 60 $^{\circ}\text{C}$, and primer extension for 70 s at 72 $^{\circ}\text{C}$.
6. The purification of the six-cycle-amplified library and the removal of potential artifacts (primer-dimers) are done using the NucleoMag NGS clean-up and Size Select kit:
 - (a) Add 130 μL (1.3 \times) Macherey-Nagel (MN) beads to the 100 μL of amplified library.
 - (b) Vortex and let sit for 5 min at RT.
 - (c) Quick spin and place on a magnetic rack; let sit for 2 min or until supernatant is clear.
 - (d) Discard liquid.
 - (e) Add 500 μL 80 % ethanol (freshly made).
 - (f) Twist tubes two or three times, until beads no longer stick to the tube's surface, and let sit for 2 min or until supernatant is clear.
 - (g) Discard ethanol with a pipet, being careful to remove as much as possible.
 - (h) Let dry for 2 min at RT on the magnetic rack.
 - (i) Remove from the magnetic rack, add 52 μL γ -irradiated water, and pipet ten times to mix well.
 - (j) Let stand for 2 min at RT away from the magnetic rack.
 - (k) Quick spin and place on the magnetic rack; let sit for 2 min or until supernatant is clear.
 - (l) Remove 50 μL to new tubes.
 - (m) Add 65 μL (1.3 \times) MN beads.
 - (n) Repeat **steps b–h**.
 - (o) Remove from the magnetic rack, add 25 μL γ -irradiated water or EBT, and pipet ten times to mix well.
 - (p) Let stand for 2 min at RT away from the magnetic rack.
 - (q) Quick spin and place on the magnetic rack; let sit for 2 min.
 - (r) Remove 22 μL of purified sample and place in a new tube.
7. To obtain a sufficient quantity of DNA while avoiding over-amplification (*see Note 30*), qPCR quantification of the pre-amplified library is performed using three serial dilutions (1:10, 1:100, 1:1000) and the modified Illumina P5 and P7 primers. For each sample, the amplification curves are analyzed and the

cycle number at the point between exponential phase and saturation is determined. This value is used to calculate how many cycles are needed to amplify the library. For a typical reaction, the correct amplification of the library will be 7–8 cycles less than the value determined with the 1/100 diluted sample (*see Note 32*).

8. To further amplify the library, mix 20 μL of the library, 61 μL of γ -irradiated water, 2 μL of 5 mM dNTPs (A,T,C,G), 3 μL of each 10 μM Illumina amplification primers P5s and P7s, 10 μL of the 10 \times PCR reaction buffer containing 20 mM MgCl_2 , and 1 μL of 5 u/ μL FastStart Taq DNA Polymerase. Gently pipet 1–2 times to mix and amplify the library using the following program: (95 $^\circ\text{C}$ for 20 s, 60 $^\circ\text{C}$ for 35 s, 72 $^\circ\text{C}$ for 70 s) for the appropriate number of cycles.
9. Purify the PCR using the Qiagen PCR purification kit.
10. Quantify the amplified libraries using a fluorescence-based quantification method; observe the size distribution of the amplified library and possible presence of artifacts on an Agilent Bioanalyzer 2100 (*see Note 33*).

3.5 Single-Stranded DNA Library Preparation

Prepare all enzymatic mixes prior to each step to avoid letting the beads dry between steps.

1. Dilute the purified DNA extract (between 1 fmol and 1 pmol of DNA) with γ -irradiated water to a final volume of 29 μL (*see Note 34*).
2. To optimize damaged DNA recovery prior to library preparation, add 29 μL of diluted DNA extract, 8 μL of 10 \times Circligase buffer, 4 μL of 50 mM MnCl_2 , and 0.5 μL 10 u/ μL Endonuclease VIII to a 1.5 mL tube and incubate for 1 h at 37 $^\circ\text{C}$ (*see Note 35*).
3. Add 1 u of FastAP to the above reaction and mix. Spin briefly and incubate for 10 min at 37 $^\circ\text{C}$. Incubate the reaction for 2 min at 95 $^\circ\text{C}$ to heat denature the DNA, then transfer tube to an ice-water bath, and leave for 1 min. Spin briefly.
4. To ligate the first adapter
 - (a) Add 32 μL 50 % PEG-4000, 1 μL 10 μM adapter oligo CL78, and 1 μL 100 u/ μL Circligase II. Incubate tube for 1.5–3 h at 60 $^\circ\text{C}$. *Samples may be frozen safely at -20°C at this point.*
 - (b) For each sample, transfer 20 μL of MyOne C1 dynabeads into a 1.5 mL tube. Use additional tubes for more than five samples. Wash beads by placing tube on a magnetic rack for 2 min. Discard supernatant and wash beads twice with 500 μL bead binding buffer (*see Note 36*).

- (c) Resuspend beads in 250 μL bead binding buffer per sample (*see Note 37*), vortex, and transfer 250 μL of beads to each ligated sample.
 - (d) Rotate tubes slowly on a rotating wheel, making sure that the beads stay in suspension, for 20 min at room temperature.
 - (e) Spin tubes briefly, place on a magnetic rack for 2 min, and discard supernatant.
 - (f) Add 200 μL wash buffer A.
 - (g) Place tubes on the magnetic rack, wash by twisting tubes (*see Note 36*), and then leave tubes on the rack for 2 min. Discard supernatant.
 - (h) Add 200 μL wash buffer B.
 - (i) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on the rack for 2 min. Discard supernatant.
5. To create the second strand
- (a) Mix together 38.5 μL γ -irradiated water, 5 μL 10 \times Thermopol buffer, 2.5 μL dNTP mixture (5 mM of each), and 1 μL 100 μM extension primer CL9.
 - (b) Add mixture (47 μL) to beads in each tube. Vortex tubes and spin down briefly.
 - (c) Incubate tubes for 2 min at 65 $^{\circ}\text{C}$, and then cool for 1 min in an ice bath.
 - (d) Place tubes in an Eppendorf thermomixer set at 15 $^{\circ}\text{C}$, and then add 3 μL Bst 2.0.
 - (e) Incubate at 15 $^{\circ}\text{C}$ for 30 min, mixing 1000 RPM (*see Note 11*).
 - (f) Place tubes on the magnetic rack for 2 min, and discard supernatant.
 - (g) Add 200 μL wash buffer A.
 - (h) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on the rack for 2 min. Discard supernatant.
 - (i) Add 100 μL stringency wash buffer, and incubate for 3 min at 45 $^{\circ}\text{C}$ in a Thermomixer, mixing at 1000 RPM.
 - (j) Spin tubes briefly, then place on the magnetic rack for 2 min, and discard supernatant.
 - (k) Add 200 μL wash buffer B.
 - (l) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on the rack for 2 min. Discard supernatant.

6. Remove 3' overhangs remaining from the extension step
 - (a) Mix 84.5 μL γ -irradiated water, 10 μL 10 \times Tango buffer, 2 μL dNTP mix (5 mM each), 2.5 μL 1 % Tween-20, and 1 μL T4 DNA polymerase.
 - (b) Add mixture (100 μL) to beads in each tube, vortex, and spin down briefly.
 - (c) Incubate for 15 min at 25 $^{\circ}\text{C}$ in a Thermomixer, mixing at 1000 RPM.
 - (d) Add 10 μL stop solution to each tube.
 - (e) Place tubes on the magnetic rack for 2 min, and discard supernatant.
 - (f) Add 200 μL wash buffer A.
 - (g) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on rack for 2 min. Discard supernatant.
 - (h) Add 100 μL stringency wash buffer, and incubate for 3 min at 45 $^{\circ}\text{C}$ in a Thermomixer, mixing at 1000 RPM.
 - (i) Spin tubes briefly, place on the magnetic rack for 2 min, and discard supernatant.
 - (j) Add 200 μL wash buffer B.
 - (k) Place tubes on magnetic rack, wash by twisting tubes, and then leave tubes on rack for 2 min. Discard supernatant.
7. To ligate the second adapter
 - (a) Mix 71 μL γ -irradiated water, 10 μL 10 \times T4 DNA ligase buffer, 10 μL 50 % PEG-4000, and 2.5 μL 1 % Tween-20.
 - (b) Add mix (93.5 μL) to each tube, vortex, and spin down briefly.
 - (c) Add 5 μL 40 μM annealed adapters CL53/CL73 and 2 μL T4 DNA ligase to each tube, vortex, and spin down briefly.
 - (d) Incubate for 1 h at 25 $^{\circ}\text{C}$ in a Thermomixer, mixing at 1000 RPM.
 - (e) Place tubes on the magnetic rack for 2 min. Discard supernatant.
 - (f) Add 200 μL wash buffer A.
 - (g) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on rack for 2 min. Discard supernatant.
 - (h) Add 200 μL wash buffer B.
 - (i) Place tubes on the magnetic rack, wash by twisting tubes, and then leave tubes on rack for 2 min. Discard supernatant.

- (j) To elute the single-stranded library, add 50 μ L EBT buffer to each tube and resuspend the beads by pipetting gently up and down.
 - (k) Incubate each tube for 1 min at 95 °C in a Thermomixer, mixing at 1000 RPM, and immediately move tube to the magnetic rack.
8. Transfer the supernatant, which contains the single-stranded library molecules, to a new tube. This non-indexed, single-stranded “proto”-library must be amplified with P7 and modified P5 indexing primers (preferably barcoded) before being ready for sequencing. Libraries can be stored at -20 °C for several months.
 9. Follow the same protocol as for double-stranded library preparation (Subheading 3.4, steps 5–10) but using indexing primers P7 and modified P5.
 10. When setting up the Illumina sequencing run, be sure to replace the Read 1 sequencing primer with the custom primer CL72 using the manufacturer’s instructions for custom primers.

4 Notes

1. **Water was decontaminated by γ -Irradiation** (at least 2 kGy) using a ^{37}Cs source that had been calibrated by Fricke dosimetry [18].
2. MixG (home-made qPCR mastermix) is prepared with a high amount of BSA or HSA to allow its use with the LightCycler Carousel systems, since those systems use glass capillaries and BSA/HSA are needed to minimize DNA binding to the glass. We use the Carousel system for its sensitivity and its rapidity and to minimize carry-over contamination, each sample being in its own tube and capped after filling. The source of the albumin depends on the analyzed species. HSA is used for ancient bovine remains, and BSA in all other cases.
3. UV irradiation decontamination of reagents is performed in thin-wall, UV clear tubes for 300 s on each side in a Spectrolinker XL 1500 UV cross-linker device equipped with 254 nm UV light bulbs (Spectronics Corp., Westbury, NY, USA) corresponding to a total energy of 4.8 J/cm² [18]. The UV bulb must be at a close distance (we use 5 cm) from the tubes since the efficiency of the destruction of DNA molecules with UV light is a function of the square of the distance [18].
4. SYBR-Green I can be replaced by EvaGreen 25 mM diluted 1/80 in DMSO [18]. The -80 °C freezing step is essential.

5. Primer sequences for IonTorrent library preparation
For sequencing on an Illumina sequencer, the adapters described in **Note 7** should be used instead.
- (a) A barcoded adapter:
- Long strand: 5' CCATCTCATCCCTGCGTGTCTCC
GACTCAGXXXXXXXXXXCGAT 3'
 - Short strand: 5' ATCGXXXXXXXXXX 3'
- (b) P1 adapter:
- Long strand: 5' CCACTACGCCTCCGCTTTCCTC
TCTATGGGCAGTCGGTGAT 3'
 - Short strand: 5' ATCACCGACTGCCC 3'
- (c) Primer PCR A: 5' CCATCTCATCCCTGCGTGTCTC 3'
- (d) Primer PCR P1: 5' CCACTACGCCTCCGCTTTCCT
CTCT 3'
6. Hybridize both strands to make double-stranded adapters by mixing 20 μL of each adapter (100 μM), 5 μL of annealing buffer (10 \times), and 5 μL of γ -irradiated water. Incubate for 30 s at 95 $^{\circ}\text{C}$ in a heating block, then turn off the heat block, and allow the tubes to come to room temperature. Annealed oligos may be stored at -20°C . Final concentration of adapter is 40 μM . For IonTorrent adapters, since only the A adapter is bar-coded, it is more convenient to prepare a premix of equal volume of A and P1 adapters for each barcode.
7. Primers for the double-strand library preparation are modified from the Illumina sequencing primers, Nextera or TruSeq barcodes, and amplification primers, to change the Y-shape adapter design back to the initial Solexa design with two different adapters, as this design minimizes dimer background when working with very low DNA amounts [19]. For sequencing on an IonTorrent sequencer, the adapters described in **Note 5** should be used instead. In the example below, amplification primers are in italics, barcodes are bold, and sequencing primers are in roman fonts.
- (a) D7XX construct:
*CAAGCAGAAGACGGC***CATACGAGAT** **XXXXXXXXXX**GT
GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
- (b) D50X construct:
*AATGATACGGCGACCA***CCGAGATCTACAC**
XXXXXXXXXXACACTCTTTCCTACACG
ACGCTCTTCCGATCT
- (c) SLP5P7: 5'-AGATCGGAAGAG-3'
8. MinElute is used to allow small elution volume (here 17 μL). Columns are purchased independently of the purification kit.

9. P5s sequence CAAGCAGAAGACGGCATAACGAGAT
P7s sequence: AATGATACGGCGACCACCGAGAT
10. Oligonucleotides [10, 25]
 - (a) CL78: single-stranded adapter: 5'[Phosphate]-AGATCGGAAGXX-[TEG-biotin] (X=C18 spacer). Alternatively, ten C3 spacers can be used in the place of two C18 spacers.
 - (b) CL9: 5'-tailed extension primer:
5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT.
 - (c) CL53: double-stranded adapter strand 1: 5'CGACGCTCTTC-[ddC] (ddC = dideoxy cytidine).
 - (d) CL73: double-stranded adapter strand 2:
5'[Phosphate]-GGAAGAGCGTCGTGTAGGGAAA
GAGTGTA.
Some protocols specify four phosphorothioate bonds at the 3' end of CL73 [25]. We obtain good results without this modification.
 - (e) Modified P5 indexing primer:
5' AATGATACGGCGACCACCGAGATCTACAC[barcode]ACACTCTTCCCTACACGACGCTCTTCC
Insert standard Illumina barcode of choice in the place of [barcode]. The modified P5 indexing primer with the P7 indexing primer is used to create, via PCR, full-length bar-coded Illumina adapters flanking the initial single-stranded library products.
IMPORTANT! Modifications in the P5 barcoded primer/adapter necessitate a corresponding custom sequencing primer, CL72, to be used in place of the Illumina Read 1 sequencing primer during sequencing (CL72: 5' ACACTCTTCCCTACACGACGCTCTTCC [25]).
 - (f) P7 indexing primer: Use a standard, non-annealed, bar-coded Illumina P7 adapter oligo as the P7 indexing primer.
11. Alternatively, a standard heating block or thermocycler with a heated lid can be used instead of Eppendorf Thermomixer Comfort, but tubes containing beads must be manually mixed periodically to prevent the beads from settling.
12. All pre-PCR work is carried out in a physically isolated high-containment laboratory in a part of the building where no DNA amplification is performed. Rooms are under positive air pressure with a gradient from low to high positive pressure from the airlock, through the extraction and purification rooms, to the pre-PCR room with the highest positive pressure. Incoming air is filtered. This installation minimizes contamination with airborne environmental DNA.

13. Many techniques exist to extract DNA from soil. Authors propose different buffers, but the general purpose is to homogenize and lyse the soil in a liquid buffer mechanically (Precellys, MoBio homogenizer, etc.) or using chemicals (SDS, NLS, etc.) while limiting the interactions between DNA and soil particles using a buffer of high ionic strength (such as NaCl 1.5 M). Different DNA extraction techniques have been found to modify the representation of bacterial populations [29, 30].
14. We use MoBio PowerSoil to obtain the least variable results and to be able to perform comparisons from one extract to another, since the various DNA purification protocols extract different bacterial populations [26].
15. Buffers usually pass through the column in a few minutes. With a vacuum pump capable of producing a vacuum of -800 to -900 mbar, buffers pass through the column in 5–10 min. Clogging of the silica membrane by particles remaining in the sample can sometimes occur during the DNA binding step. When this occurs, the remaining sample can be passed through one or more fresh columns and combined after elution. If clogging continues to occur, columns can be centrifuged multiple times ($10,000 \times g$ usually for 1 min) using 700 μL of buffer per centrifugation.
16. For maximal recovery, elution can be performed using 2×75 μL preheated (50 $^{\circ}\text{C}$) EB as follows:
 - (a) Load spin columns with 75 μL preheated EB and incubate for 2 min.
 - (b) Centrifuge for 1 min at $12,500 \times g$.
 - (c) Load spin columns with 75 μL preheated EB and incubate for 1 min in the thermo-block at 50 $^{\circ}\text{C}$.
 - (d) Centrifuge for 2 min at $12,500 \times g$.
 - (e) Pool the eluates and store at -20 $^{\circ}\text{C}$.
17. The situation is more complicated since the skeletal remains are chemically not homogeneous but rather consist of multiple chemical microenvironments, each with its specific chemistry, in which DNA preservation can be variable.
18. Since there is no specific lysis procedure to open up microbial cells, it is likely that the DNA recovery from live microbial cells is low.
19. A complete overview of the process can be seen in a movie, available at http://www.univ-paris-diderot.fr/Mediatheque/spip.php?article246&var_mode=calcul, especially after the seventh minute.
20. In environmental samples, many soil compounds can interfere with molecular techniques used downstream, such as humic and fulvic substances.

21. The 16S rRNA gene includes both conserved regions, which can be used for designing amplification primers across taxa, and nine hypervariable regions (VI–V9), which can be effectively used to discriminate between taxa [27]. Nearly every hypervariable region or combination thereof has been studied. Owing to the short size of DNA fragments in degraded ancient samples, we selected V5 (28 bp long in *Escherichia coli*). Among published primers, we selected the pair providing the best coverage according to SILVA TestPrime and SILVA SSU refNR r114 [28]. We selected a forward primer (E786F) from Baker et al. [29] and a reverse primer (926r) from Watanabe et al. [30], the pair producing a 141 bp long fragment [29].
22. To prevent carry-over contamination, we systematically use dUTP instead of dTTP in qPCR mixes [31]. Incorporation of dUTP during PCR allows for elimination of amplicons from previous PCR steps when incubation with uracil-N-glycosylase (UNG) precedes each PCR. We selected codUNG from ArcticZymes (Tromsø, Norway) for its enhanced thermolability and high efficiency [18].
23. 20 pmol of adapters are present in subsequent preparation steps and must be in excess with respect to PCR products. A maximum amount of 5 pmol of amplicon products is recommended per sample.
24. This step will create blunt-ended 5' phosphorylated DNA using two enzymes: T4 DNA polymerase and T4 polynucleotide kinase. T4 DNA polymerase fills in 5'-protruding ends and removes 3'-protruding ends, thus producing blunt ends. T4 polynucleotide kinase phosphorylates the 5'-ends of DNA.
25. We use a TECAN EVO 100 and NucleoSpin 96 PCR clean-up (Macherey-Nagel ref. 740658) to achieve automated 96-well purification.
26. This step ensures ligation of DNA fragments with adapters. We use Quick Ligase to increase the ligation reaction and reduce incubation time.
27. E-Gel is preferred when amplicons have similar sizes (within about 50 bp) whereas the Caliper XT is better suited when the amplicon sizes are more heterogeneous.
28. This step allows the removal of the small fragment of the adapters and the fill-in of the 3'-protruding end of the ligated adapter. We use OneTaq from NEB, a blend of Taq and Deep VentR™ DNA Polymerases.
29. Save 1 µL for bioanalyzer analysis if desired.
30. If libraries are amplified beyond the point at which PCR starts saturating, multimers of PCR products may form due to the

cross-hybridization of library molecules via their adapter sequences, both preventing the proper determination of fragment size distribution and DNA concentration using an electrophoresis-based system and causing abnormal chimeric sequences.

31. By using 1 μL of the 40 μM adapters, considering an average size of the ancient DNA fragment in the extract of 50 bp, the adapter excess is 2.5 \times if the starting input is 500 ng and 1250 \times if the starting input is 1 ng. The adapter concentration can be modified according to the quantity of the input material.
32. For example, considering a qPCR amplification plot obtained from quantifying the library with the 1:100 dilution: (1) qPCR was performed in a 10 μL reaction volume, whereas library amplification is performed in 100 μL . Thus, 3.5 cycles should be added to allow for ten times more end product. (2) One microliter of a 1:100 library dilution was used for measurement, whereas 20 μL of the library is used for the library amplification (2000 times more). This corresponds to roughly 11 cycles that should be deducted. Thus, 7.5 cycles should be deducted from the number of cycles just prior to the beginning of the saturation phase (around roughly 75 % of the plateau height) of the 1:100 dilution amplification curve. In this way we estimate the optimal cycle numbers for PCR.
33. If excessive adapter dimers or small inserts are present in your library, additional size selection may be desired following the library purification. If AMPure XP SPRI or NucleoMag beads are used, two rounds of purification using a bead volume of 1.3 \times sample volume are recommended to best remove adapter dimers while preserving inserts of 30 bp and above, as described in Subheading 3.4, step 6.
34. For optimal results, the purified DNA extract should first be quantified with a fluorescence-based quantification method, and length distribution can be observed in most cases using an Agilent 2100 BioAnalyzer. Positive and negative control libraries should be included with each procedure. A positive control oligonucleotide should be 5'-phosphorylated and have internal primers to allow quantification with qPCR. Use 29 μL of water for a negative control.
35. This step will remove DNA with abasic sites to maximize incorporation of damaged DNA molecules into the library. To additionally remove uracils, the result of DNA cytosine deamination, 0.5 μL codUNG (1 U) may also be added at this step. If uracils are not removed, some cytosines may be improperly rendered as thymines in the final sequence. 1 μL USER enzyme (NEB) may be used in place of Endonuclease VIII and UDG.

36. To wash beads, twist tube three turns while seated in the magnetic rack.
37. For example, if preparing three samples, resuspend beads in 750 μ L.

Acknowledgments

EAB was supported by the Labex « Who am I ? ». The sequencing facility is supported by grants from the University Paris Diderot, the Fondation pour la Recherche Médicale, and the Région Ile-de-France.

References

1. Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, Morel N et al (2007) Freshly excavated fossil bones are best for amplification of ancient DNA. *Proc Natl Acad Sci U S A* 104:739–744
2. Fortea J, de la Rasilla M, Garcia-Tabernero A, Gigli E, Rosas A, Lalueza-Fox C (2008) Excavation protocol of bone remains for Neandertal DNA analysis in El Sidron Cave (Asturias, Spain). *J Hum Evol* 55:353–357
3. Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7:e34131
4. Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K et al (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104:14616–14621
5. Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins MJ (2001) Neanderthal DNA. Not just old but old and cold? *Nature* 410:771–772
6. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G et al (2005) Genomic sequencing of Pleistocene cave bears. *Science* 309:597–599
7. Geigl EM (2002) On the circumstances surrounding the preservation and analysis of very old DNA. *Archaeometry* 44:337–342
8. Geigl EM (2005) Why ancient DNA research needs taphonomy. In: O'Connor T (ed) *Biosphere to lithosphere, new studies in vertebrate taphonomy*. Oxbow Books, Oxford, pp 79–86
9. Salamon M, Tuross N, Arensburg B, Weiner S (2005) Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc Natl Acad Sci U S A* 102:13783–13788
10. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S et al (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226
11. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M et al (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74–78
12. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S et al (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49
13. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY et al (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060
14. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I et al (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514:494–497
15. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK et al (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478:506–510
16. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI et al (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341:179–183
17. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J et al (2014) Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* 46:336–344
18. Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl EM (2010) An efficient

- multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5:e13042
19. Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl EM, Grange T (2014) Library construction for ancient genomics, single strand or double strand? *Biotechniques* 56:289–290, 292–286, 298, passim
 20. Renaud G, Stenzel U, Kelso J (2015) *leeHom*, adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* 42(18):e141
 21. Jonsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) *mapDamage2.0*, fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684
 22. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
 23. Lutfalla G, Uze G (2006) Performing quantitative reverse-transcribed polymerase chain reaction experiments. *Methods Enzymol* 410:386–400
 24. Pruvost M, Geigl E-M (2004) Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *J Archaeol Sci* 31:1191–1197
 25. Gansauge MT, Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* 8:737–748
 26. Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P et al (2011) Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* 77:1315–1324
 27. Clarridge JE 3rd (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17:840–862
 28. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M et al (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1
 29. Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55:541–555
 30. Watanabe I, Kodama Y, Harayama S (2001) Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of Microbiological Methods* 44:253–262
 31. Pruvost M, Grange T, Geigl EM (2005) Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples: application to ancient DNA studies. *Biotechniques* 38 (4):569–575

2.1.2. Protocole d'extraction d'ADN à partir d'os frais

Les os frais, essentiellement des fémurs, proviennent de souris et sont systématiquement d'un poids inférieur ou égal à 100 mg. Après pesée de la pièce osseuse, il est ajouté 2 mL d'EDTA 0,5M pH8 et 1%β-mercaptoéthanol (20 µL pour un volume final de 2 mL) et 10 µL de pK à 20 mg/mL. La pièce osseuse est digérée sous agitation orbitale dans une étuve pendant 24/48h à 50°C. S'il reste des tissus après 24h, il est ajouté de nouveau 10 µL de pK à 20 mg/mL. Après digestion, les échantillons sont centrifugés 10 minutes à 8.000 g pour faire sédimenter les éventuels restes tissulaires. Le surnageant est ensuite traité pour purifier l'ADN en solution selon le protocole décrit dans « Ancient DNA in environmental ecology ».

2.2. Analyses taphonomiques

Afin de réaliser les expériences de taphonomie, nous utilisons des souris auxquelles nous injectons des agents bactériens non pathogènes afin de simuler une septicémie.

2.2.1. Protocole d'injection

Le site d'injection est la veine caudale. Afin de faciliter l'injection, les souris sont placées plusieurs minutes avant sous une lampe infrarouge, ce qui favorise la dilatation des vaisseaux veineux. La suspension de bactéries, titrée préalablement, est injectée à l'aide d'une seringue à insuline. Deux volumes sont testés, 250 et 500 µL, ce dernier étant selon la littérature le maximum compatible avec une souris.

2.2.2. Protocole d'immuno-suppression

Afin d'obtenir des souris immuno-déprimées, nous injectons du cyclophosphamide (Endoxan ® -Figure 30) à la dose de 250mg/kg à quatre souris Balb/C. Compte-tenu du poids des souris, l'injection intrapéritonéale consiste en 100µL à 40mg/mL. L'immuno-suppression est due à l'effet alkylant du cyclophosphamide sur l'ADN, empêchant sa réplication. Ainsi, les cellules immunitaires circulantes sont intoxiquées, entraînant pendant deux semaines une leucopénie.

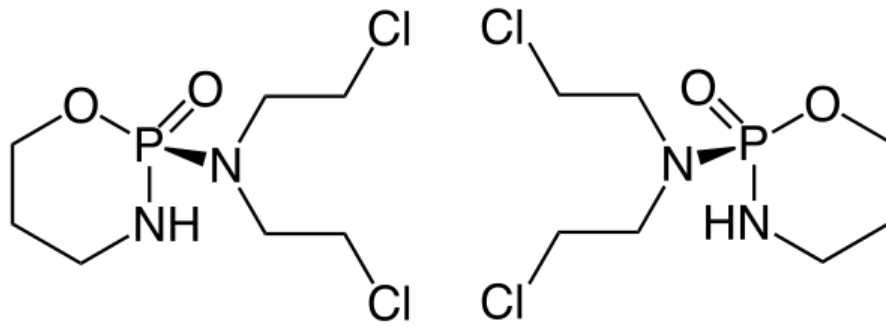


Figure 30 : formule chimique du cyclophosphamide

Trois jours après l'injection de cyclophosphamide, les souris sont soumises à injection de suspensions bactériennes.

2.2.3. Protocole expérimental

A J1, les souris subissent une injection IV de cyclophosphamide pour provoquer une immune-suppression temporaire. La dose injectée est de 200µL à 20mg/mL. Les fèces présentes dans les cages sont collectées et congelées. A J+4, 24 souris subissent une injection IV intra-caudale de 200 µL de suspension de *X. arboricola* et 24 souris de *B. atrophaeus*. Les fèces présentes dans les cages sont collectées et congelées. A J+5, les souris sont sacrifiées par injection IP de pentobarbital. Les fèces présentes dans les cages sont collectées et congelées. 12 cages métalliques sont préparées avec deux souris *X. arboricola*, 12 avec deux souris *B. atrophaeus*, 12 avec une souris saine. Les cages *X. arboricola* sont identifiées par un fil bleu, les *B. atrophaeus* par un fil rouge et les saines par un fil vert/jaune (neutre). Un lot est constitué de trois boîtes, une de chaque type, reliées entre elles et séparées de 20 cm, en triangle (Figure 31).

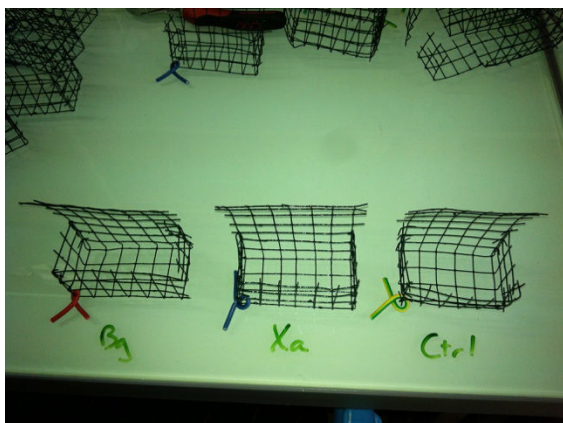


Figure 31 : exemple de cages métalliques préparées pour l'expérimentation et disposition de ces cages

Ces cages ont été enterrées le 25/01/2013 dans l'emprise de DGA MNRBC, sur un site non bâti (Figure 32) signalé et entouré d'une rubalise. Afin de compenser la disparition d'un lot (animal fouisseur, travaux, etc...), 4 lots supplémentaires ont été prévus, soit un

total de 12 lots de 5 souris. 11 lots ont été enfouis à une profondeur de 50 cm séparés de 50 cm et un lot utilisé en tant que T0. Afin de connaître l'évolution de la température dans le sol au niveau des souris, une sonde « Escort » est placée dans un contenant plastique dans une fosse à la même profondeur que les souris.

Le calendrier de prélèvements est le suivant : T0, T0+15 jours, T+1 mois, T+3 mois, T+6 mois, T+12 mois, T+18 mois, T+24 mois et T+30 mois, soit 9 lots de souris. Deux lots supplémentaires ont été prévus en cas de perte. A chaque échéance, le protocole prévoit un prélèvement de terre témoin, le relevé de température de l'Escort, et le prélèvement d'un lot, soit trois cages.

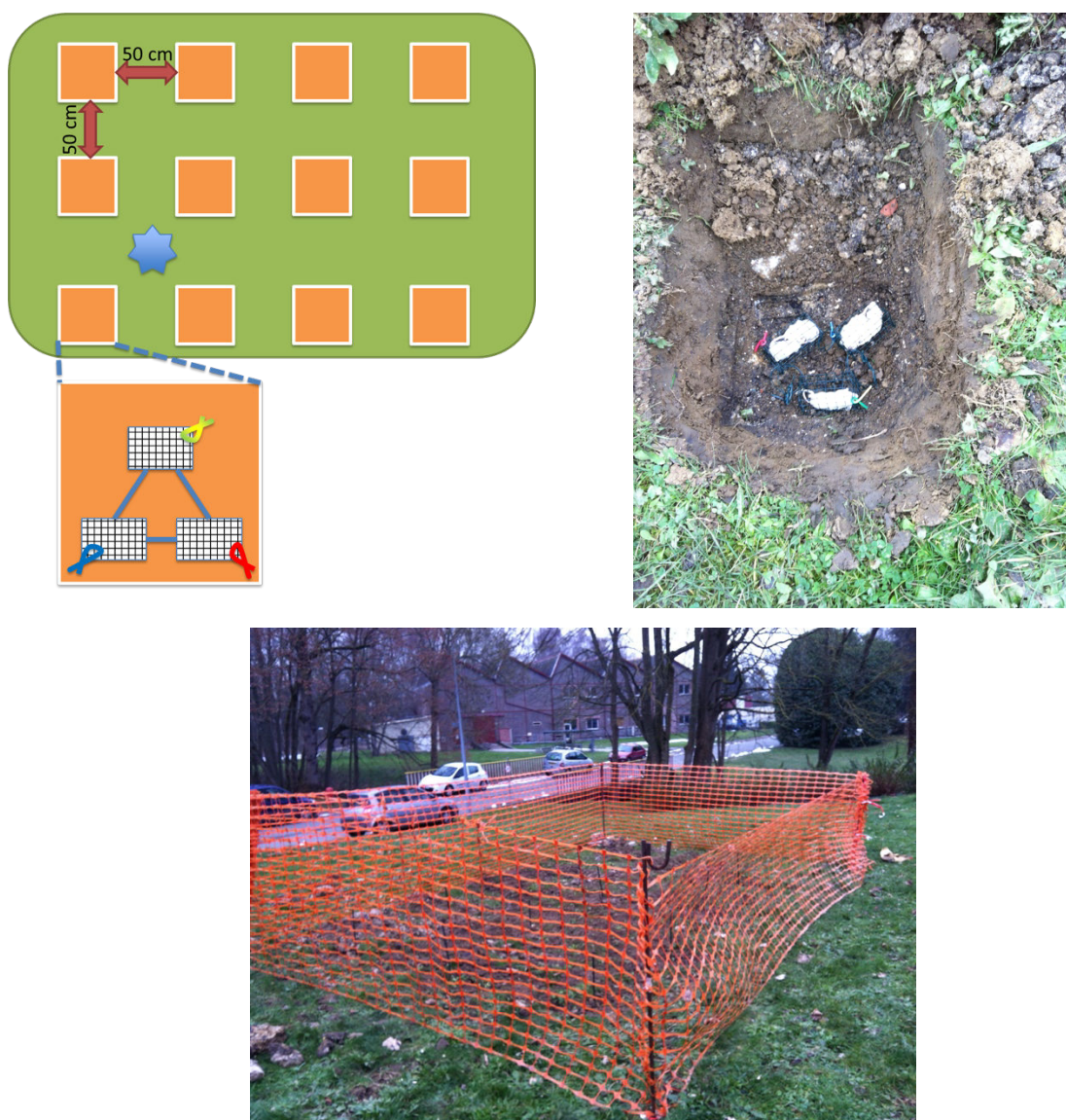


Figure 32 : plan du site d'expérimentation, avec focus sur la disposition dans une fosse et indication de la localisation du capteur de température. A gauche, disposition des cages dans les fosses d'inhumation et en bas, zone d'expérimentation

Les mesures réalisées immédiatement après prélèvement sont le pH, le dénombrement des communautés bactériennes aérobies et anaérobies dans le sol témoin et dans les

sols au contact des souris. Les souris sont ensuite disséquées et un fémur prélevé pour digestion. Un aliquot de sol est également mis de côté pour extraction d'ADN et d'ARN. Après digestion des fémurs, l'ADN est purifié et il est mené sur cet extrait la recherche des bactéries injectées, la mesure de la dégradation de l'ADN murin, la quantification globale de bactéries (gène 16S).

2.2.4. Détermination de la communauté bactérienne cultivable

Le protocole destiné à déterminer les espèces et la quantité des bactéries aérobies viables est inspiré du mode opératoire CEB N° 200514 S-CAT « culture microbiologique sur boîte de Pétri à partir d'un biocollectat ». La culture sur gélose Trypticase Soja Agar (TSA) reste la technique la plus pratique dès lors qu'il s'agisse d'estimer la charge en micro-organismes cultivables d'un échantillon environnemental en général. L'utilisation de gélose TSA permet la culture d'un spectre des micro-organismes présents le plus large possible.

Des échantillons de sol sont prélevés dans une zone témoin et à proximité immédiate des souris. Le sol étant naturellement riche en bactéries, il est nécessaire dans un premier temps de le diluer, puis de réaliser des dénombrements sur des boîtes de Pétri. Environ 1 gramme de sol (plus ou moins un tube Eppendorf de 1,5 mL) est prélevé et transféré dans un tube Falcon préalablement taré. Après pesée, l'échantillon est dilué 10 fois Poids/Volume. Ensuite, des dilutions en série de 10 en 10 sont réalisées jusqu'à 10^{-7} . 100 μ L de diluat sont étalés en dupliquat sur quatre boîtes de Petri remplies de gélose TSA. Les géloses sont ensuite placées à l'envers (couvercle en bas) dans les étuves. Par dilution, deux sont placées à l'étuve à température ambiante et deux dans une jarre anaérobie, elle-même placée dans l'étuve à température ambiante. La culture se fait à température ambiante, plus favorable pour le développement de bactéries environnementales. Le dénombrement se fait après 48 heures, afin de laisser suffisamment de temps aux colonies bactériennes pour croître et être visibles.

2.2.5. Suivi de la dégradation de l'ADN

Afin de suivre la dégradation au cours du temps de l'ADN murin, celui-ci a été amplifié en utilisant la PCR quantitative en temps réel (qPCR) à l'aide d'amorces spécifiques des espèces étudiées, en ciblant notamment l'ADN mitochondrial. Les couples d'amorces ont permis d'amplifier des fragments d'ADN de différentes longueurs et donc permettront la détermination et la quantification de l'état de dégradation des molécules d'ADN dans les

échantillons selon la stratégie de Deagle [174]. La quantité d'ADN présent dans un échantillon est mesurée par qPCR avec des couples d'amorces amplifiant des fragments de taille croissante. Selon un modèle de dégradation aléatoire, la quantité de fragments diminue exponentiellement avec l'augmentation de longueur du fragment considéré et la fréquence de dommage de l'ADN (λ) peut être estimée par la mesure de la pente de la courbe de diminution.

Une étude préliminaire menée dans notre laboratoire (Figure 33) a montré que la détermination fiable de la taille moyenne des fragments d'ADN amplifiés est possible, même si l'ADN cible est à l'état de traces au sein d'un mélange complexe.

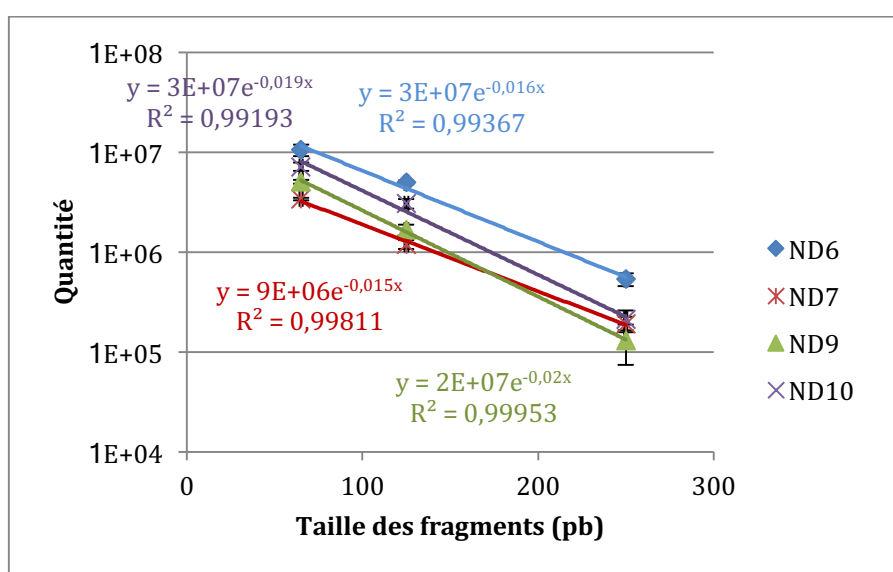


Figure 33 : Corrélation entre taille du fragment d'ADN et quantité initiale d'ADN dans des os ayant subi des traitements taphonomiques différents (résultats non publiés)

Les séquences de ces trois couples d'amorces sont décrites dans le Tableau 2. Dans le même esprit, des amplifications seront menées sur les bactéries injectées. A l'aide du logiciel Oligo7 [175], des couples d'amorces amplifiant des fragments d'ADN de taille croissante des deux bactéries utilisées sont dessinées. 7 kits ont été dessinés pour *Xa* sur deux gènes spécifiques de cette bactérie, *qumA* et *NP_Xa*. Pour *Bg*, 6 couples sont dessinés sur le seul gène spécifique de *Bg*, *Bgl1R*. Ces couples sont testés sur des extraits d'ADN de ces deux bactéries pour vérifier leur efficacité. Celle-ci est évaluée en comparant les Ct de chaque couple d'amorces pour une gamme d'ADN en dilution successive de 10 en 10. Un couple spécifique déjà existant pour chaque bactérie est rajouté. Le Tableau 2 identifie les couples retenus.

couple	cible	gène	taille		séquence F	séquence R
			mplicon			
Rd73F-Rd73R	<i>Mus musculus</i>	Mito-chondrie	73		GCATACTGGAAAGTGTGCTTG	TCTTCTGGGTGTAGGCCAGAT
Rd125F-Rd125R			125		GACCTCGATGTTGGATCAGGA	AATAGATAGAAACCGACCTGGATT
Rd154F-Rd154R			154		CTCTTAAATAGGATTGCGCTGTT	ATTTCGGTTGGGGTGACCTC
Xarb_232U18-288L18	<i>X. arboricola</i>	<i>qumA</i>	74		CGGTCAGRTCAGATCTCAA	TGACYTCGCCGCCACTGG
Xarb_5974U22-6112L22			160		CGCCTGTTACTCAGTCGCAATC	ATCGACCTCTCCTCCCGTCT
Xarb_5698U22-P8-10_R			469		CGCAGGACAAGGCTGAGCTTAC	TCAAGTCGCGTGAGCACTTC
Bg571F-635R	<i>B. atrophaeus</i>	<i>Bgl1R</i>	65		Sur demande	Sur demande
Bg_441U21-533L22			114		CCCTAGAGACAGTGATTACGA	AGGTCCTTGGATTGTTTGCTGA
Bg_189U21-533L22			366		GATGAATTATCCACCGTTAGA	AGGTCCTTGGATTGTTTGCTGA

Tableau 2 : sélection des couples d'amorces amplifiant des fragments de taille croissante dans les souris et les deux bactéries sélectionnées.

Ainsi, au fur et à mesure du temps, il a été possible d'évaluer la dégradation de l'ADN, comme décrit pour les souris, en comparant les vitesses de disparition des trois fragments de taille différente.

2.3. Recherche de pathogènes animaux par qPCR

2.3.1. Recherche de *Bacillus anthracis*

Bacillus anthracis appartient au groupe *Bacillus cereus*, qui regroupe *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. weihenstephanensis*, *B. mycoides*, *B. pseudomycoïdes* et *B. cytotoxicus* [176]. Génétiquement très proche de ces autres bacillus, il se distingue par la présence de deux plasmides porteurs de gènes de virulence, pXO1 et pXO2, ainsi qu'une mutation ponctuelle dans le gène *plcR*, régulant l'expression de facteurs de virulence [177]. La détection de *B. anthracis* repose donc essentiellement sur la mise en évidence des facteurs de virulence plasmidique, en particulier les gènes de virulence *pag*, *cya* et *lef* portés par le plasmide pXO1 et les gènes responsables de la formation de la capsule (*capA*, *capB* et *capC*) portés par pXO2 [178].

Plusieurs couples d'amorces existent et dans cette étude nous avons utilisé deux couples développés à DGA Maîtrise NRBC, le premier ciblant *capA* et le second *pagA*. Les amplicons sont de 90 pb pour *pagA* et 66 pb pour *capA*

2.3.2. Recherche de *Mycobacterium tuberculosis*

Historiquement, *Mycobacterium tuberculosis* a été le premier pathogène mis en évidence dans des échantillons anciens [13]. Plusieurs cibles ont été décrites dans la littérature, ciblant principalement les éléments d'insertion réputés spécifiques de *M. tuberculosis*, IS6110 et IS1081, qui présentent de plus l'avantage d'être présents en plusieurs copies

dans le génome [179]. Les éléments répétés en tandem, utilisés pour le spoligotypage et qui se sont révélés depuis être des CRISPRs, sont également multicopies et pourraient permettre une assignation phylogénétique [180, 181] ; Suite à la mise en évidence de ces IS dans d'autres mycobactéries [84], Bos *et al.* [40] ont proposé un nouveau gène candidat pour la détection de *M. tuberculosis*, katG, pour lesquels nous avons développé des amorces et validé un couple. Le Tableau 3 indique les séquences employées pour rechercher *M. tuberculosis* dans les échantillons anciens de cette étude.

couple d'amorces	Séquence Forward	Séquence Reverse	Réf.
IS3F-IS4R	TTCGGACCACCAGCACCTAA	TCGGTGACAAAGGCCACGTA	[38]
katG_F2-R	CGCGTGGACCTGGTCTTC	CCAGTCCTGGTTTCCGCC	cette étude
katG_F4-R4	TGCCGCCATAGAGAAAAGCAG	TTCCACGTCGGTTTGTTCCT	cette étude
Mtub_DRa-DRb	GGTTTGGGTCTGACGAC	CCGAGAGGGGACGGAAAC	[180]
mpt40_F-R2	ATAGGGAATGCTCGGCAAC	ACTTTGCGACTCCACCGAAT	cette étude
mpt40_F7-R7	AGTTCGACCTCGGATCAGCT	GTGTACCAGAACAAGGGCCT	cette étude
Mtub_IS1081_90U20_141L20	GCAGCGATGAGCGGTCCAAT	GATGGTTGCGGCACGGGTGT	cette étude
IS6110-FR	CCTGCGAGCGTAGGCGTCGG	CTCGTCCAGCGCCGCTTCGG	[182]

Tableau 3 : liste des cibles et amorces utilisées pour la détection par qPCR de *M. tuberculosis*

2.4. Recherche de spores sur ossements

Dans le cas des échantillons de Bure, Bourogne et Vellechevreux, l'archéozoologue, O. Putelat, avait pris le soin de prélever le sol au contact des ossements, qui eux-mêmes n'avaient pas été lavés. Par conséquent, dans l'hypothèse d'une infection par *B. anthracis*, en tenant compte de la capacité de cette bactérie à former des spores, il a été décidé de rechercher si des spores de *B. anthracis* pouvaient être remises en culture. L'ensemencement de milieux nutritifs par une suspension de sol se traduit par une grande diversité de bactéries et de champignons se développant sur ce milieu nutritif. Dans le cas du groupe *Bacillus cereus*, il existe des milieux sélectifs qui n'autoriseront que la pousse de bacilles appartenant à ce groupe, dont *B. anthracis*. Le milieu retenu est le « CEREUS-Ident_Agar », commercialisé par la société MERCK MILLIPORE [183]. Selon la fiche technique, le milieu favorise la pousse de *B. cereus* et *B. thuringiensis*, observables sous la forme de colonies bleu-vert de quelques millimètres de diamètre. La coloration bleu-vert est liée à la digestion d'un substrat chromogénique, le X-myoinositol-1-phosphate (Figure 34). *B. anthracis* présente des colonies de quelques millimètres de diamètre, mais blanches. Les autres espèces du genre *Bacillus* voient leur croissance inhibée ou ralentie, et les colonies sont blanches (pas de digestion du substrat chromogénique). Il est possible de trouver des souches de *B. cereus* et

B. thuringiensis ayant perdu la capacité à digérer le substrat chromogène, produisant des colonies blanches identiques à celles de *B. anthracis*.

Afin de rechercher les éventuelles spores de *B. anthracis*, des écouvillonnages secs et humides sont réalisés sur les ossements, en particulier aux endroits où les sédiments adhèrent à l'os. Des sédiments sont également recueillis dans les sacs de prélèvement.



Figure 34 : milieu sélectif Cereus-Ident-Agar ensemencé avec *B. cereus* [183]

Le protocole requiert par échantillon un bouillon BHI (Brain Heart Infusion – milieu de culture très riche) et 3 géloses sélectives (Cf. annexe 2). La stratégie de remise en culture est illustrée sur la Figure 35. Le Tableau 4 fait la liste des différents échantillonnages réalisés.

Localisation	Ossement	Echantillon
Bure	1272	Soil
		Wet Swab
		Dry Swab
	890	Soil
		Wet Swab
		Dry Swab
	1093	Wet Swab
		Dry Swab
902	Wet Swab	
	Dry Swab	
Bourogne	180-0164	Wet Swab
		Dry Swab
Vellechevreux	292-007	Wet Swab
		Dry Swab
	119-001/2/3	Wet Swab
		Dry Swab
	24-012	Soil
	fosse_26	Soil
	24-010/011	Soil
Wet Swab		
		Dry Swab

Tableau 4 : liste des échantillons réalisés à partir des prélèvements fournis par O. Putelat sur les ossements de la porte de Bourgogne.

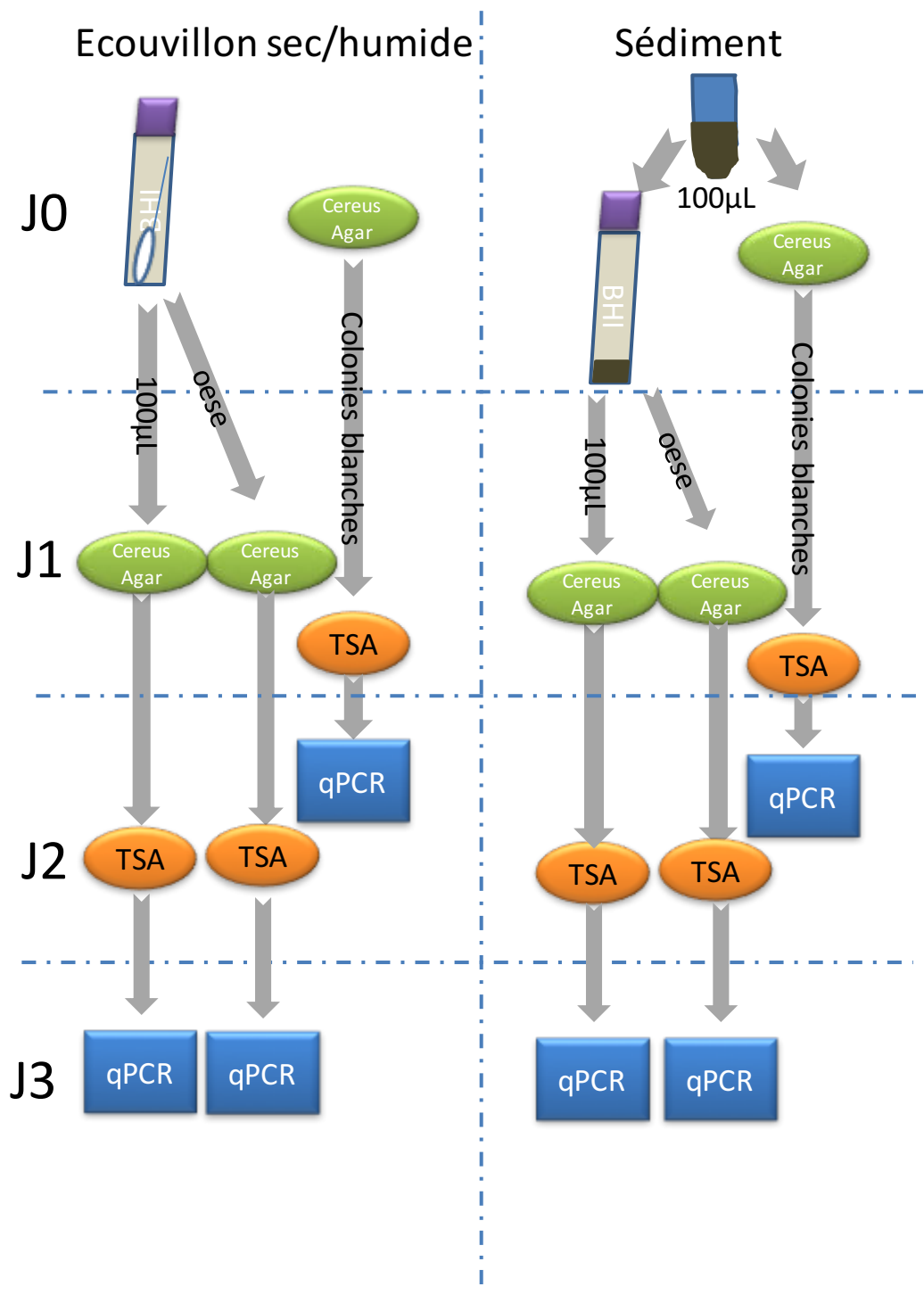


Figure 35 : stratégie de prélèvement sur échantillon sec (*i.e. os*) et de sédiments accompagnant ces os, afin de chercher à mettre en évidence des spores de *B. anthracis*.

Considérant la classification de *B. anthracis* comme agent pathogène de classe 3, la mise en culture a été réalisée dans le laboratoire BSL-3 de la DGA à Vert-Le-Petit.

Comme indiqué dans le protocole, la présence de *B. anthracis* a été vérifiée sur culture par qPCR, en cherchant à amplifier deux gènes de virulence, *pag* et *capA*, présents chacun sur un des deux plasmides de *B. anthracis*. Cette recherche s'est faite avec des

amorces et sondes développées à la DGA, sur un appareil GenExpert (Cepheid), qui procède à l'extraction des acides nucléiques et à l'amplification des cibles dans une cartouche, à l'intérieur du laboratoire BSL-3, ce qui facilite l'éventuelle mise en évidence de l'agent pathogène recherché.

2.5. Sélection des amorces 16S

Afin d'évaluer la diversité des bactéries présentes dans les échantillons, et en particulier rechercher les pathogènes, il est possible de s'appuyer sur le gène de l'ARN ribosomal 16S. Ubiquitaire et présentant des zones conservées, il est un outil puissant pour caractériser les populations bactériennes. Ce gène code pour l'ARN de la sous-unité 16S (SSU pour *Small Sub Unit*) du ribosome procaryote (bactéries et archées). Il est très utilisé pour l'identification bactérienne car il est le mieux conservé au sein du règne bactérien, il est relativement petit (1,5 kb) et est suffisamment différent de son homologue eucaryote (18S) pour permettre le dessin d'amorces spécifiques. Les ribosomes (et en conséquence l'ADN qui code pour eux) ont vu leur structure très peu modifiée au cours des âges en raison de leur rôle primordial, la traduction d'ARNm en protéines. Certaines parties ont été conservées plus que d'autres. Cela est dû à la structure du ribosome lui-même (Figure 36). Les zones repliées, ayant établi des ponts, correspondent aux régions très conservées, tandis que d'autres régions forment des boucles et correspondent aux régions hypervariables (HVR) au nombre de 9.

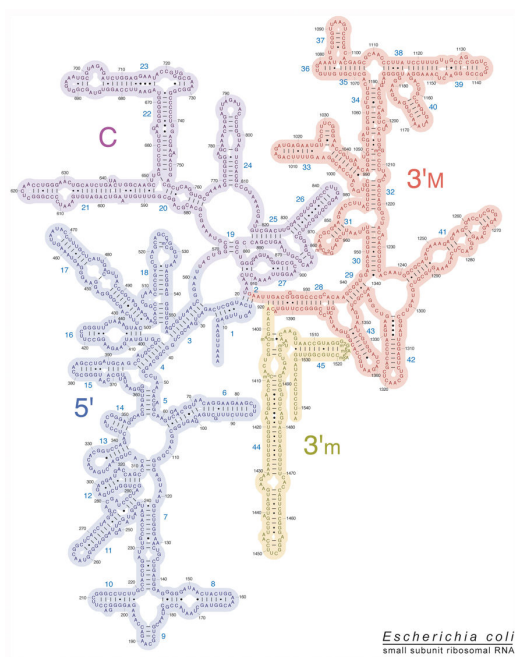


Figure 36 : structure secondaire de l'ARNr 16S

La Figure 37 représente la distribution des zones hypervariables au sein du gène ARNr 16S et l'emplacement des amorces sélectionnées pour amplifier les fragments V5 et V3-4-5 de ce gène chez *Escherichia coli*.

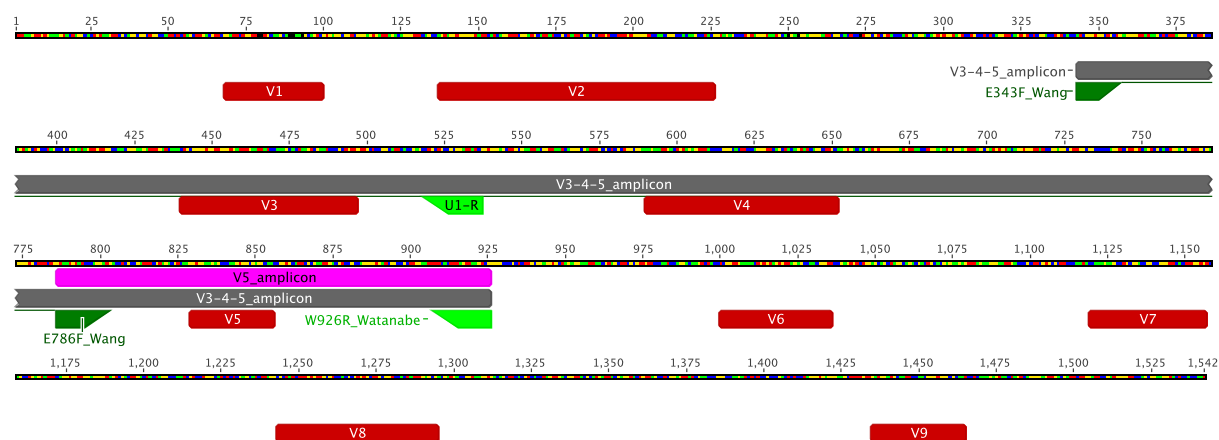


Figure 37 : séquence consensus du gène de l'ADNr 16S de *E. coli*, établi à partir des 7 opérons 16S présents dans la souche K12 - MG1655. Les 9 zones hypervariables sont de couleur rouge, les amorces de couleur verte et les amplicons de couleur violette (V5) et gris (V3-4-5).

L'identification d'espèces se base sur la diversité de séquence dans ces zones hypervariables, en comparant la séquence obtenue à celles contenues dans une base de données. Trois bases de données principales sont répertoriées : RDP (Ribosomal Database Project, <http://rdp.cme.msu.edu>), GreenGenes (<http://greengenes.lbl.gov>) et SILVA (<http://www.arb-silva.de>).

De nombreuses publications ont proposé des amorces universelles, c'est-à-dire capables d'amplifier un fragment du gène 16S chez toutes les bactéries, voire bactéries et archées. En règle générale, un fragment hypervariable ne peut être assigné à une espèce unique. Dans le cas d'analyses environnementales, où l'on s'intéresse à la richesse d'un échantillon en espèces bactériennes, ce n'est pas nécessairement un inconvénient, selon le degré de précision recherché et le niveau d'assignation taxonomique (OTU : Operational Taxonomic Unit) attendu. On s'attachera alors à amplifier un fragment hypervariable et à séquencer l'ensemble des fragments obtenus jusqu'à épuisement de la diversité. Les fragments les plus utilisés sont V3, V4 et V5-V6. Jusqu'à ces dernières années, un seul fragment était amplifié, du fait des limites technologiques liées à la longueur de fragment séquençable. Il est maintenant possible, avec les dernières évolutions des kits Illumina, de séquencer un fragment regroupant V3, V4 et V5, ce qui accroît nettement le niveau d'assignation taxonomique. Dans le cadre de cette étude, après une revue de la littérature pour collecter la majorité des amorces réputées universelles [166, 184-194], une étape de sélection *in silico* a été réalisée à l'aide de l'outil

de sélection de couples d'amorces présent dans l'outil en ligne SILVA TestPrime [186]. Cette validation a été réalisée en 2013 avec la base de données SSU r114, en sélectionnant la version NR (non redondant), qui contenait 239.346 séquences 16S issues de bactéries. Le résultat des tests est donné dans le Tableau 5.

couple d'amorces	taille	HVR	sans mismatch		1 mismatch, hors 3 dernières bases en 3'		retenu
			séquences	couverture	séquences	couverture	
E343F-R534	192	V3	189257	79	203192	84,5	
E343F-U1R	195	V3	175620	73,4	202113	84,5	✓
E517F-U806R	290	V4	176825	73,9	216059	88,2	✓
E518F-U806R	289	V4	177433	74,1	216008	88,1	
E783F-E939R	157	V5	121057	50,6	181650	75,9	
E786F-W926R	141	V5	169429	70,8	208838	87,1	✓
E939F-U1053R	130	V6	89616	37,4	204250	85,3	
E949F-U1053R	120	V6	88052	36,8	205272	85,8	
E969F-U1053R	100	V6	80165	34	204250	85,3	
E969F-U1063R	110	V6	153500	62,9	207899	85,5	✓
E970F-U1053R	99	V6	71304	30	218270	89,4	
E343F-U806R	464	V3-V4	175552	73	207515	86,7	
E343F-K785R	462	V3-V4	193234	81	216008	90,2	
K341F-K785R	465	V3-V4	200755	84	225460	91,4	✓
E343F-W926R	584	V3-V5	418320	83,5	467956	93,5	✓
K341F_Wi1064R		V3-V6	202323	83	40021		
Wi356F-Wi1064R	744	V3-V6	197688	83	2315		
E786F-Wi1064R	296	V5-V6	166952	70	37640	15,4	
E786F-U1053R	283	V5-V6	185221	77	221590	88,9	✓

Tableau 5 : liste des couples d'amorces testés via TestPrime afin de déterminer les meilleures associations. A noter que le couple E343F-W926R a été testé sur SSU r123 RefNR, riche de 500.769 séquences bactériennes.

La séquence de chacune des amorces retenues est donnée dans le Tableau 6. Il est à noter la nécessité d'employer des bases dégénérées dans un certain nombre de cas, afin de couvrir au mieux la diversité bactérienne, même dans les zones les plus conservées.

Amorces	Séquence (5'→3')
E343F [192]	TACGG R AGGCAGCAG
K341F [186]	CCTACGGG N GGCWGCAG
Wi356F [191]	AC W CCTACGGG W GGC W GC
U1R[193]	CGTATTACCGCGGCTGCTGG
C515F [195]	GTGCCAGC M GCCGCGGTAA
E517F [192]	G Y C AGC M GCCGCGGTAA
R534 [192]	ATTACCGCGGCTGCTGGC
K785R [186]	GACTAC H VGGGTATCTAATCC
E786F [184]	GATTAGATACCCTGGTAG
U806R [194]	GGACTACCAGGGTATCTAAT
T347R [188]	TGGCCGTACTCCCCAGGCG
W926R [191]	CCGTCAATTC M TTTGAGTTT
E939F [184]	CTTGTGCGGGCCCCCGTCAATTC
E969F [166]	ACGCGA R GAACCTTACC
U1053R [184]	CTGACG R C R GCCATGC
E1063R [166]	CTCACG R CACGAGCTGACG
Wi1064R [191]	AYCTCACGRCACGAGCTGAC

Tableau 6 : liste des amorces analysées pour déterminer les meilleurs couples. Les bases dégénérées sont repérées par leur couleur rouge. La localisation des séquences est donnée par leur position 5' sur *E. coli*.

Plusieurs couples ont été choisis par fragment retenu. Ces couples ont été testés sur un panel de bactéries afin de vérifier leur efficacité. Il a été sélectionné une souche d'*E. coli*, une souche de *B. atrophaeus* et une souche de *M. tuberculosis*.

Les conditions de PCR ont été standardisées dans la mesure du possible pour chacun des couples.

2.6. Bio-informatique

Les amplifications ont été réalisées sur thermocycleurs de type LightCycler Carousel (1.5 et 2.0) et LightCycler480, commercialisés par la société Roche. Les données de qPCR ont été analysées avec les logiciels de pilotage des thermocycleurs, en particulier LCS v3.1, LCS v4.1 et LCS480 v1.5.1.

L'analyse des données produites par les séquenceurs de nouvelle génération, en particulier le MiSeq de la société Illumina, requiert des outils informatiques particuliers. Les différentes étapes de traitement de données peuvent être réalisées à l'aide d'algorithmes ou de pipelines. Dans cette étude, trois pipelines principaux ont été développés, les deux premiers dédiés à l'analyse de données shotgun, le dernier à l'exploitation taxonomique des amplicons 16S. Plusieurs logiciels ou scripts sont

communs à tous les pipelines, tandis que certains sont spécifiques. Avant de décrire les pipelines sont décrits tous les outils informatiques employés.

2.6.1. FastQC

Ce logiciel a pour but de réaliser les premiers contrôles qualité sur les séquences brutes produites par les séquenceurs de seconde génération. Le logiciel dénombre la quantité de séquence, leur gamme de longueur, le taux de GC et propose des représentations graphiques des scores de qualité, de la répartition en GC, des bases indéterminées, et de la distribution de la longueur des fragments. Il peut aussi identifier la présence d'adaptateurs ou de séquences répétées.

2.6.2. leeHom

leeHom [172] permet de reconstruire les fragments à l'origine des banques de séquençage. Il permet de couper les adaptateurs encore présents et de fusionner les amplicons issus de séquençage « Paired End ». Par défaut, le logiciel est optimisé pour la reconstruction de séquences à partir d'ADN ancien, mais son application peut être beaucoup plus large. Ce logiciel utilise une approche probabiliste bayésienne a posteriori, pour reconstruire les séquences à partir de fragments se recouvrant. Il peut être utilisé pour reconstruire des fragments longs à partir d'ADN moderne, mais il est alors nécessaire de déterminer la distribution de taille des inserts.

2.6.3. PandaSeq

PandaSeq [196] est un logiciel permettant la reconstruction des inserts lorsqu'il y a recouvrement de séquences lors de lecture *Paired-End*. Il est possible de déterminer un taux d'erreur autorisé dans les adaptateurs à exciser. Nous l'avons surtout utilisé pour exciser les amorces d'amplification du 16S.

2.6.4. Cutadapt

Cutadapt [197] permet de couper les adaptateurs, en 3' et/ou en 5', en pouvant moduler la taille et le taux d'erreur présent dans ces séquences à exciser. Il permet également d'appliquer des critères de qualité, telle qu'une longueur minimale du fragment après excision ou un indice de qualité (score phred). Il peut également éliminer un nombre déterminé de bases en 5' et en 3'. Le logiciel, depuis la version 1.9, est capable de travailler sur des séquences appariées et de maintenir cet appariement au fil de la sélection basée sur la longueur et la qualité des séquences.

2.6.5. FastX toolkit

Cet outil est une compilation de scripts permettant la manipulation de données de séquençage, afin de les affiner et de faire des tris sur la qualité, d'exciser les adaptateurs, d'enlever des bases en 5' et en 3'. Il est possible également de modifier les formats (fastq vers fasta) et de réaliser des analyses simples sur les séquences, comme la distribution des tailles de fragments séquencés ou les taux des différents nucléotides.

2.6.6. BWA

bwa (pour Burrow-Wheeler Alignment tool) [198] est un outil d'alignement de courtes séquences issues de séquenceurs de seconde génération sur une séquence de référence. Cette séquence de référence peut être un chromosome, un génome, un assemblage de séquences ou tout groupe de séquences fasta. Cette référence est indexée par un outil de *bwa*. Deux outils permettent ensuite l'alignement des séquences obtenues à la séquence de référence, *bwa mem* et *bwa backtrack*. Les deux outils peuvent travailler avec des séquences *Paired-End*, *bwa mem* étant optimisé pour les fragments de plus de 70 pb tandis que *bwa backtrack* (connu aussi comme *bwa sampe/samse*) est optimisé pour des fragments de taille maximale 100 pb. On obtient en sortie un fichier sam, fichier tabulé qui contient les informations relatives à la comparaison entre les fichiers de séquençage et la séquence de référence.

2.6.7. Picard

picard ([http://picard.sourceforge.net./](http://picard.sourceforge.net/)) est un groupe d'outils en ligne de commande écrits en java permettant la manipulation de données produites par un logiciel d'alignement tel que le tri de séquences, le filtrage selon la qualité, l'exclusion des duplicats, l'analyse des séquences (taille d'insert, taux de GC, qualité, nombre de séquences, SNPs, couverture...).

2.6.8. MetaPhlAn

MetaPhlAn [199] est un programme qui permet une assignation rapide des données de séquençage massif en comparant les séquences obtenues à une base de données restreinte, constituée de séquences spécifiques de chaque niveau taxonomique. Ainsi, chaque séquence obtenue sera comparée à cette base de données. Une séquence largement partagée (gène de ménage, ARNr 16S, ...) sera écartée. Lorsque la séquence sera présente dans la base de données, elle permettra de produire un résultat, qui aura

un niveau taxonomique défini selon la spécificité de la séquence (espèce, genre, famille, ...).

2.6.9. Samtools

Samtools [200] est un outil de manipulation des fichiers sam et bam obtenus après comparaison de séquence entre fichiers de séquençage et séquence de référence. Il permet de compresser ces fichiers, d'éliminer les séquences non alignées, isoler un alignement particulier, supprimer les duplicats et produire des rapports statistiques sur les quantités et longueurs des séquences analysées.

2.6.10. Krona

Krona [201] est un outil de visualisation de données sous forme de camembert interactif hiérarchisé. La représentation graphique est créée à partir de fichiers excel ou csv mis en forme et est visualisé dans un navigateur web, ce qui permet pour son application aux données métagénomiques une exploration dynamique des différents niveaux taxonomiques obtenus à partir des données de séquençage.

2.6.11. usearch

usearch [202, 203] est un outil d'analyse de séquences permettant à partir d'amplicons 16S la recherche, le nettoyage et l'identification taxonomique des amplicons. Il est composé de plusieurs algorithmes qui peuvent être utilisés seuls ou en cohérence, permettant l'analyse depuis la banque de séquençage jusqu'à l'attribution taxonomique de données issues d'amplicons.

2.6.12. mapDamage

L'ADN ancien est abimé, et il a été démontré que le taux de conversion de cytosine en thymine est plus élevé en 5' des fragments. *mapDamage* [49, 50] recherche ces profils de mutations dans les séquences d'ADN issues de séquençage massif. Il permet ainsi de vérifier l'authenticité des fragments séquencés. Il mesure également la fréquence des bases autour des extrémités et permet de visualiser la fréquence de cassures des fragments en fonction des bases.

2.6.13. Scripts

2.6.13.1. fastq_trimmer_by_quality.py

Ce script en langage python permet un filtrage des fragments séquencés selon leur qualité sur une fenêtre glissante. Il a été repris des scripts intégrés dans Galaxy [204].

2.6.13.2. GC-boxplot

Ce script en langage R a été développé par Tito Candelli à l'Institut Jacques Monod pour l'équipe Grange/Geigl. A partir d'une liste de séquences, ce script calcule la longueur et le taux de GC de chaque séquence, les répartit par intervalles de longueur et dessine un graphe boxplot des pourcentages en GC pour chacun des intervalles et associe une couleur reflétant le nombre de séquences par intervalle. Ce script permet de comparer les distributions de longueur et de taux de GC de différents échantillons et permet de comparer par exemple des échantillons avant et après alignement sur un génome de référence.

2.6.14. Pipelines bioinformatiques (annexe)

Comme indiqué plus haut, trois « pipelines » bioinformatiques ont été créés, deux dédiés à l'analyse de données génomiques, le troisième aux amplicons 16S.

Le premier pipeline va utiliser les données pré-traitées, à l'étape de fichier sam. Après traitement par leeHom et cutadapt, les séquences ayant passé ces filtres sont traitées par bwa mem contre un génome artificiel, concaténant l'ensemble des génomes bactériens disponibles dans la base NCBI en décembre 2012 (4200 séquences). Le fichier résultant est filtré pour ne retenir que les séquences ayant obtenu un équivalent dans le génome synthétique. Le script sam2taxo.pl est ensuite exécuté sur ce fichier sam. A l'issue, on obtient une liste pondérée des bactéries identifiées à partir des fragments séquencés avec leur assignation taxonomique. On recueille également les données statistiques basiques comme le nombre de séquences par bactérie, la longueur de ces séquences, la longueur moyenne, ...

Le second pipeline va utiliser les données prétraitées par leeHom et cutadapt et les faire analyser par metaphlan. A l'issue, on obtient une table d'assignation taxonomique des fragments séquencés, une heatmap comparant les échantillons si plusieurs sont analysés en parallèle, et un fichier regroupant les séquences ayant trouvé une correspondance dans la base de données metaphlan, avec leur identification.

Le troisième pipeline va analyser les données de séquençage d'amplicons 16S. Le script `usearch2taxo` est exécuté sur ces fichiers. Ce script va permettre, après avoir filtré les fragments séquencés, excisé les amorces ayant servi à l'amplification et rassemblé les fragments Paired End par `pandaseq`, d'exclure les chimères, les singletons, de proposer une assignation taxonomique et de produire des fichiers csv et tsv utilisables par `krona` pour produire des graphes représentant la diversité taxonomique. Les fichiers seront également concaténés pour pouvoir être analysés sous excel et permettre une analyse temporelle.

TAPHONOMIE – RÉSULTATS ET DISCUSSION

1. VALIDATION DE LA PRESENCE D'ADN BACTERIEN DANS LES OS

Il a été vérifié que l'ADN de *Bacillus anthracis* était détectable dans des os de souris exposées à un aérosol, prouvant la pertinence de la recherche d'acides nucléiques dans les os (Figure 38).

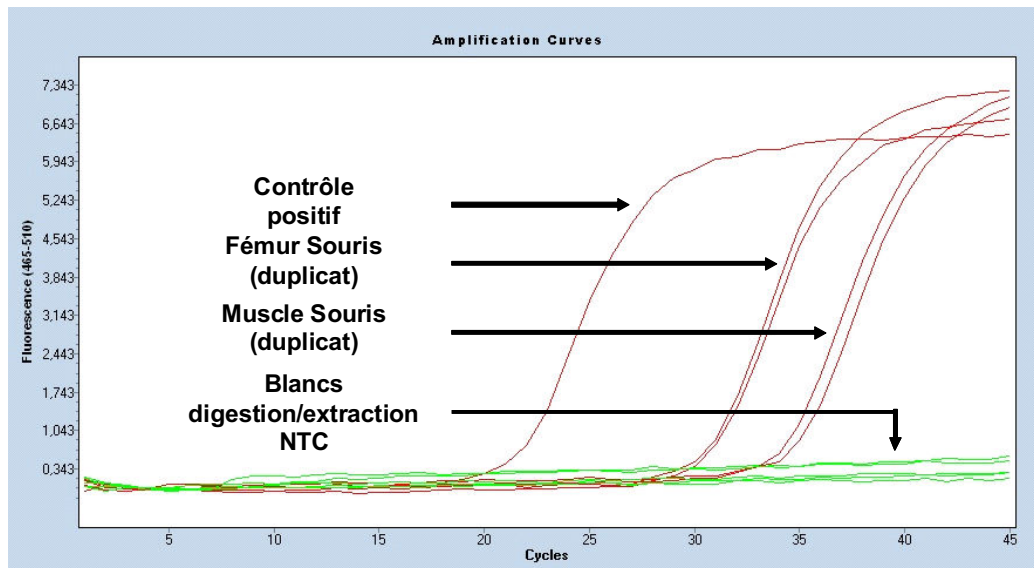


Figure 38 : détection du plasmide pXO1 dans le fémur et un muscle de la cuisse d'une souris exposée au bacille du charbon (données non publiées).

La présence de *B. anthracis* est plus importante dans le fémur que dans le muscle, mais il n'avait pas alors été recherché de corrélation entre la masse de l'échantillon de départ et la quantité d'ADN de *B. anthracis* détecté. Les souris sont décédées suite à l'exposition à *B. anthracis*, mais on note une concentration faible d'ADN de cette bactérie dans l'os (Ct aux environs de 30).

2. VALIDATION DU PROTOCOLE DE SEPTICEMIE

4 souris BALB/C ont été soumises à injection de deux suspensions bactériennes, Bg et Xa à $2,9 \cdot 10^9$ et $1,4 \cdot 10^9$ UFC/mL respectivement. Deux volumes différents ont été injectés, 250 et 500 μ L. La souris ayant reçu $1,5 \cdot 10^9$ UFC de *B. atrophaeus* est décédée presque immédiatement tandis que la souris ayant reçu $7 \cdot 10^8$ CFU de *X. arboricola* est décédée 20 heures après injection. Le fémur, la mandibule inférieure, le muscle de la cuisse et un tronçon de queue ont été prélevés. La concentration en ADN total a été mesurée par Qubit et la recherche des bactéries par qPCR, en visant les gènes décrits (*qumA* et *NP_A* pour *X. arboricola*, *Bgl1R* pour *B. atrophaeus*). Le Tableau 7 décrit les résultats obtenus.

	Queue			Muscle		Femur		Mandibule	
	ADN total	eq.genome/g de tissu		ADN total	eq.genome/g de tissu	ADN total	eq.genome/g de tissu	ADN total	eq.genome/g de tissu
souris 1	X.a 7.10 ⁸	9,12.10 ⁴	2,4.10 ⁸	ND	2,3.10 ⁶	1,76.10 ⁴	2,8.10 ⁵	7,5.10 ³	2,3.10 ⁵
souris 2	X.a 3,5.10 ⁸	8,12.10 ⁴	6,7.10 ⁹	5,18.10 ⁴	0,0E+00	1,63.10 ⁴	0,0E+00	8,5.10 ³	0,0E+00
souris 3	B.g 1,5.10 ⁹	8,78.10 ⁴	3,2.10 ⁶	8,16.10 ⁴	6,7.10 ⁵	6,12.10 ⁴	1,8.10 ⁵	3,48.10 ⁴	6,5.10 ⁵
souris 4	B.g 7,5.10 ⁸	9,08.10 ⁴	2,1.10 ⁶	4,16.10 ⁴	0,0E+00	2,02.10 ⁴	0,0E+00	1,11.10 ⁴	0,0E+00

Tableau 7 : ADN total et équivalents génomes des bactéries injectées. ND : Non déterminé. En vert les échantillons négatifs pour les agents injectés.

L'agent injecté a été détecté dans la queue (site d'injection) pour toutes les souris. Cette quantité était très importante pour *X. arboricola*, ce qui pouvait être le signe d'une infection locale. Les souris ayant reçu les doses les moins importantes ont complètement éliminé les bactéries de leurs tissus, hors site d'injection. Pour les doses importantes, les souris sont décédées immédiatement ou dans les minutes suivant l'injection. Il était donc impossible de simuler une septicémie avec ces souris BALB/C. Nous avons alors testé des souris immuno-déprimées.

Afin d'obtenir des souris immuno-déprimées, nous avons injecté du cyclophosphamide (Endoxan®) à la dose de 250 mg/kg à quatre souris BALB/C. Compte-tenu du poids des souris, l'injection intrapéritonéale a consisté en 100 µL à 40 mg/mL. L'immunosuppression est due à l'effet alkylant du cyclophosphamide sur l'ADN, empêchant sa réplication. Ainsi, les cellules immunitaires circulantes sont intoxiquées, entraînant pendant deux semaines une leucopénie.

Trois jours après l'injection de cyclophosphamide, les souris ont été soumises à injection de bactéries. Les suspensions bactériennes (*B. atrophaeus* : 1,6.10¹⁰ bactéries/mL et *X. arboricola* : 6,7.10¹⁰ bactéries/mL) ont été diluées au 1/10. Le volume d'injection est réduit à 100 µL pour éviter un choc septique. Les souris immunodéprimées ont été remises en animalerie jusqu'au lendemain. Le lendemain, une des souris ayant reçu *B. atrophaeus* est morte, tandis que les autres sont prostrées et respirent difficilement. Le fémur, la mandibule inférieure, le muscle de la cuisse et un tronçon de queue ont été prélevés après euthanasie.

		Queue	Muscle	Femur	Mandibule
		eq.genome/g de tissu	eq.genome/g de tissu	eq.genome/g de tissu	eq.genome/g de tissu
souris 5	X.a 6,7.10 ⁹	2,8.10 ⁶	4,0.10 ³	1,8.10 ⁴	6,6.10 ³
souris 6	X.a 6,7.10 ⁹	7,4.10 ⁴	8,8.10 ³	1,5.10 ⁵	3,7.10 ⁴
souris 7	B.g 1,6.10 ⁹	1,1.10 ³	3,3.10 ²	2,2.10 ³	8,7.10 ²
souris 8	B.g 1,6.10 ⁹	8,1.10 ²	6,0.10 ²	1,6.10 ³	5,1.10 ²

Tableau 8 : dosage de *B.atrophaeus* et *X.arboricola* dans les échantillons de souris immunodéprimées

Dans le cas des souris immuno-déprimées, les bactéries restent détectables dans l'ensemble des tissus 24 heures après l'injection. Leur concentration est comparable dans tous les tissus, sauf dans la queue, site d'injection (Tableau 8). Ce modèle peut donc être utilisé.

3. PROTOCOLE TAPHONOMIE A COURT TERME

La première expérience d'enfouissement a débuté en janvier 2013 et les mois suivants ont été caractérisés par un temps froid, comme attesté par la Figure 39.

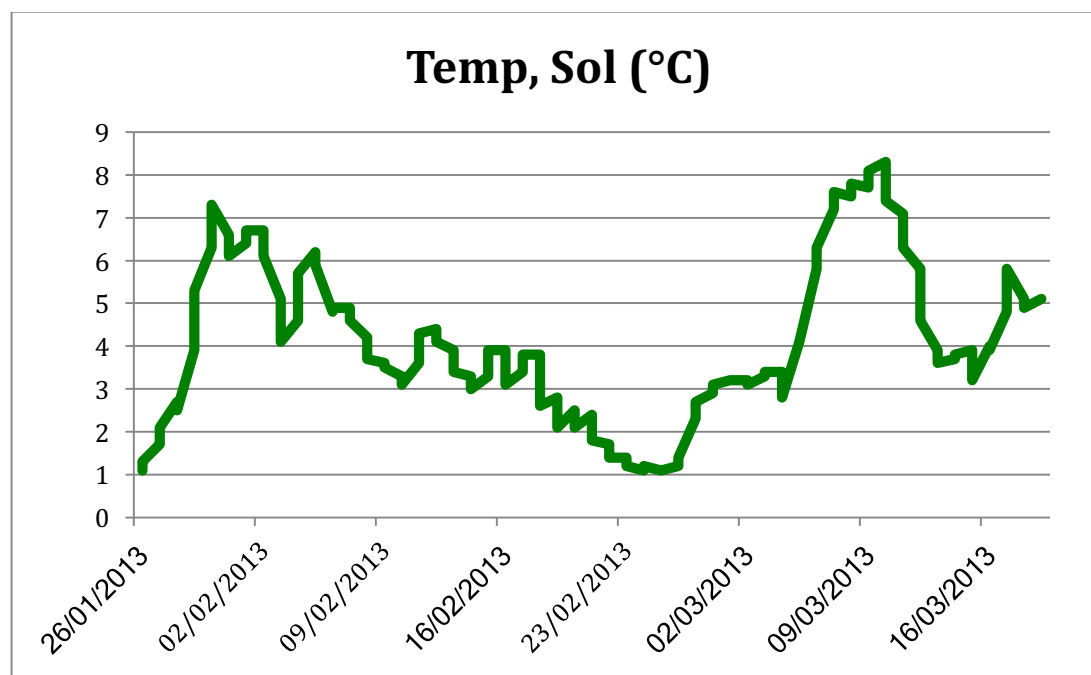


Figure 39 : évolution de la température dans le sol au cours des quinze premières semaines de la première expérimentation.

Les phénomènes de dégradation ont été ralentis lors des premiers prélèvements, du fait de ce froid, et se sont accélérés à partir de début avril 2013. L'aspect des souris exhumées après 15 jours en témoigne (Figure 40). C'est comme si les dépouilles étaient restées au réfrigérateur.



Figure 40 : aspect des souris après 15 jours d'inhumation dans un sol froid. A gauche deux souris dans la cage de contention, à droite une souris après ouverture de la cage.

Afin de tracer plus finement les phénomènes précoces de dégradation, nous avons mis en place une deuxième expérimentation, à une période plus chaude, et en modifiant le protocole. Comme nous cherchions à caractériser les phénomènes de dégradation et de colonisation des restes par la communauté bactérienne tellurique ou intestinale à court terme, les souris n'étaient plus infectées par les bactéries *X. arboricola* et *B. atrophaeus*, ce qui ne nécessite plus d'immunosuppression. Un article publié après le lancement du premier protocole montrant des différences entre souris posées sur un sol stérile et sur un sol naturel [155], nous avons aussi décidé de placer des souris à la surface du sol tout en étant protégées des charognards. Ainsi le protocole prévoyait 5 lots de souris enterrées, 5 lots de souris en surface et un calendrier de prélèvement permettant de suivre les phénomènes précoces de dégradation sur deux mois, à savoir J0, J+7, J+14, J+28, J+42 et J+56. La température de l'air et celle du sol ont été enregistrées et les prélèvements identiques à ceux du premier protocole ont été effectués.

3.1. Suivi de la température

L'évolution de la température est plus régulière, avec un passage de 7 à 13°C dans le sol (Figure 41). Les variations dans l'air étaient importantes entre le matin et le soir, mais la courbe de tendance, mesurée selon une loi polynomiale, est corrélée à la courbe dans le sol, les montées en température dans l'air précédant celles dans le sol, du fait de l'inertie relative de ce dernier.

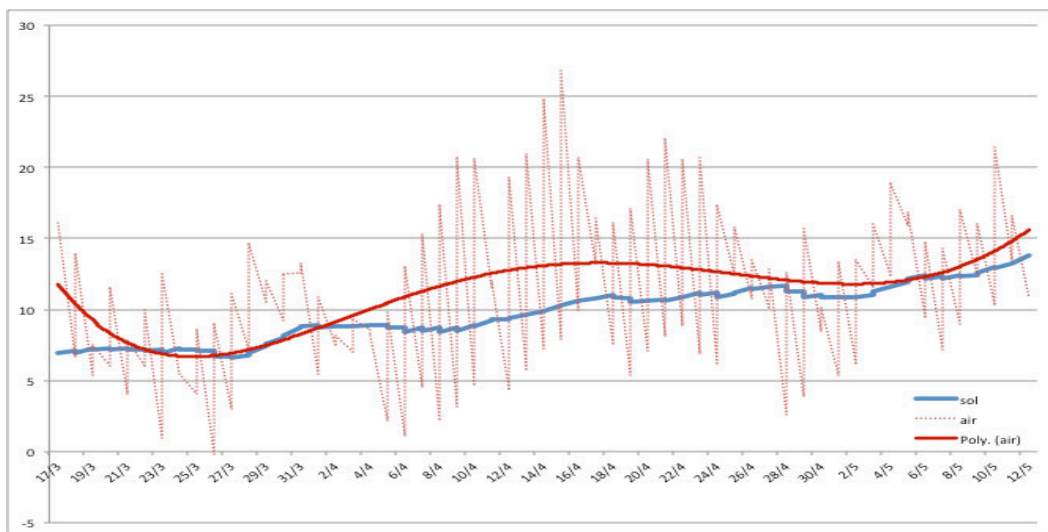


Figure 41 : évolution de la température dans le sol et à la surface au cours de la seconde expérimentation. Les fortes variations de températures pour l'air (courbe rouge en pointillé) étaient liées à la fréquence de prélèvement, biquotidienne (6H-18H).

3.2. Observations macroscopiques

Le Figure 42 présente en parallèle l'évolution des carcasses de souris laissées à la surface et enterrées tout au long de l'expérimentation.

Au fur et à mesure du temps, l'évolution des carcasses a été différente selon qu'elles soient restées à la surface ou enfouies.

A J+7, les souris en surface étaient d'aspect encore comparables à J0. Elles commençaient à se recroqueviller, mais il n'y avait pas de signes extérieurs éloquentes. Les mêmes observations ont été faites pour les souris enterrées.

A J+14, des boursouflures ont pu être observées au niveau de l'abdomen, mais il n'a pas semblé y avoir de rupture de l'enveloppe corporelle. Pour les souris enfouies, le processus de dégradation était beaucoup plus avancé, avec la peau qui se séparait des tissus sous-jacents, et une fonte des chairs.

A J+28, la dépouille à l'extérieur grouillait d'asticots, résultant de l'éclosion des œufs pondus par divers insectes depuis le dépôt de la dépouille. La carcasse était cireuse, le poil collé au corps. Pour la souris enfouie, elle ressemblait à la dépouille à J+14, avec une perte plus importante de chairs et une perte complète d'adhérence de la peau.

A J+42, le corps des souris en surface était cireux, les os commençant à poindre à travers la peau, et les chairs semblant avoir fondu autour des os. Pour les souris enfouies, il ne restait plus que la peau et les os, la silhouette était très fine. Des asticots étaient présents dans le corps et alentour.

Surface



Temps

J0

Enfouie



Système de protection des souris de surface contre les charognards opportunistes



J7

Surface : pas de modification visible

Enfouies : pas de modification visible



J+14

Surface : apparition de boursouflures

Enfouies : peau se sépare des tissus



J+28

Surface : carcasse cireuse, nombreux asticots

Enfouies : comparable à J14





J+42

Surface : disparition des chairs, os percent la peau

Enfouies : chairs ont disparu, présence d'asticots



J+56

Surface : seuls la peau et les os demeurent

Enfouies : peau partiellement dégradée, os apparaissent sans chair adhérente



Figure 42 : dispositif expérimental et évolution de l'aspect des souris au cours des 56 jours de l'expérimentation. Les souris laissées à la surface sont à gauche et les souris enfouies à droite.

A J+56, la dégradation est devenue très importante, avec dans la cage comme restes essentiellement la peau, la queue et les os. Les os étaient nettement visibles. Dans un cas, toute la carcasse avait été digérée et seul le crâne et quelques os parmi les plus importants étaient encore présents dans la cage. Pour les souris enfouies, la peau se dégradait, les os commençant à apparaître sans chair adhérente.

3.3. Observations microscopiques

Nous avons pu travailler en partenariat avec le Dr. Yolanda Fernandez-Jalvo, taphonomiste du département de paléobiologie du Museum National des Sciences Naturelles de Madrid, spécialisée en imagerie microscopique et électronique d'ossements et de dents. Plusieurs échantillons d'ossements, avec des durées d'enfouissement différentes dans le sol, lui ont été soumis pour analyse. L'état de dégradation est quantifié selon un indice, appelé OHI pour Oxford Histological Indice [205]. Le Tableau 9 rappelle les différents degrés d'altération des os, et les OHI correspondants.

Index	Approx. % of intact bone	Description
0	<5	No original features identifiable, other than Haversian canals
1	<15	Small areas of well preserved bone present, or some lamellar structure preserved by pattern of destructive foci
2	<33	Clear lamellate structure preserved between destructive foci
3	>67	Clear preservation of some osteocyte lacunae
4	>85	Only minor amounts of destructive foci, otherwise generally well preserved
5	>95	Very well preserved, virtually indistinguishable from fresh bone

Tableau 9 : valeurs d'index histologique d'Oxford correspondant au degré de modification diagénétique (adapté de [205]).

Nous lui avons fourni des os frais, sans passage par le sol. Ces os ont été extraits de souris sacrifiées et ont été grossièrement décarnés très peu de temps après la mort de l'animal, mais sans traitement dédié à l'enlèvement des tissus associés à l'os, afin d'éviter de modifier la surface de l'os. Les os ont été plongés dans de l'éthanol pur afin d'empêcher la croissance de microorganismes pendant le transport. Parmi les clichés pris, nous avons retenu deux photographies (Figure 43).

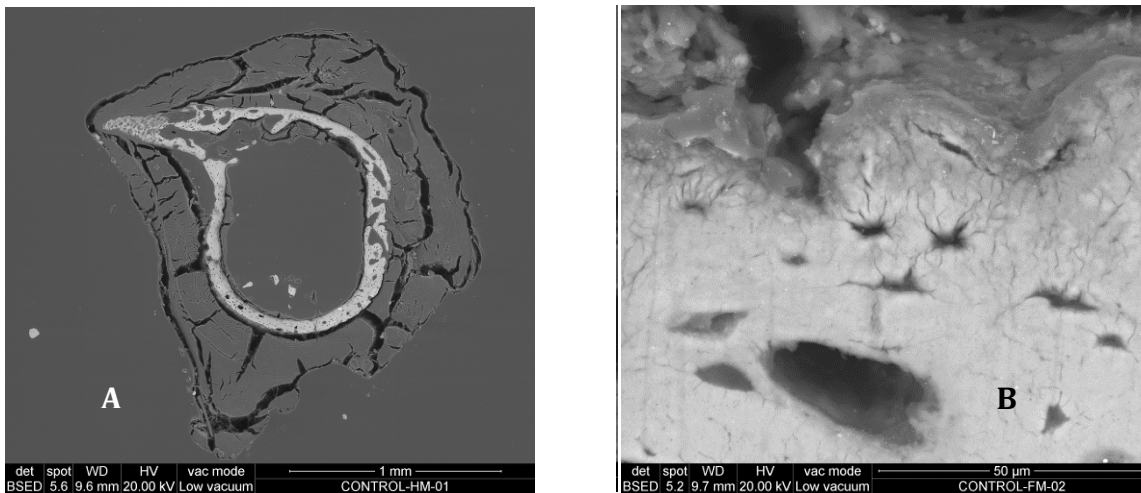


Figure 43 : le cliché A montre la section de l'humérus, où l'on voit les chairs entourant encore l'os. Le cliché B montre la surface de l'os, avec un canal haversien (flèche rouge) et des ostéoblastes (clichés : Yolanda Fernandez-Jalvo).

Le second échantillon analysé a été prélevé après 6 mois d'enfouissement (Figure 44).

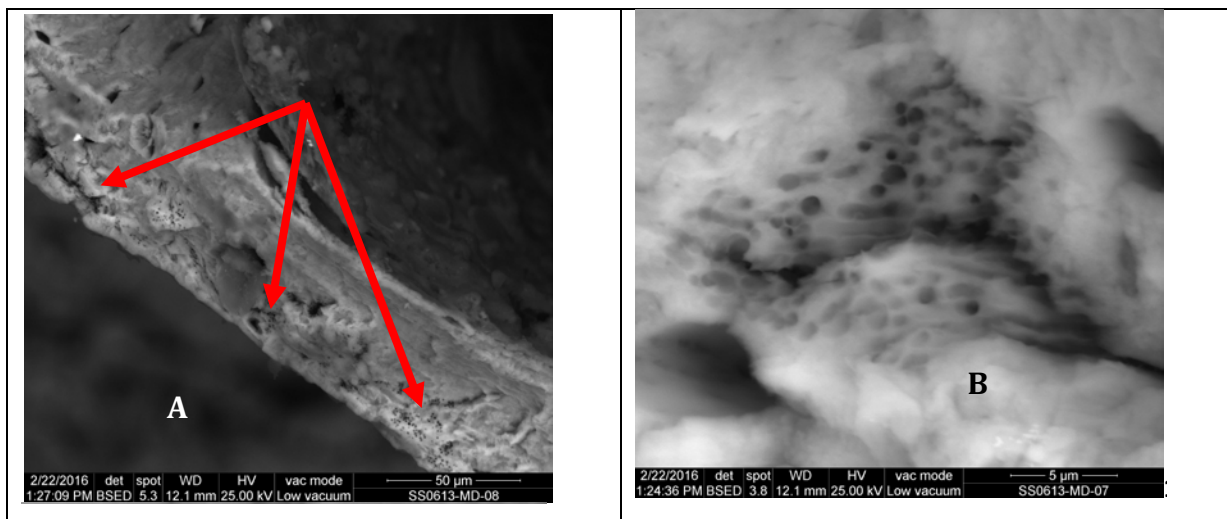


Figure 44 : mandibule analysée après six mois d'enfouissement. Les flèches rouges identifient des zones où les attaques bactériennes étaient visibles. Le cliché B, au grossissement x12.000, met en évidence des bactéries à la surface des zones érodées (clichés : Yolanda Fernandez-Jalvo).

Son OHI a été évalué à 3-4, c'est-à-dire que l'os était globalement bien préservé mais que quelques dégradations apparaissaient sur moins d'un tiers de l'os. Ces attaques ont été repérées par des flèches rouges sur la Figure 44. En augmentant le grossissement, il a été possible d'observer des points de petite taille dans les zones corrodées, pouvant être des bactéries.

Après 12 mois d'enfouissement, les modifications sont devenues plus nettes. Sur le premier cliché de la Figure 45, correspondant à un humérus, des racines sont observées, pénétrant à l'intérieur des os, également colonisés par des diatomées. Les articulations, souvent plus

fragiles car composées de cartilages, et les crêtes osseuses ont été également érodées (second cliché).

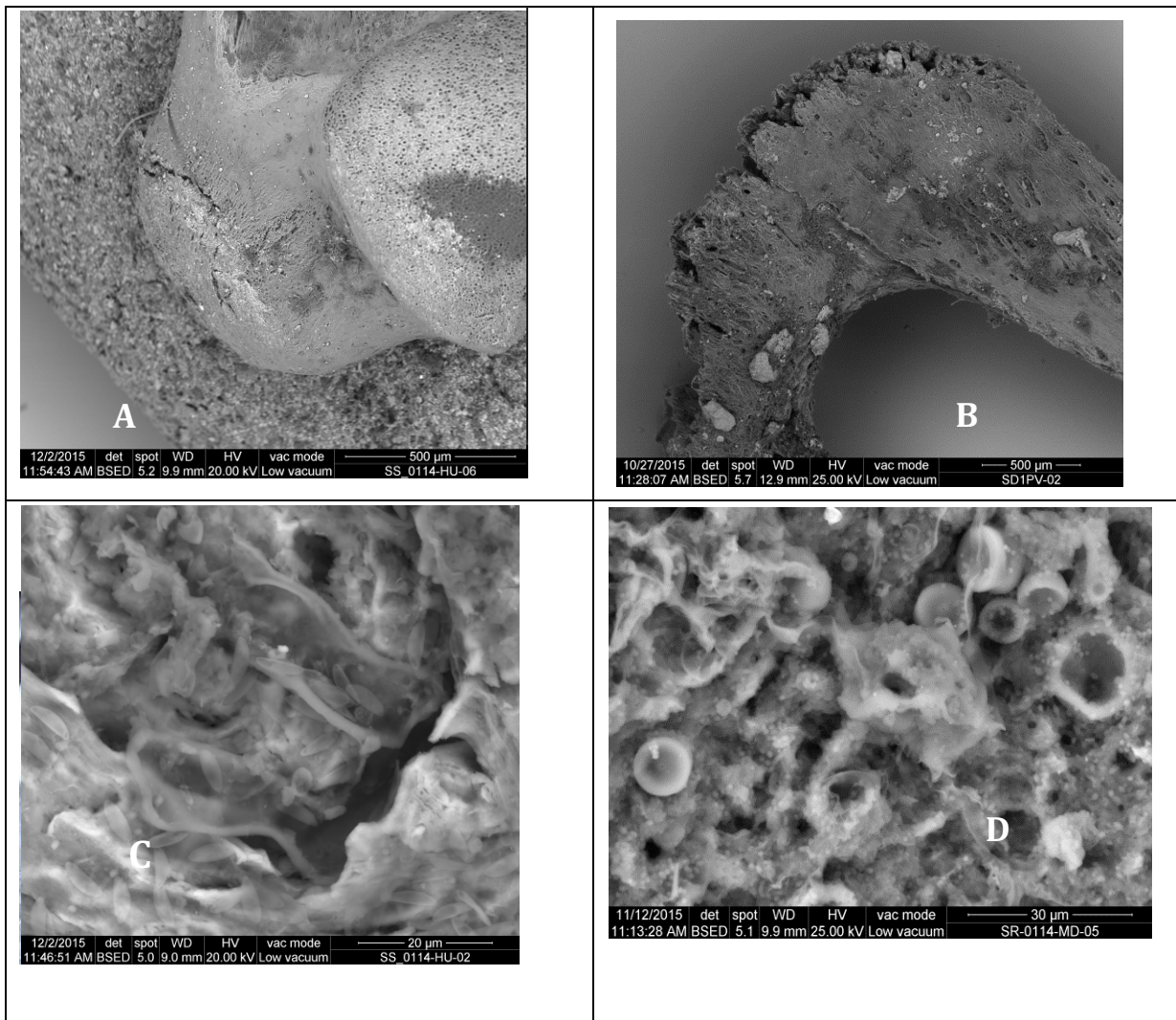


Figure 45 : ossements analysés après douze mois d'enfouissement. Le cliché A montre la corrosion des articulations et le cliché B l'érosion observée sur les crêtes osseuses. A un plus fort grossissement, on peut observer sur le cliché C la présence de racines à l'intérieur des os, os qui étaient couverts de bactéries et de spores fongiques (cliché D) (clichés : Yolanda Fernandez-Jalvo).

En passant à un plus fort grossissement, nous pouvons observer la présence de racines végétales dans les os (cliché C de la Figure 45), correspondant probablement à la fétuque rouge, plante identifiée par barcoding (Cf. 4.2). Enfin, au plus fort grossissement, il est possible d'observer la présence de bactéries et de spores fongiques.

Une intensification des signes de dégradation a été observée après 18 mois (Figure 46). Le cliché A montre une érosion importante du pelvis, avec disparition d'une partie de la structure osseuse. Analysée à un fort grossissement, il est observé une importante colonisation bactérienne et la présence de spores fongiques (cliché B).

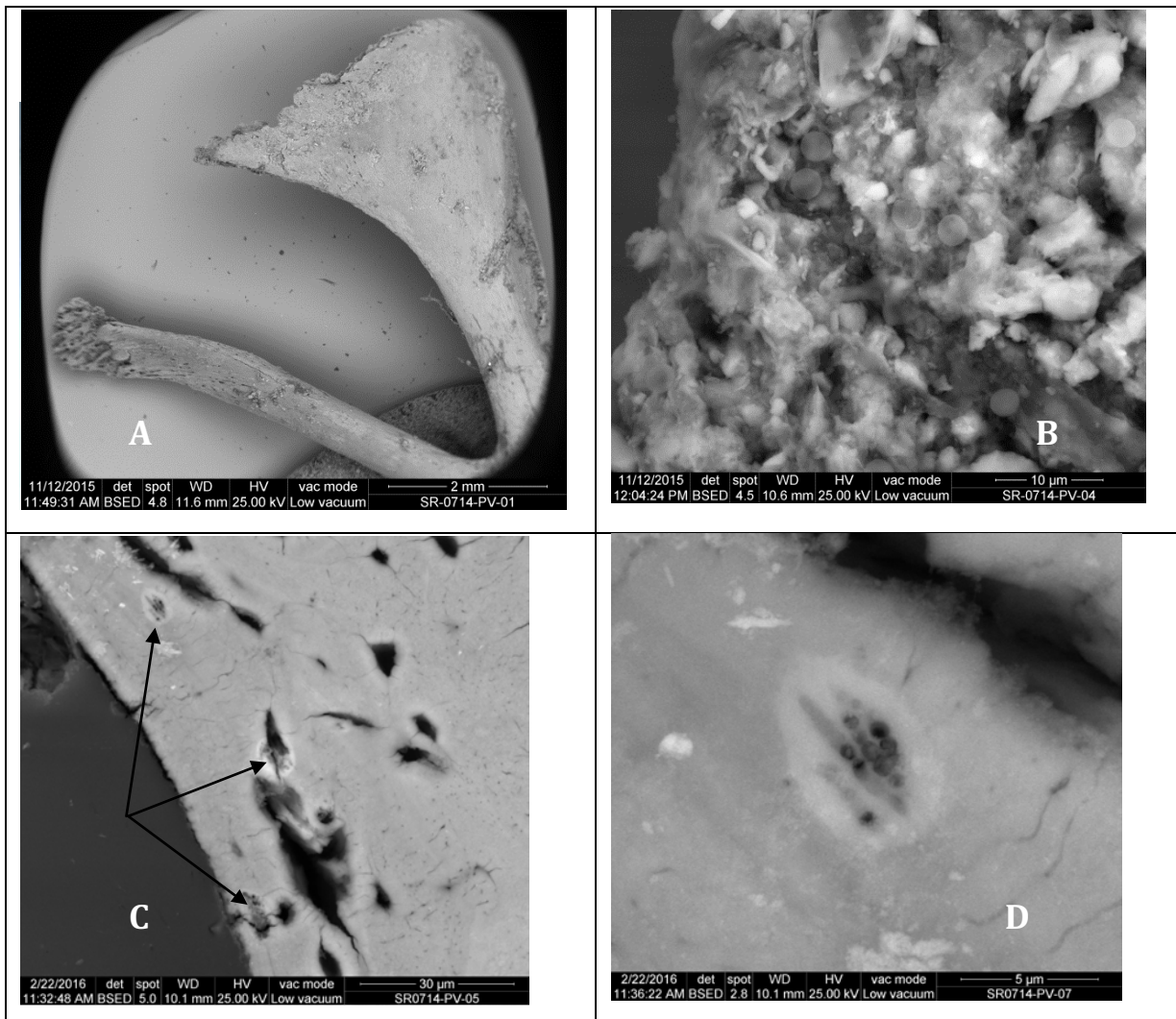


Figure 46 : ossements analysés après dix-huit mois d'enfouissement. Le cliché A montre la corrosion d'un os pelvien et le cliché B la prolifération de microbes à la surface de cet os. A un plus fort grossissement (flèches noires sur le cliché C), on peut observer la colonisation des canaux osseux par des bactéries, plus visibles sur le cliché D (agrandissement de la zone indiquée par l'une des flèches) (clichés : Yolanda Fernandez-Jalvo).

A plus fort grossissement, la colonisation bactérienne est retrouvée dans différents canaux osseux, qui sont autant de voies d'invasion pour les micro-organismes du sol. Cela est visible dans les clichés C et D, qui présentent des signes de présence bactérienne.

Après deux ans d'enfouissement, il est possible d'observer dans les canaux ayant servi à irriguer l'os des bactéries qui ont utilisé ces canaux pour envahir ce tissu. Ces bactéries sont visibles dans les deux clichés du haut dans la Figure 47.

Enfin, après 30 mois, les clichés présentent une forte similarité, montrant la rémanence de la présence bactérienne. Il n'est pas été possible dans les deux cas, à partir de ces clichés, d'identifier les bactéries présentes, mais il est confirmé que les os deviennent des matrices de support à la croissance de bactéries, quelles qu'elles soient, dans le sol.

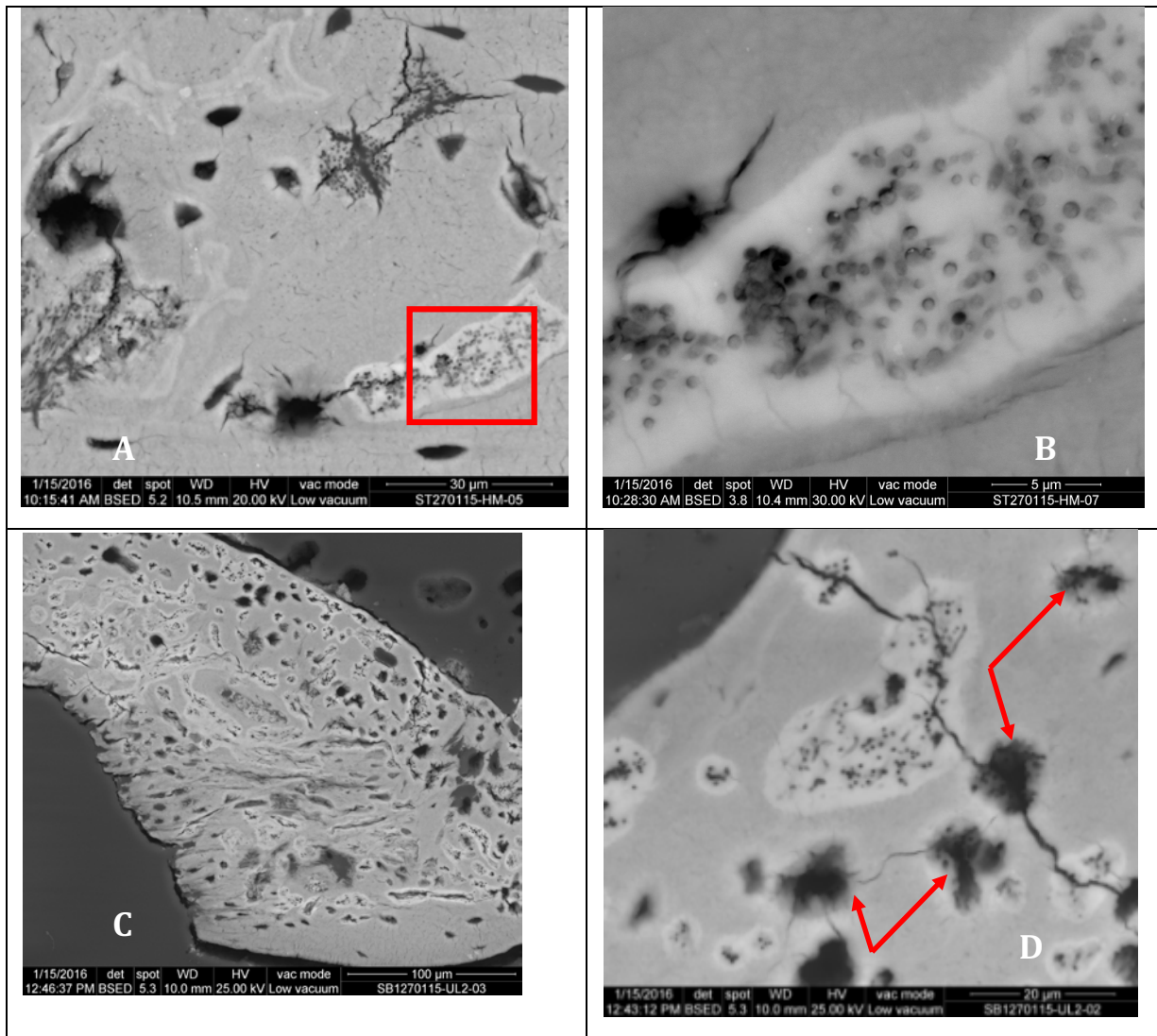


Figure 47 : clichés de l'invasion bactérienne après 24 mois (clichés A et B) et 30 mois (clichés C et D) d'enfouissement (clichés : Yolanda Fernandez-Jalvo).

Nous nous sommes intéressés à la quantification des bactéries présentes dans les ossements, à différents intervalles de temps, en amplifiant les fragments V5 et V345 dans les extraits d'ADN réalisés à partir des fémurs, contaminés ou non par les bactéries simulant une septicémie. Il est noté, comme attendu, une absence d'amplification des fragments V5 et V345 à J0. L'amplification reste faible à J+15 jours, corrélée avec l'aspect des dépouilles et la basse température du sol. Il est ensuite observé un pic de concentration de bactéries à J+3mois, tant en amplifiant le fragment V5 que le fragment V345, avant d'observer une décroissance de la présence de bactéries dans les os, que l'on peut corrélérer avec la baisse de bactéries, en particulier anaérobies, dans le sol environnant les souris. Après 18 mois, l'interprétation devient plus délicate, les quantités devenant élevées par moment, et non forcément corrélées entre fragments V3 et V345. Ces valeurs ont été représentées dans la Figure 48, qui indique les concentrations en ng/ μ L/g de fémur pour les souris enfouies.

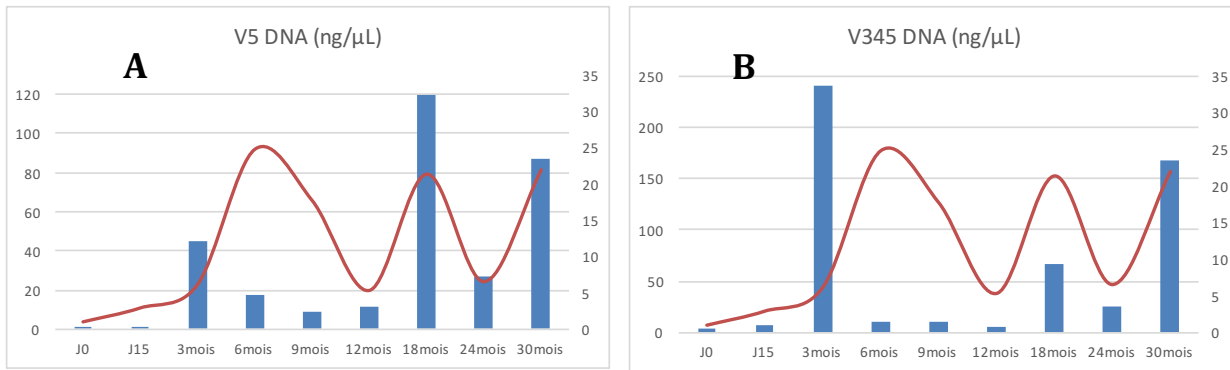


Figure 48 : évolution de la quantité d'ADN bactérien dans les os en ng/μL calculé par gramme d'os. La figure A correspond au fragment V5 et la figure B au fragment V345. L'axe secondaire indique les valeurs de température, rendues sous forme d'une courbe rouge.

Dans les deux cas, il devient possible de détecter de l'ADN bactérien de manière fiable après 3 mois. C'est le moment où l'on observe le pic de bactéries cultivables dans les sols environnant les souris enfouies (Figure 49). Cette quantité diminue ensuite mais recommence à augmenter, dans des proportions différentes, entre V5 et V345. Il n'a pas été trouvé d'explication évidente à ce phénomène, sauf éventuellement un effet de la saison, les pics à 18 et 30 mois correspondant aux mois d'été, visualisés sur la Figure 49 par les pics de température. La plus petite taille de l'amplicon V5 permet l'amplification d'ADN plus fragmenté, et donc éventuellement de bactéries en cours de décomposition ou déjà mortes. Cela pourrait expliquer la forte quantité d'amplicons observés lors des mois d'intense activité, la succession de bactéries actives étant mieux rendue en amplifiant un fragment court, qui sera en mesure de montrer les bactéries vivantes à cet instant, mais également les bactéries qui ont été présentes dans les semaines ou mois différents. Il permettrait en quelque sorte de remonter dans le temps, cette capacité à remonter dans le temps étant inversement proportionnelle à la longueur du fragment amplifié. Plus prosaïquement, il peut aussi y avoir une variabilité de la purification ou une différence d'efficacité des amorces avec les différents taxons qui se succèdent au cours du temps.

3.4. Détermination de la communauté bactérienne

Les sols témoin et au contact des bactéries ont été mis en suspension dans de l'eau stérile. La suspension a ensuite été étalée sur gélose TSA et mise en culture dans une étuve à température ambiante pour 72 heures afin de dénombrer la quantité de bactéries viables. Il a alors été possible de dénombrer les bactéries ayant poussé en aérobiose et anaérobiose. Le Tableau 10 présente les données collectées lors des six temps de prélèvements.

prélèvement	pH	dilution	numération aérobie				numération anaérobie				
			duplicat		quantif°	moyenne	duplicat		quantif°	moyenne	
sol témoin J0	7,16	-3	86	97	9,15E+05	1,51E+06	54	55	5,45E+05	4,65E+05	
		-4	9	13	1,10E+06		9	8	8,50E+05		
		-5	4	1	2,50E+06		0	0	0,00E+00		
sol témoin J+7	6,21	-3	ind	ind	2,55E+06	ND	ND	1,50E+05	1,50E+05		
		-4	19	23		2,10E+06	1			2	1,50E+05
		-5	2	4		3,00E+06	0			0	0,00E+00
récolte terre dans cage souris exhumée J+7	7,22	-3	155	171	1,63E+06	1,74E+06	ND	ND	1,25E+05		
		-4	25	17	2,10E+06		3	2		2,50E+05	
		-5	1	2	1,50E+06		0	0		0,00E+00	
sol témoin J+14	6,17	-3	ind	ind	2,85E+06	ND	ND	3,50E+05	3,50E+05		
		-4	26	18		2,20E+06	4			3	3,50E+05
		-5	4	3		3,50E+06	0			0	0,00E+00
récolte terre dans cage souris exhumée J+14	6,66	-3	ind	ind	5,78E+06	ND	ND	3,50E+05	3,50E+05		
		-4	55	66		6,05E+06	5			2	3,50E+05
		-5	7	4		5,50E+06	0			0	0,00E+00
sol témoin J+28	5,83	-3	ind	ind	6,93E+06	8	5	6,50E+04	5,75E+04		
		-4	56	51		5,35E+06	1	0		5,00E+04	
		-5	11	6		8,50E+06	0	0		0,00E+00	
récolte terre dans cage souris exhumée J+28	8,78	-3	ind	ind	3,84E+08	ind	ind	7,60E+07			
		-4	ind	ind		ind	ind		ind		
		-5	361	406		3,84E+08	69		83	7,60E+07	
sol témoin J+42	6,61	-3	ind	ind	3,90E+06	15	8	1,15E+05	1,58E+05		
		-4	32	24		2,80E+06	1	3		2,00E+05	
		-5	6	4		5,00E+06	0	0		0,00E+00	
récolte terre dans cage souris exhumée J+42	7,75	-4	ind	ind	7,00E+08	ind	ind	1,58E+08			
		-5	ind	ind		ind	87		113	1,00E+08	
		-6	74	66		7,00E+08	21		22	2,15E+08	
sol témoin J+56	6,33	-3	ind	ind	9,65E+06	50	40	4,50E+05	5,00E+05		
		-4	86	80		8,30E+06	2	8		5,00E+05	
		-5	12	10		1,10E+07	0	0		0,00E+00	
récolte terre dans cage souris exhumée J+56	7,99	-5	ind	ind	8,05E+08	54	50	5,20E+07	5,35E+07		
		-6	67	55		6,10E+08	7	4		5,50E+07	
		-7	9	11		1,00E+09	0	0		0,00E+00	

Tableau 10 : données de dénombrement sur les prélèvements issus du second protocole. ND : non déterminé. Ind : indénombrable, le nombre de colonies étant trop élevé pour être compté.

A partir de ces données, un graphe a été dessiné (Figure 49), permettant de visualiser l'évolution des populations bactériennes aérobies et anaérobies.

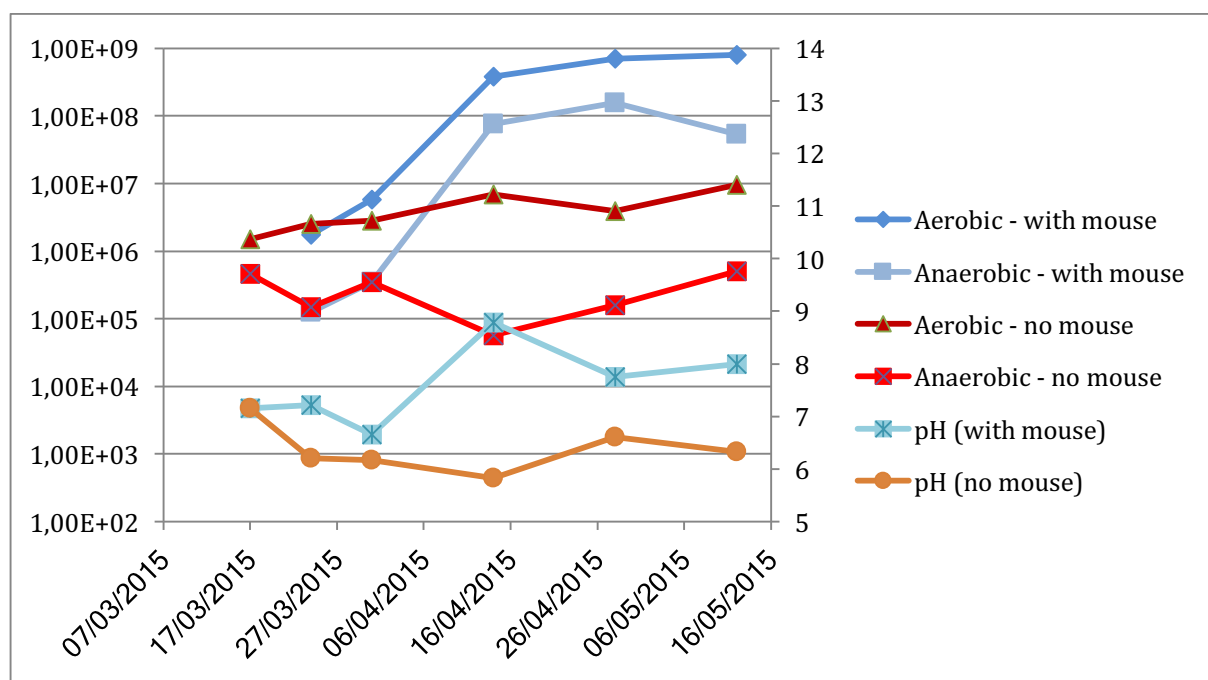


Figure 49 : évolution du pH et des populations bactériennes dans un sol témoin (courbes rouges) et à proximité immédiate de souris (courbes bleues). L'axe principal suit une échelle logarithmique.

En observant le sol témoin (courbes rouge), il est noté une croissance régulière du nombre de bactéries, essentiellement aérobies. Le pH évolue peu, restant entre 6 et 7, légèrement acide, ce qui correspond à la nature du sol, argilo-marneux. Cet accroissement du nombre de bactéries est régulier, et suit la courbe de température.

En observant la courbe des bactéries aérobies dénombrées dans le sol à proximité immédiate de la souris, il a été noté dès la deuxième semaine une augmentation de la population bactérienne, augmentation qui s'est accélérée pendant le premier mois puis qui s'est stabilisée avec 100 fois plus de bactéries que le sol témoin. L'accroissement des bactéries anaérobies a également été très sensible, dans les mêmes ordres de grandeur, mais avec un décalage d'une semaine par rapport aux bactéries aérobies. Le pH a augmenté pour se stabiliser entre 8 et 9, en accord avec ce qui est décrit dans la littérature [165].

3.5. Devenir de l'ADN murin

L'ADN et l'ARN ont été extraits d'échantillons de sol témoin et de sol au contact immédiat de souris enfouies, ainsi que les fémurs des souris enfouies après pesée et élimination des chairs adhérentes par un traitement à la trypsine. L'ADN global a été mesuré par Qubit. La Figure 50 représente l'évolution de la quantité d'ADN dans le sol au cours de

l'expérimentation. Cette quantité est restée stable dans le cas du sol témoin, entre 50 et 100 ng/μL d'extrait, tandis que la concentration en ADN a commencé à augmenter dans le sol au contact des souris à partir du 28^{ème} jour (en corrélation avec la croissance des populations bactériennes) et cette augmentation s'est poursuivie au cours du temps, pour atteindre 300 ng/μL d'extrait, soit près de 10 fois plus que le sol témoin.

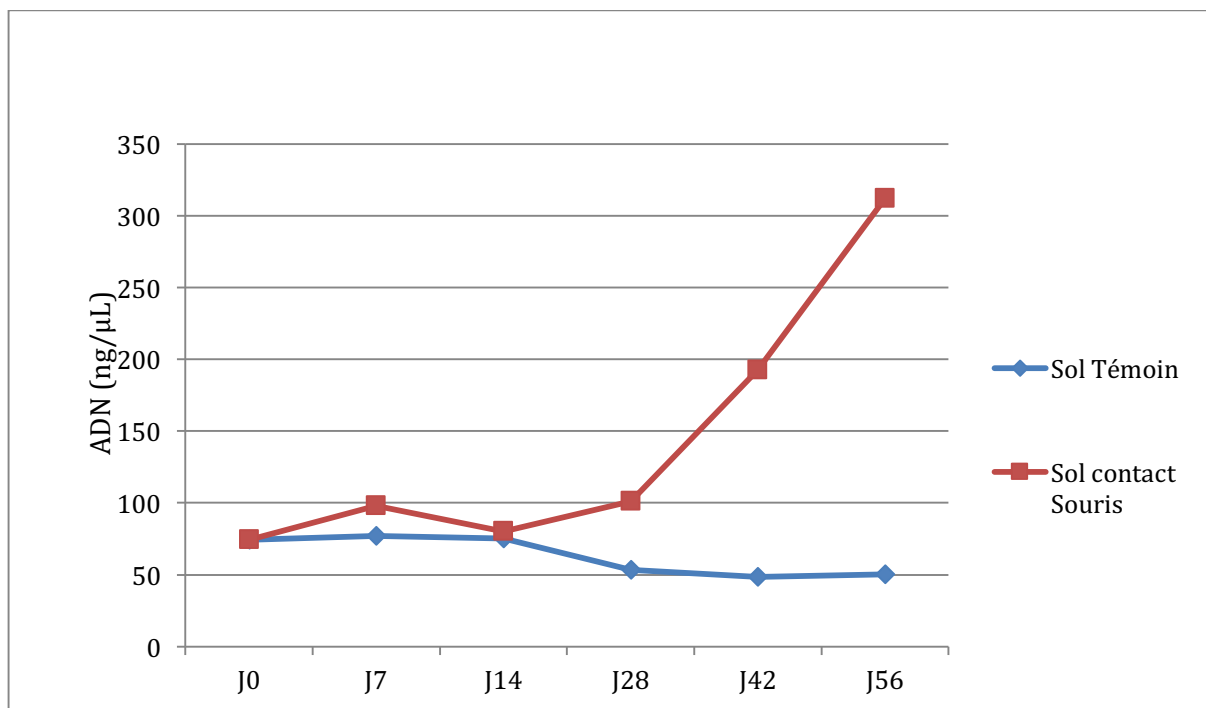


Figure 50 : évolution de la quantité brute d'ADN extrait à partir d'échantillons de sol au contact de la souris (courbe rouge) et de sol témoin (courbe bleue).

A partir des extraits d'ADN global, la concentration en ADN spécifiquement murin a pu être mesurée et sa dégradation a été évaluée en comparant les concentrations de quatre fragments de taille différente (73 - 125 - 154 - 513 bp).

Avec les échantillons les moins dégradés, il a été observé une sorte d'auto-amorçage de l'amplification, conduisant dès les premiers cycles à une augmentation linéaire de la quantité d'ADN, et nuisant à la mesure précise de la concentration en ADN (Figure 51). De plus, ce phénomène se produisait même en absence d'amorces murines (Figure 52).

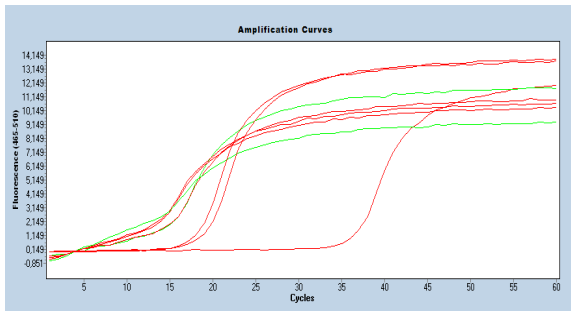


Figure 51 : courbes d'amplification d'échantillons murins avec le couple d'amorces Rd73.

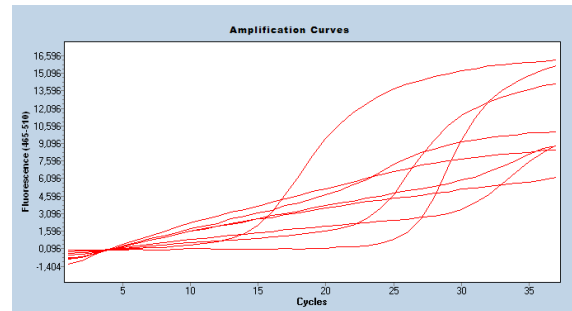


Figure 52 : courbes d'amplification d'échantillons murins avec le couple d'amorces V5

Afin de pallier ce phénomène, et au vu de son extrême précocité, tous les échantillons de souris ont été dilués au 1/100^{ème}. Ils ont ensuite été amplifiés par les trois couples d'amorces déjà décrits amplifiant les quatre fragments de taille croissante.

Lors de l'analyse des courbes de quantification, il a été possible d'observer, quel que soit le fragment amplifié, une augmentation initiale de la quantité de molécules d'ADN dans les os, puis une décroissance. Cette évolution est constatée sur les souris enterrées (Figure 53) et en surface (Figure 54), avec une cinétique comparable.

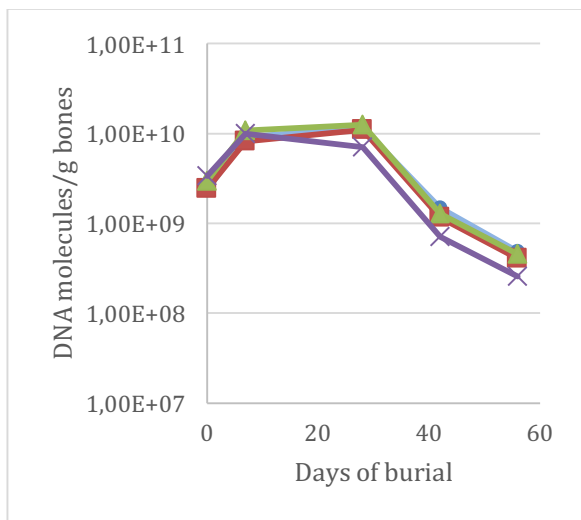


Figure 53 : Evolution du nombre de molécules d'ADN murin de taille variable en fonction du temps d'enfouissement. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb

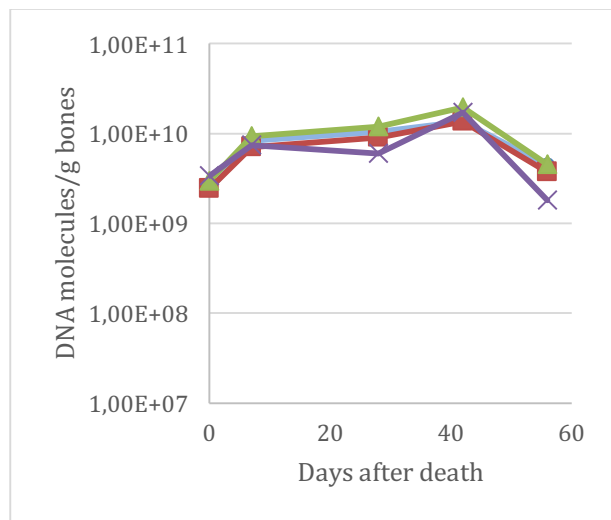


Figure 54 : Evolution du nombre de molécules d'ADN murin de taille variable en fonction du temps d'enfouissement. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb

Le fragment de plus grande taille, 513 bp, a semblé apparemment se dégrader plus vite que les autres fragments, plus courts.

Cette évolution a pu être confirmée en comparant le nombre de molécules d'ADN par gramme d'os, en fonction du temps et de la longueur des fragments (Figure 55). A J0 et à J7, pour le fémur enfoui, le nombre de molécules était comparable quelle que soit la longueur du fragment amplifié. Cela est montré par la pente de la courbe, mesurée entre 0,0002 et

0,0007, soit quasi-nulle. Par contre, pour J+28, J+42 et J+56, les courbes ont présenté une pente négative (entre -0,001 et -0,002), montrant l'appauvrissement relatif en molécules de grande taille.

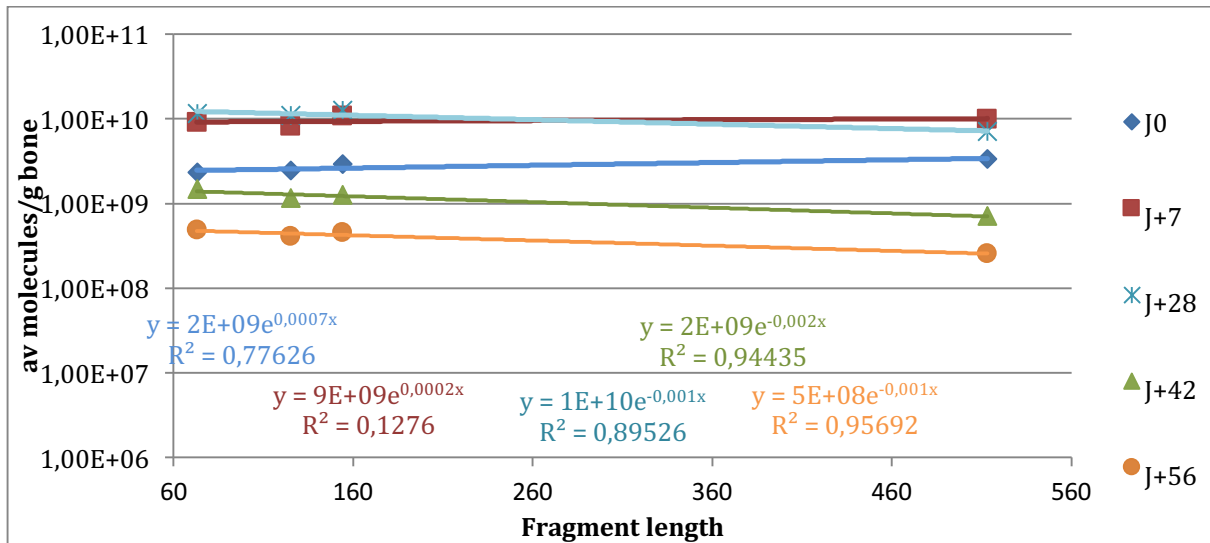


Figure 55 : évolution du nombre de molécules d'ADN présentes en fonction du temps, dans les fémurs enfouis. Bleu : J0, rouge J+7, Bleu clair : J+28, vert : J+42 et orange : J+56. Les équations des courbes ont été données dans la même couleur.

De même, pour les souris laissées en surface, il n'a pas été mesuré de variation du nombre de molécules selon leur taille entre J0 et J+28 (Figure 56). Par contre, il a été observé une décroissance pour les souris après 42 et 56 jours.

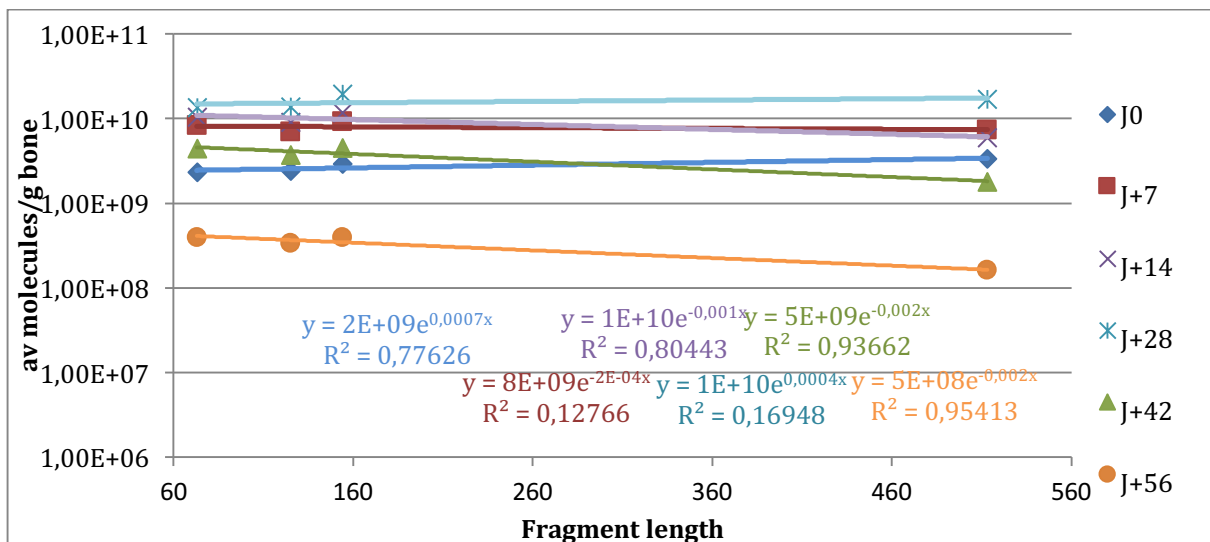


Figure 56 : évolution du nombre de molécules d'ADN présentes en fonction du temps, dans les fémurs des souris restées en surface. Bleu : J0, rouge J+7, Bleu clair : J+28, vert : J+42 et orange : J+56. Les équations des courbes ont été données dans la même couleur.

La corrélation entre les quantités de différents marqueurs était médiocre lors des premiers temps de mesure. Elle s'est nettement améliorée ensuite.

Parmi les échantillons enfouis, le fémur analysé à J+14 donnait des résultats aberrants, avec une quantité d'ADN inférieure d'un ordre de grandeur aux extraits précédents et suivants. Il a été décidé d'éliminer cette mesure de la comparaison, l'extraction ayant probablement été mal réalisée ou défectueuse dans un de ses réactifs.

Le sol au contact des souris a également été analysé, et comparé à un sol témoin. Il a été recherché dans le sol de l'ADN de souris par amplification de deux marqueurs murins, Rd73 et Rd125, tandis que deux fragments d'ADNr 16S ont été amplifiés dans le sol et dans les fémurs, le premier, court, encadrant le fragment hypervariable V5 (141bp), tandis que le second comprend les fragments V3, V4 et V5 (580bp).

Pour le sol, le contrôle d'inhibition a montré une inhibition présente dans tous les échantillons, plus ou moins intense. En conséquence, les échantillons de sol témoin à J+28, J+42 et J+56 ont été dilués à moitié, tandis que l'ensemble des autres échantillons de sol a été dilué au quart.

Dans le sol, il a été observé un pic de présence d'ADN murin au 28^{ème} jour (Figure 57). Ce pic est fugace, et il ne reste que des traces d'ADN murin deux semaines plus tard.

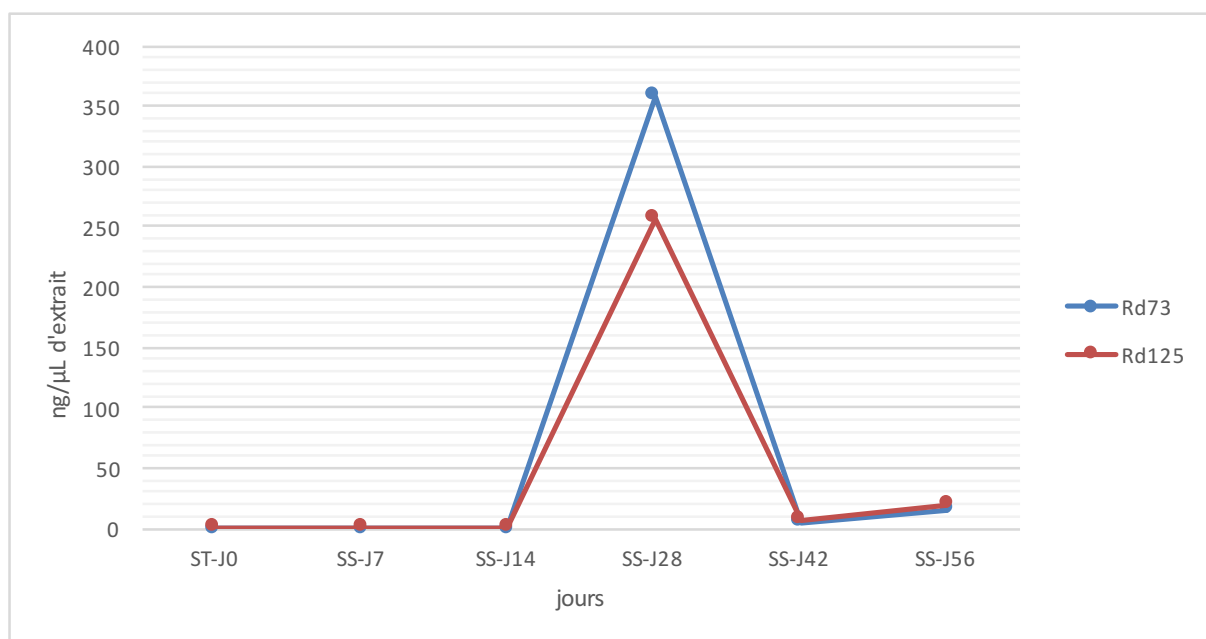


Figure 57 : évolution de la concentration d'ADN murin dans le sol au contact des cadavres de souris au cours du temps. La quantité d'ADN est mesurée par deux marqueurs, Rd73 (courbe bleue) amplifiant un fragment de 73 bp, et Rd125 (courbe rouge) amplifiant un fragment de 125 bp.

Ce pic est synchronisé avec l'élévation de concentration de l'ADN total dans le sol, la multiplication de la communauté bactérienne aérobie et anaérobie. Il est possible que ces trois paramètres soient corrélés, avec une rupture de l'enveloppe corporelle, entraînant une libération importante de fluides riches en composés nutritifs dans le sol, d'où une

multiplication des populations saprophytes du sol, et donc une élévation de la quantité d'ADN total. Ces fluides corporels seraient riches en cellules murines, et donc cela expliquerait également l'élévation de la concentration d'ADN murin dans le sol, élévation transitoire car ces cellules et ces fluides seraient rapidement digérées par les organismes saprophytes présents dans le sol.

3.6. Analyse des Acides nucléiques par séquençage massif

Les extraits de fémur, de sol témoin et de sol au contact des souris ont été amplifiés par deux couples d'amorces, l'un amplifiant le fragment V5, l'autre un fragment comprenant V3, V4 et V5.

A l'issue de l'amplification par PCR, des banques ont été préparées selon le protocole décrit dans [172], en normalisant les quantités initiales de produits d'amplification. Le séquençage est réalisé sur appareil Illumina MiSeq, en multiplexant jusqu'à 94 échantillons avec un kit V3 600pb (ref. MS-102-3003). Les résultats de séquençage ont été traités selon le script « usearch2taxo ». Les résultats d'amplification ont été rassemblés dans le Tableau 11. Le nombre de séquences obtenues par échantillon a été très variable, avec néanmoins une tendance à un plus grand nombre de séquences avec les extraits issus des amplicons V345.

Les séquences obtenues avec l'extrait FT-170315 étaient très peu nombreuses, et le taux de GC très différent des autres. En théorie, il ne devrait pas être amplifié d'ADN bactérien sur l'extrait FT, qui est le fémur témoin, et donc non exposé à la décomposition. Néanmoins, l'autopsie et les manipulations ultérieures ont été susceptibles d'avoir apporté des traces d'ADN bactérien, et l'amplification a pu également entraîner la synthèse de séquences non spécifiques. Il est donc probable que les assignations taxonomiques issues de ces deux échantillons ne soient pas un reflet de la réalité mais un artefact. Les résultats n'ont donc pas été utilisés pour les comparaisons. Dans un premier temps, il a été observé par analyse en composantes principales (PCA) si les banques issues des deux amplicons présentaient un profil comparable. Le résultat est présenté en Figure 58. Les deux populations étaient dissemblables, les extraits amplifiés par V5 étant plus proches entre eux que ceux amplifiés par V345. Chaque couple d'amorces amplifie préférentiellement des populations bactériennes différentes, du fait de la différence de taille des séquences amplifiées. Il est plausible que le couple V5 puisse identifier des ADN bactériens dégradés, mais avec une précision taxonomique moindre. Les analyses ultérieures seront faites séparément sur V5 ou V345.

échantillon	séquences obtenues	séquences conservées après nettoyage	Séquences alignées	longueur moyenne	%G C	% alignés	nombre d'OTUs (niveau du genre, après exclusion des singletons)
F-BalbC	31284	30857	27389	139,349	50	87,5%	1036
F-Bg	32889	32567	29559	139,491	51	89,9%	1178
F-Xa	64640	63302	57894	139,22	51	89,6%	1642
F-cycloP	49532	49116	45405	139,459	51	91,7%	1912
FSE-120515-V345	653369	625976	533587	144,2	53	81,7%	20067
FSE-120515-V5	253098	171796	134038	167,4	53	53,0%	8672
FSE-140415-V345	394092	376582	314596	143,5	52	79,8%	11861
FSE-140415-V5	187785	120310	90546	187,2	53	48,2%	6260
FSE-240315-V345	441277	425512	377166	141,7	50	85,5%	10005
FSE-240315-V5	132610	99403	80399	155,4	51	60,6%	4903
FSE-280415-V345	452677	430651	353900	145,2	53	78,2%	14684
FSE-280415-V5	245176	141202	107241	169,9	53	43,7%	7449
FSE-310315-V345	395376	379362	285759	142,1	50	72,3%	8157
FSE-310315-V5	124148	88188	88164	153,1	51	71,0%	4078
FSS-120515-V345	527166	507215	423032	143,1	52	80,2%	18207
FSS-120515-V5	193180	130599	99969	154,1	53	51,7%	7470
FSS-140415-V345	618457	590860	491037	143,8	53	79,4%	18101
FSS-140415-V5	258679	165949	130353	154,1	53	50,4%	7949
FSS-240315-V345	796689	766501	684369	142,1	51	85,9%	15325
FSS-240315-V5	279712	171574	138333	151,0	52	49,5%	7113
FSS-280415-V345	595748	568384	476057	145,4	53	79,9%	16803
FSS-280415-V5	295653	212690	167410	153,5	53	56,6%	10209
FSS-310315-V345	357403	350199	343553	144,3	51	96,1%	12868
FSS-310315-V5	381744	250608	70169	185,9	51	18,4%	11251
FT-170315-V345	59761	59017	57407	151,3	35	96,1%	761
FT-170315-V5	2950	2269	1906	155,7	39	64,6%	93

Tableau 11 : Résultats du séquençage des banques créées à partir des extraits osseux et des fèces. Le nombre d'OTUs est donné après élimination des séquences présentes en une seule copie (singleton), au niveau taxonomique du genre (niveau le plus bas obtenu avec usearch). FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; FT : Fémur Témoin. F-xxx : banques obtenues à partir de fèces.

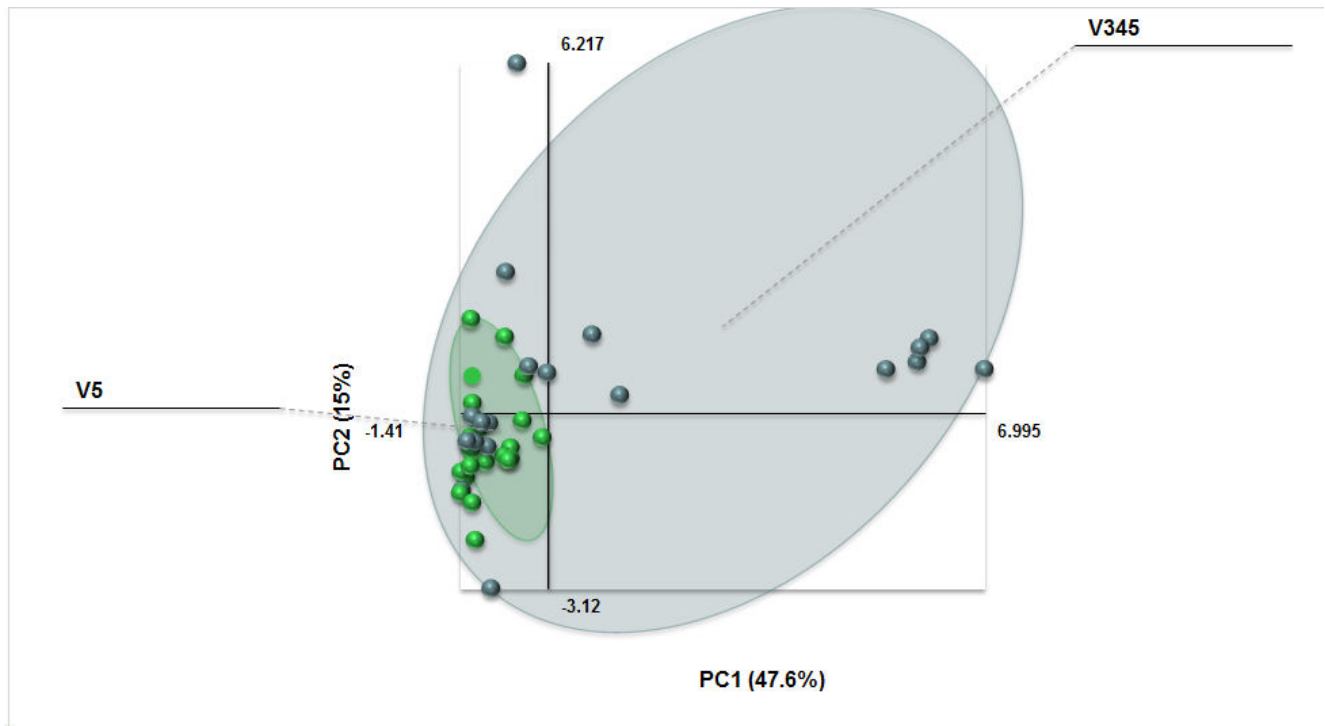


Figure 58 : Représentation en PCA des deux populations amplifiées, par V5 (points verts) et par V345 (points gris).

3.6.1. Analyse du fragment HVR-5

La première analyse a concerné les échantillons amplifiés par le fragment V5. A partir des assignations taxonomiques, il a été possible dans un premier temps de représenter les proportions des différents phyla et leur évolution au cours du temps. La Figure 59 donne une représentation des résultats obtenus, au niveau taxonomique du phylum.

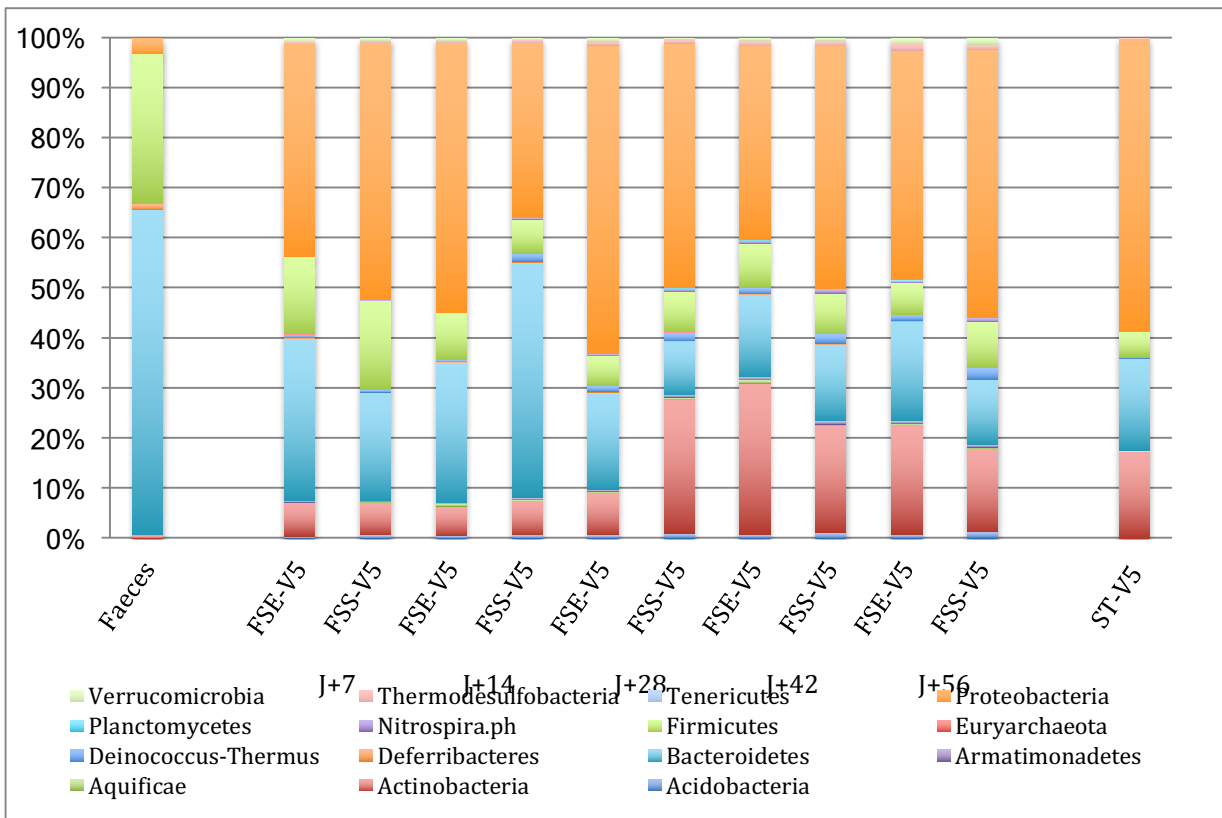


Figure 59 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau du phylum. Les extraits de fémur ont été regroupés deux à deux en fonction du délai de décomposition. Un échantillon de sol témoin est représenté à droite et de fèces à gauche. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin

Les fèces ont donné des profils très différents des échantillons osseux, avec une prédominance de Bacteroidetes, ce qui est attendu, ce phylum étant prépondérant dans le microbiote intestinal. Le second phylum le plus représenté était constitué de Firmicutes, connus pour être également dominants dans le microbiote intestinal.

En observant les populations bactériennes dans les os au cours du temps (Figure 59), une évolution a pu également être observée. Dans le cas des fémurs issus de souris enfouies (FSE), la population majoritaire était initialement composée de Protéobactéries, de Bacteroidetes et de Firmicutes par ordre de prévalence. Les Bacteroidetes sont devenus majoritaires à J+14, avant que les protéobactéries ne soient redevenues majoritaires à partir de J+28 tandis que les Actinobactéries ont commencé à croître. A partir de ce moment, le profil a tendu à se stabiliser et à ressembler à un profil tellurique, comme représenté dans l'échantillon ST-V5. Néanmoins, plus de phyla étaient présents que dans le sol (Chlamydiae, Chlorobi, Gemmatimonadetes et Planctomycetes). Dans le cas des souris laissées en surface, l'évolution a été un peu différente, avec une proportion plus importante de Bacteroidetes dès 14 jours, qui ont diminué des 3/4 entre J+14 et J+28, remplacées par les Actinobactéries

et les Protéobactéries. La croissance des Actinobactéries a été plus tardive, avec un pic à J+42, avant que les Protéobactéries ne soient redevenues majoritaires entre J+42 et J+56. A cette date, le profil est devenu proche de celui des souris enfouies, et de celui du sol, à la différence d'une présence notable du phylum *Deinococcus-thermus*.

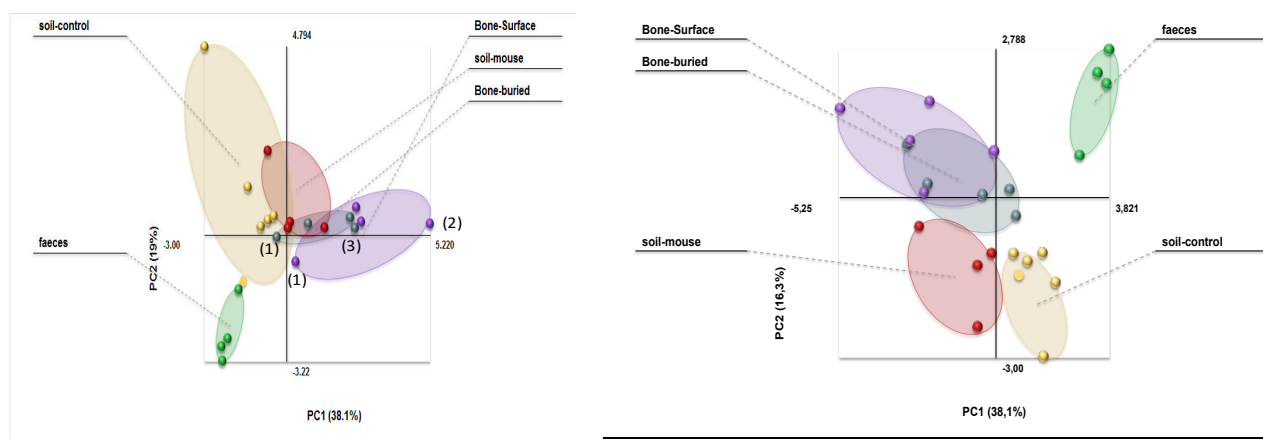


Figure 60 : séparation par PCA (à gauche) et par PLS-DA (à droite) des différentes populations analysées, au niveau du phylum, à partir des amplicons HVR5. En jaune le sol contrôle, en rouge le sol au contact des souris, en vert les fèces, en gris-bleu les os enfouies et en violet les os en surface. (1) : J7 ; (2) : J42 ; (3) : J42-J56.

Lorsqu'est réalisée une analyse en composantes principales (ACP ou PCA) sur ces échantillons, toutes les catégories, hors fèces, se sont partiellement recouvertes (Figure 60) et il n'apparaît pas de répartition liée au temps de dégradation entre les différents points. A titre d'exemple, pour les os de surface (orbe violette), les deux points les plus distants correspondaient à J+7 et J+42, J+14, J+28 et J+56 étant colocalisés. Pour les os enterrés (orbe grise), les deux extrémités de l'orbe correspondaient à J+7 d'une part, et J+42/J+56 d'autre part. En réalisant une analyse discriminante par régression des moindres carrés partiels (PLS-DA)², les groupes ont été mieux séparés, seuls les os enterrés et en surface présentant un recouvrement partiel. Les échantillons de sols contrôle ont été mieux regroupés, les fèces également. Par contre, il n'y avait pas de corrélation entre position dans l'orbe et temps de décomposition, quel que soit le type d'échantillon considéré. La PLS-DA a semblé être l'analyse la mieux adaptée à nos échantillons, mais il est possible que l'amplicon HVR5

² Cette méthode d'analyse est utilisée afin de mettre en évidence des relations entre plusieurs groupes, caractérisés par au moins une valeur qualitative (catégorie - ici le genre ou l'espèce bactérienne) dans le cas de l'analyse discriminante. Elle est utilisée pour la clusterisation supervisée, c'est-à-dire que les échantillons sont répartis en groupes connus et que la méthode s'attache à déterminer si les groupes sont réellement différents, et dans l'affirmative quels sont les paramètres permettant cette différenciation.

n'apporte pas assez de résolution. Les amplicons issus du fragment couvrant les HVR 3, 4 et 5 ont donc été à leur tour analysés.

3.6.2. Analyse du fragment HVR-345

La Figure 61 donne une représentation de l'évolution des populations bactériennes dans les os de souris au cours du temps, mises en évidence par l'amplification du fragment HVR 345.

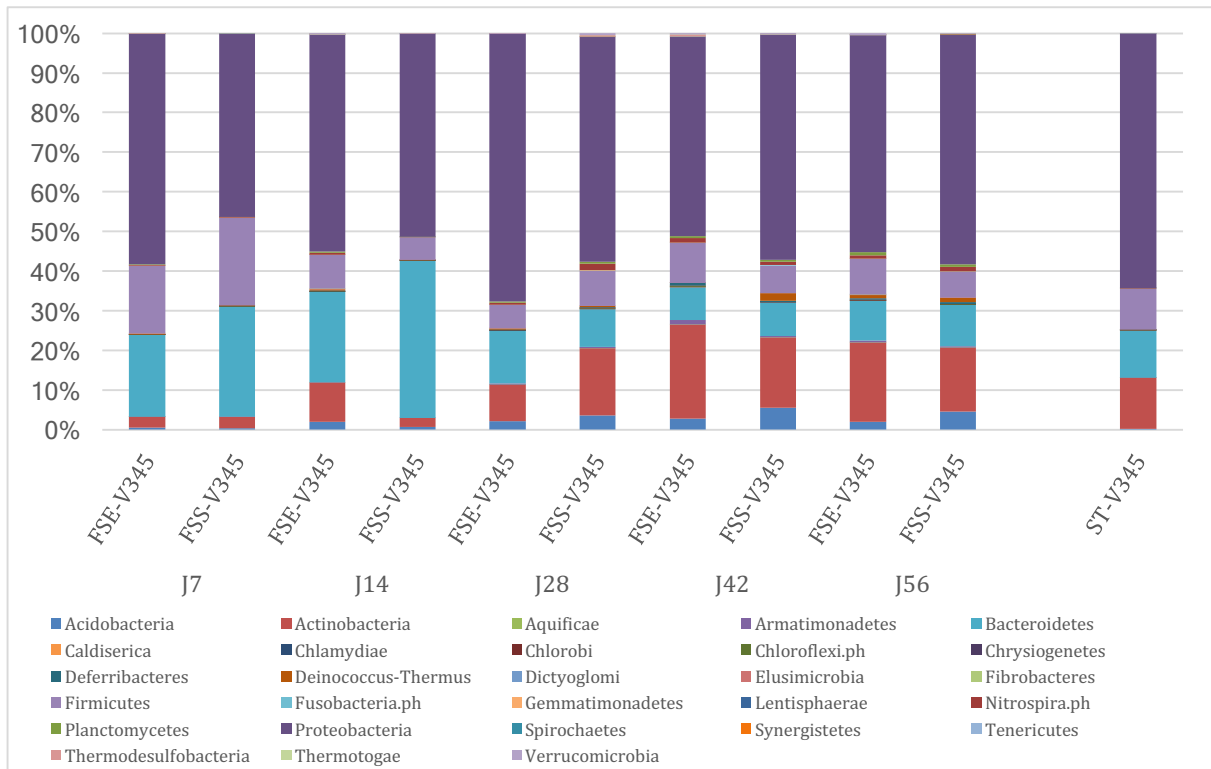


Figure 61 : représentation de la diversité bactérienne déduite des amplicons V345, au niveau du phylum. Les extraits de fémur ont été regroupés deux à deux en fonction du délai de décomposition. Un échantillon de sol témoin est représenté à droite. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin

Comme dans le cas des amplicons V5, les fèces étaient essentiellement composées de Bacteroidetes et de Firmicutes. Par contre, au cours du temps, les populations bactériennes observées dans les fémurs de souris enfouies ou laissées en surface ont moins différencié que ce qui avait été observé avec les amplicons V5. Cette observation s'est retrouvée dans les PCA et surtout en PLS-DA (Figure 62).

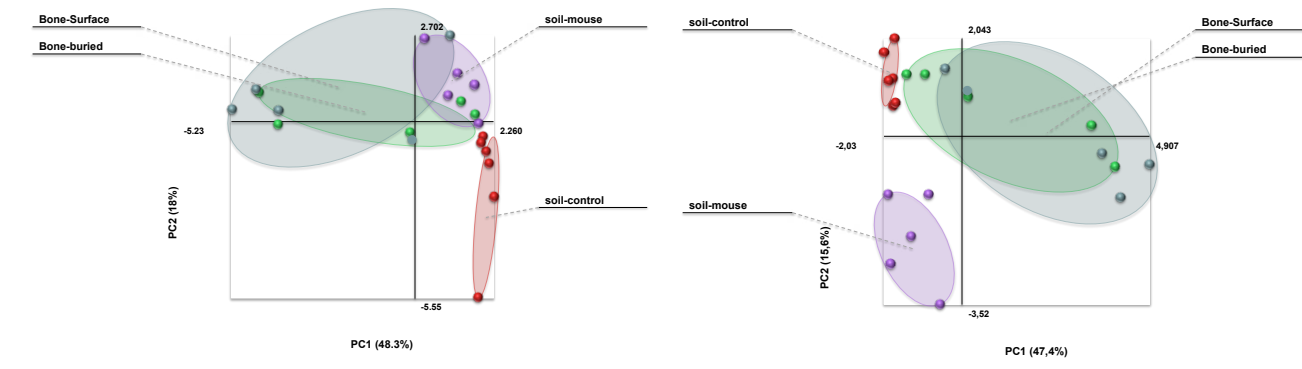


Figure 62 : séparation par PCA (à gauche) et par PLS-DA (à droite) des différentes populations analysées, au niveau du phylum, à partir des HVR 3, 4 et 5. En rouge le sol contrôle, en violet le sol au contact des souris, en vert les os enfouis et en gris les os en surface.

En effet, pour cette dernière, il a été observé un très bon recouvrement des profils issus des os, qu'ils soient en surface ou enterrés. Les sols contrôles ont bien été regroupés et les sols au contact des souris évoluent au fur et à mesure du temps de décomposition. Cependant, cette évolution n'a suivi aucun des axes identifiés par la PCA ou la PLS-DA. Les méthodes statistiques retenues ne semblaient pas permettre de modéliser les évolutions temporelles des populations observées. D'autres méthodes, comme la PCoA, sont mises en œuvre au laboratoire pour analyser les changements de populations et mettre en évidence leurs éventuels patrons d'évolution.

En descendant d'un niveau taxonomique, les profils se sont complexifiés, mais il est apparu qu'à la fin de l'expérimentation, les classes bactériennes présentes étaient comparables entre échantillons osseux et sol environnant, quel que soit le fragment hypervariable considéré (Figure 63 et Figure 64). Cette évolution a été nette à partir de J+28. Au niveau des fèces, les classes présentes étaient, pour les Bacteroidetes, *Bacteroidia*, et pour les Firmicutes les Clostridiales. Ces deux classes sont essentiellement anaérobies, et leur présence dans les fèces est bien décrite dans la littérature.

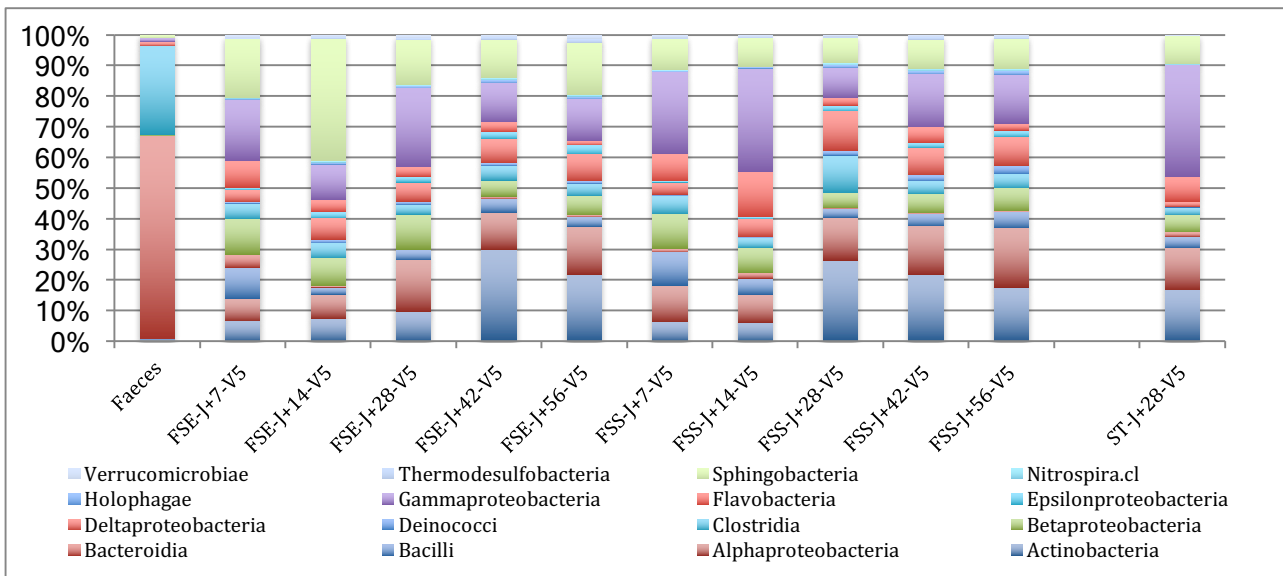


Figure 63 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification du fragment hypervariable V5, au sein des populations en surface et enterrées. Un échantillon de sol témoin est représenté à droite et de fèces à gauche. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin

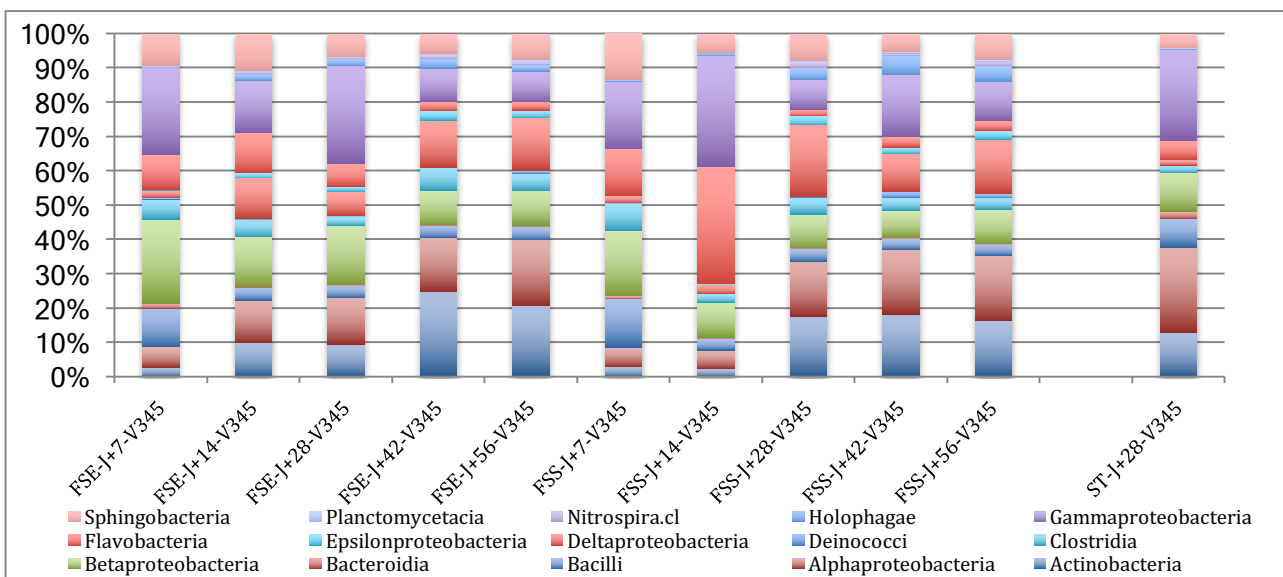


Figure 64 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification d'un fragment englobant les régions hypervariables 3, 4 et 5, au sein des populations en surface et enterrées. Seules les 15 classes majoritaires ont été représentées. Un échantillon de sol témoin est représenté à droite. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin

Les proportions des classes mises en évidence variaient selon l'amplicon considéré, avec une moindre diversité dans le cas de V345, qui pourrait être attribué à une meilleure assignation taxonomique du fait de la plus grande taille du fragment analysé.

Les mêmes amplifications ont été réalisées sur des échantillons de sol, témoin et au contact des souris (Figure 65).

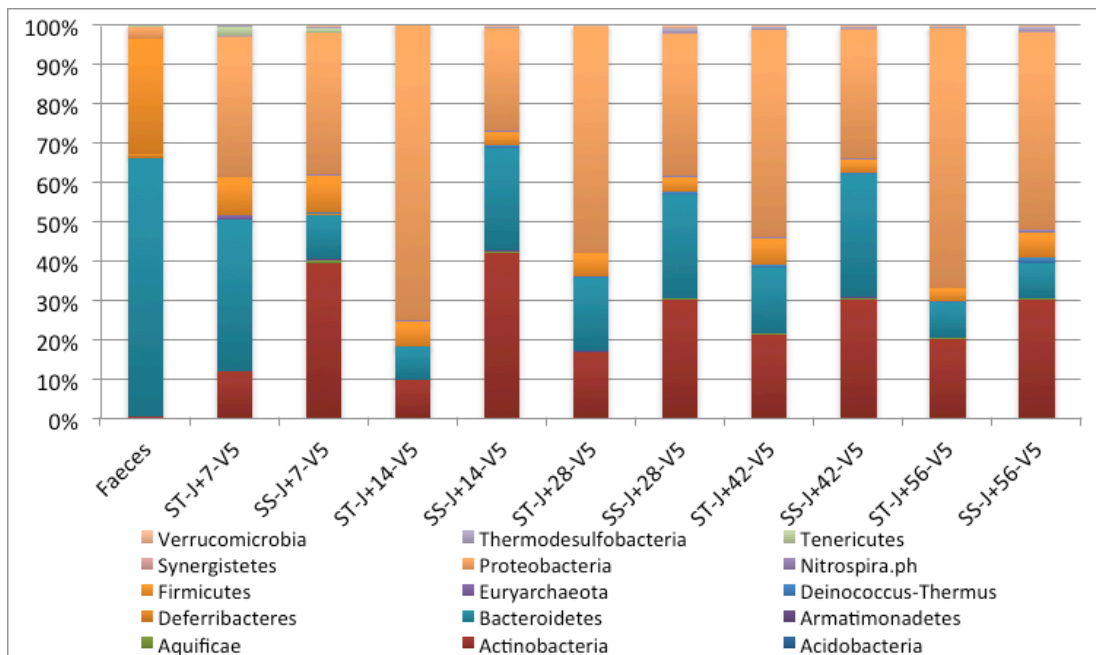


Figure 65 : évolution des populations bactériennes au cours du temps post-mortem, inférées par l'amplification du fragment hypervariable 5, au sein des sols témoin et contact. Seules les 15 classes majoritaires ont été représentées. Un échantillon de faeces témoin est représenté à droite. ST : Sol Témoin ; SS : Sol Souris.

Les histogrammes alternent sols témoin et sols contact, afin de comparer à la même date les profils bactériens. La composition globale, au niveau du phylum, a assez peu varié pour le sol témoin, avec les Protéobactéries globalement en majorité, puis Actinobactéries ou Bacteroidetes et Firmicutes. Le sol témoin J+7 a présenté un profil particulier, avec un excès de Bacteroidetes. Les Bacteroidetes ont dépassé nettement en proportion les Actinobactéries en une seule occasion sur sept (J+6 mois). La différence avec le sol est apparue dès J+7, avec une augmentation des Actinobactéries et des Bacteroidetes, tandis que les Firmicutes et les Protéobactéries tendaient à diminuer. Après 56 jours, les profils ont commencé à se rapprocher, mais avec toujours un excès d'Actinobactéries et une sur-représentation des Firmicutes dans le sol au contact des cadavres. Un échange s'est donc mis en place entre les os et le sol environnant, avec colonisation des os par les bactéries telluriques, certaines d'entre elles étant stimulées par l'apport de nutriments représenté par le cadavre.

Lorsque la même analyse a été réalisée sur les amplicons V345, les résultats ont été très proches (Figure 68).

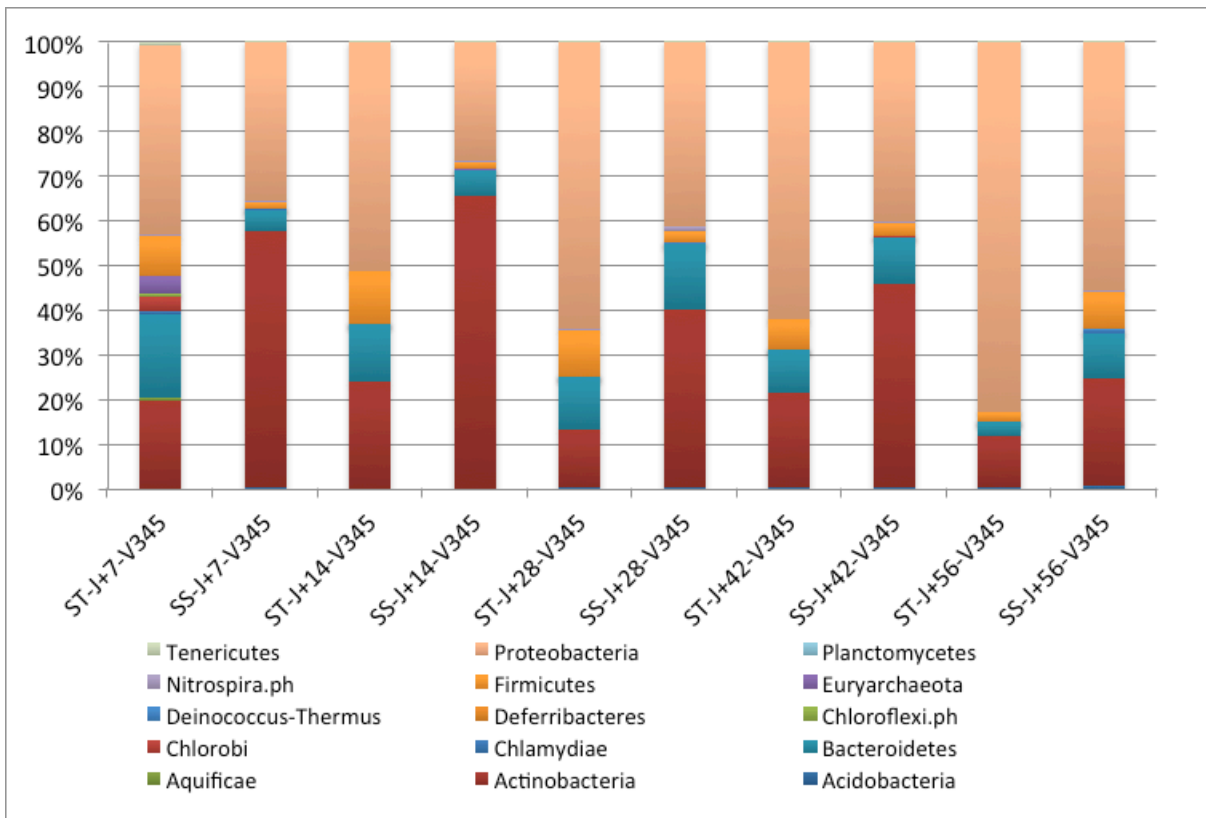


Figure 66 : évolution des populations bactériennes au cours du temps post-mortem, inférées par l'amplification d'un fragment englobant les régions hypervariables 3, 4 et 5, au sein des sols témoin et contact. Seules les 15 classes majoritaires ont été représentées. ST : Sol Témoin ; SS : Sol Souris.

Le sol témoin à J+7 a présenté également un profil particulier. Par contre, les Firmicutes étaient plus représentés dans le sol témoin que dans le sol contact, à l'exception du dernier point de prélèvement, tandis que les Bacteroidetes étaient proportionnellement plus représentés dans le sol témoin qu'avec le fragment V5. Les Actinobactéries ont très vite augmenté dans le sol contact, avec un pic en proportion à J+14, avant que les Bacteroidetes n'aient augmenté à leur tour.

3.6.3. Etude de l'activité métabolique bactérienne :

L'amplification des fragments HVR 5 et HVR 345 dans les extraits d'ADN a permis d'identifier les bactéries présentes dans les échantillons, qu'elles soient en pleine activité, en dormance voire mortes. L'amplification de l'ARN montre les bactéries métaboliquement actives, et donc peut permettre de mieux comprendre les phénomènes cataboliques se déroulant au niveau des cadavres. Une RT-q-PCR a donc été réalisée sur les extraits d'ARN réalisés en parallèle des extraits d'ADN.

La Figure 67 montre l'évolution des populations bactériennes au sein du sol témoin au cours de l'expérimentation, à partir des amplicons V5, sur les extraits d'ADN et d'ARN.

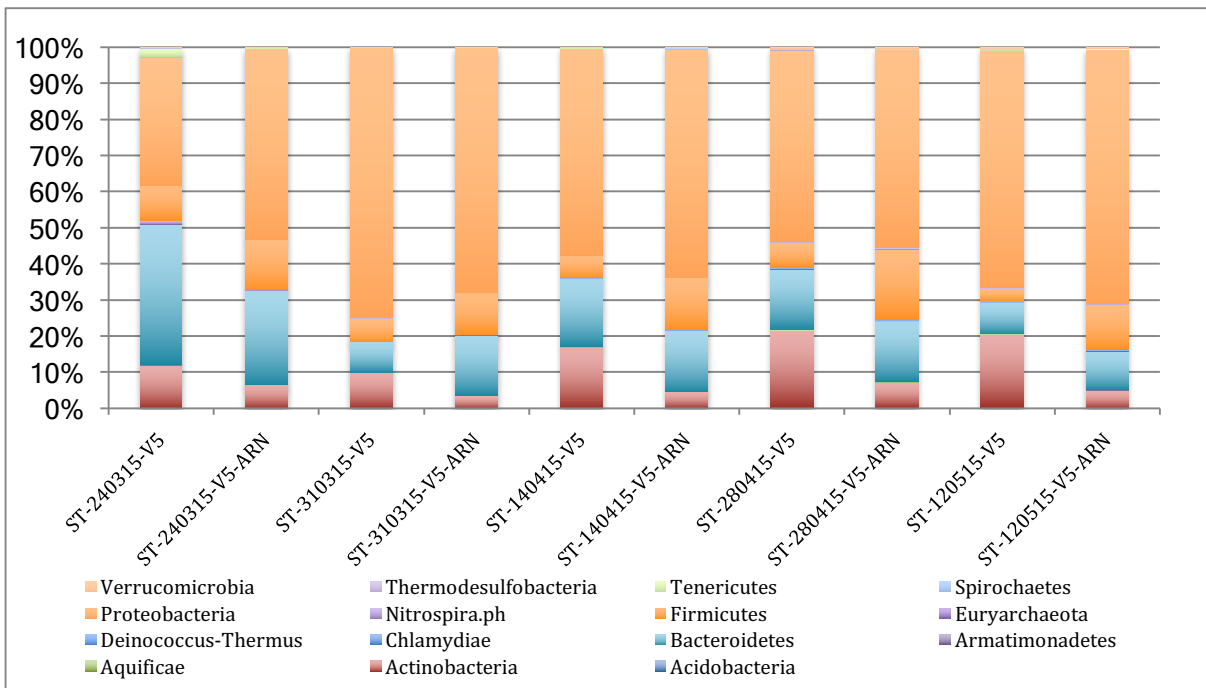


Figure 67 : évolution des populations bactériennes dans le sol témoin, à partir des amplicons du fragment HVR5. Les 15 phyla majoritaires ont été représentés. ST : Sol Témoin ; SS : Sol Souris.

Il est montré des différences entre populations présentes (mesurées par l'amplification ADN de HVR5) et populations actives (mesurées par l'amplification par RT-PCR de HVR-5 de l'extrait d'ARN du sol). En permanence, il a été détecté plus d'Actinobactéries par ADN que par ARN. *A contrario*, il a été observé plus de Firmicutes par ARN que par ADN. Cela tendrait à dire que les Firmicutes étaient proportionnellement plus actifs que les autres phyla détectés, et les Actinobactéries plus présents qu'actifs. La Figure 68 présente les profils obtenus à partir des sols au contact des souris. Comme pour le sol témoin, les Actinobactéries étaient plus présentes qu'actives. C'est également le cas pour les Bacteroidetes. Les Firmicutes étaient actives, ainsi que les Protéobactéries. Ce phylum bactérien semblait être le seul à évoluer de manière différente entre sol témoin et sol au contact de souris enfouies. Nous avons donc exploré la diversité au sein de ce phylum. La Figure 69 donne une représentation de la distribution des différents ordres dans le phylum des Protéobactéries pour le sol témoin, tandis que la Figure 70 donne la représentation pour les sols contact.

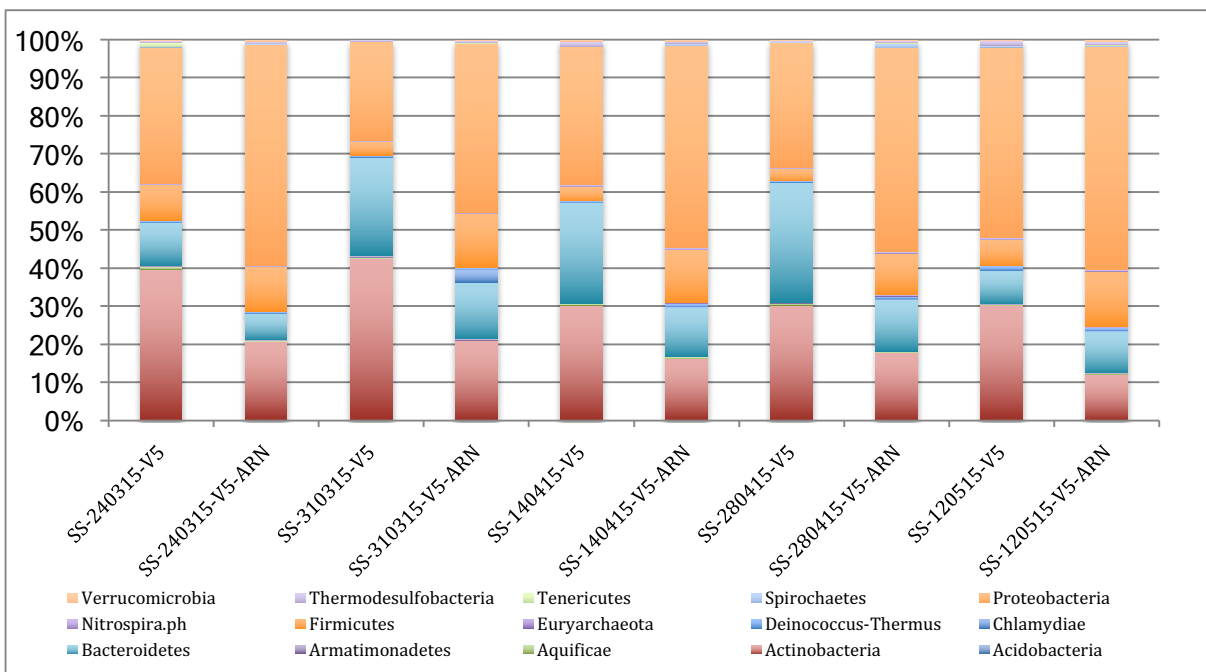


Figure 68 : évolution des populations bactériennes dans le sol au contact des souris, à partir des amplicons du fragment HVR5. Les 15 phyla majoritaires ont été représentés. ST : Sol Témoin ; SS : Sol Souris.

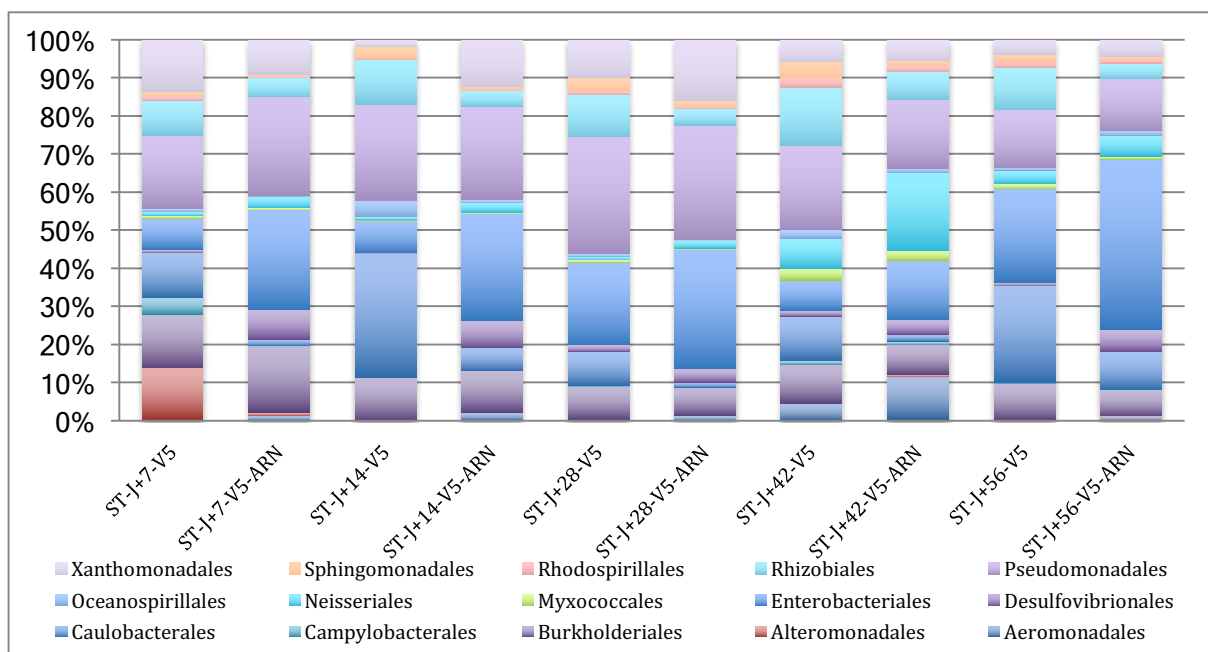


Figure 69 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols témoins. Les assignations ont été faites à partir des amplicons du fragment HVR5. ST : Sol Témoin.

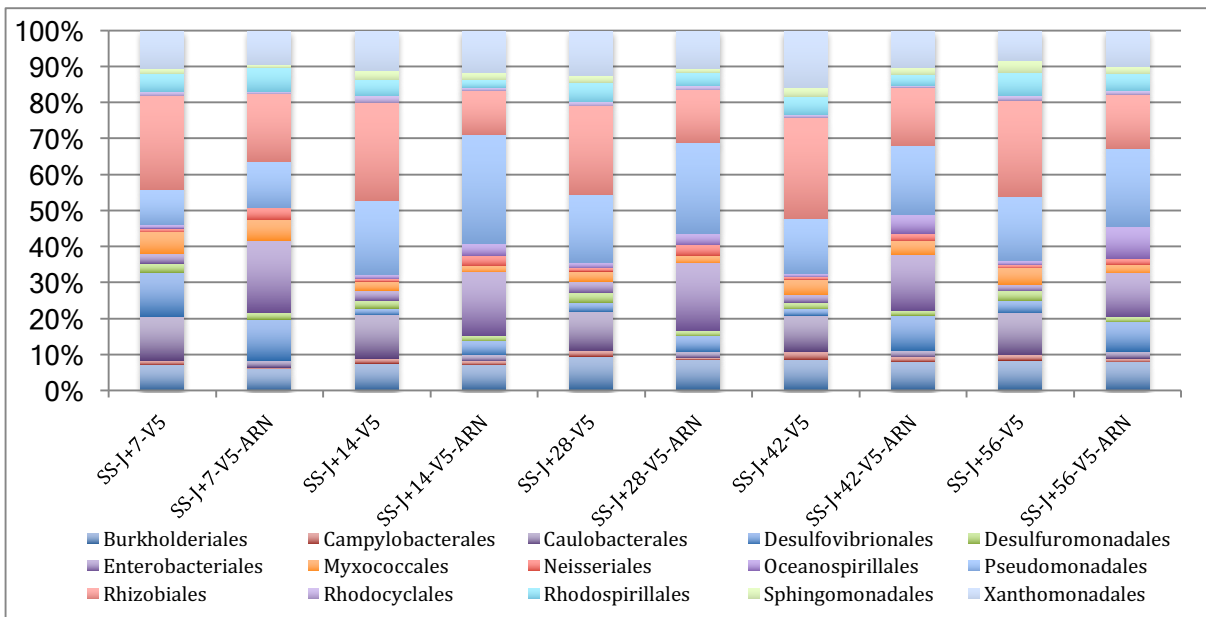


Figure 70 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols contact. SS : Sol Souris.

La dichotomie déjà observée entre bactéries présentes et bactéries actives a été retrouvée ici. Pour le sol témoin, les Caulobactérales, les Rhizobiales, les Sphingomonadales et les Xanthomonadales étaient présentes mais peu actives, tandis que les Entérobactériales et les Neisseriales étaient actives. Les Caulobactérales sont plutôt associées à des milieux aqueux, tandis que les Rhizobiales ont la capacité de fixer l'azote et sont des symbiontes des plantes. Les Sphingomonadales sont très largement répandus dans les différents milieux naturels et les Xanthomonadales associées aux plantes.

Globalement, six ordres (sur 41 présents) étaient responsables de 85% des séquences obtenues, avec dans l'ordre d'importance les Pseudomonadales, les Caulobactérales, les Entérobactériales, les Burkholdériales, les Rhizobiales et les Xanthomonadales.

Dans le cas du sol contact, les populations étaient plus diversifiées, avec 9 ordres pour représenter 86% des séquences obtenues. Les 3 ordres supplémentaires étaient les Desulfovibrionales, les Myxococcales et les Rhodospirilles. Le Tableau 12 résume ces pourcentages, en comparant entre sol témoin et sol contact. Comme pour le sol témoin, les Caulobactérales et les Rhizobiales étaient présentes mais peu actives, tandis que les Desulfovibrionales, les Entérobactériales et les Pseudomonadales étaient actives. Ces trois premiers ordres sont connus pour participer à la décomposition des substrats dans le sol.

Ordre	% dans le sol témoin			% dans le sol contact		
	total	ADN	ARN	total	ADN	ARN
Burkholderiales	10,55%	10,85%	9,21%	7,52%	8,14%	6,67%
Caulobacterales	22,64%	26,76%	4,39%	6,94%	10,81%	1,54%
Desulfovibrionales	1,42%	0,53%	5,35%	8,46%	4,13%	14,49%
Enterobacterales	13,85%	10,31%	29,51%	7,60%	2,52%	14,68%
Myxococcales	0,66%	0,59%	0,95%	3,28%	3,78%	2,58%
Pseudomonadales	23,31%	23,80%	21,17%	17,13%	15,88%	18,87%
Rhizobiales	10,44%	11,71%	4,83%	20,22%	25,28%	13,17%
Rhodospirillales	0,69%	0,66%	0,83%	4,25%	4,91%	3,34%
Xanthomonadales	4,52%	3,56%	8,76%	10,54%	11,73%	8,88%

Tableau 12 : pourcentage des séquences assignées, au sein du phylum protéobactéries, à partir des amplicons du fragment HVR5. Les ordres sélectionnés représentent 85% des séquences assignées pour le sol témoin et le sol contact. Les cellules grisées représentent les ordres minoritaires dans le sol témoin, qui ont fortement crû dans le sol au contact des cadavres. Les valeurs en rouge indiquent les populations actives.

Ce phénomène se déroulant en permanence dans le sol, on observe en effet un ratio ARN/ADN supérieur à 1 pour ces populations dans le sol contact et, pour deux d'entre elles, dans le sol témoin. Seules les Pseudomonadales ont montré une activité plus importante dans le sol contact. Lorsque l'analyse était focalisée sur l'ordre des Pseudomonadales, deux genres bactériens ont représenté plus de 95% des séquences assignées (Figure 71). Ces deux genres étaient Acinetobacter et Pseudomonas. Tous deux sont des bactéries Gram négatif en bâtonnets, aérobies. Présentes dans le sol, elles participent à la dégradation des composés organiques, ayant un arsenal enzymatique très élaboré.

Genre	% dans le sol témoin			% dans le sol contact		
	total	ADN	ARN	total	ADN	ARN
Acinetobacter	17,07%	14,19%	31,41%	9,80%	6,85%	13,25%
Azotobacter	0,14%	0,01%	0,78%	0,05%	0,06%	0,04%
Cellvibrio	0,22%	0,15%	0,60%	0,13%	0,13%	0,13%
Dasania	0,03%	0,02%	0,05%	0,03%	0,02%	0,05%
Enhydrobacter	0,90%	1,07%	0,06%	0,40%	0,56%	0,20%
Moraxella	0,03%	0,03%	0,03%	0,04%	0,06%	0,02%
Pseudomonas	81,34%	84,37%	66,24%	87,37%	90,50%	83,69%
Psychrobacter	0,16%	0,06%	0,68%	1,97%	1,65%	2,34%
Rhizobacter	0,04%	0,04%	0,04%	0,07%	0,12%	0,02%

Figure 71 : pourcentage des séquences assignées au niveau du genre, au sein de l'ordre des pseudomonadales, à partir des amplicons du fragment HVR5. Les ordres sélectionnés représentent 85% des séquences assignées pour le sol témoin et le sol contact. Les valeurs en rouge indiquent les populations actives, c'est-à-dire présentant un ration ARN/ADN supérieur à 1.

A ce niveau de résolution, il n'a plus été observé de différence de populations actives entre sol témoin et sol contact, les mêmes genres bactériens étant impliqués dans les phénomènes de dégradation.

Lorsque le sol contact est analysé avec l'amplicon des fragments HVR345, la distribution des ordres bactériens dans le phylum des Protéobactéries est apparue encore plus claire (Figure 72). En comparant ADN et ARN, il est apparu une augmentation nette de l'activité des Entérobactériales, mais surtout des Pseudomonadales, ce dès la première semaine. L'activité des Pseudomonadales est restée élevée tout au long de l'expérimentation. Plus discrète, l'augmentation des Burkholdériales était néanmoins visible et a persisté tout au long de l'expérimentation, tandis que les Entérobactériales ont décliné en activité.

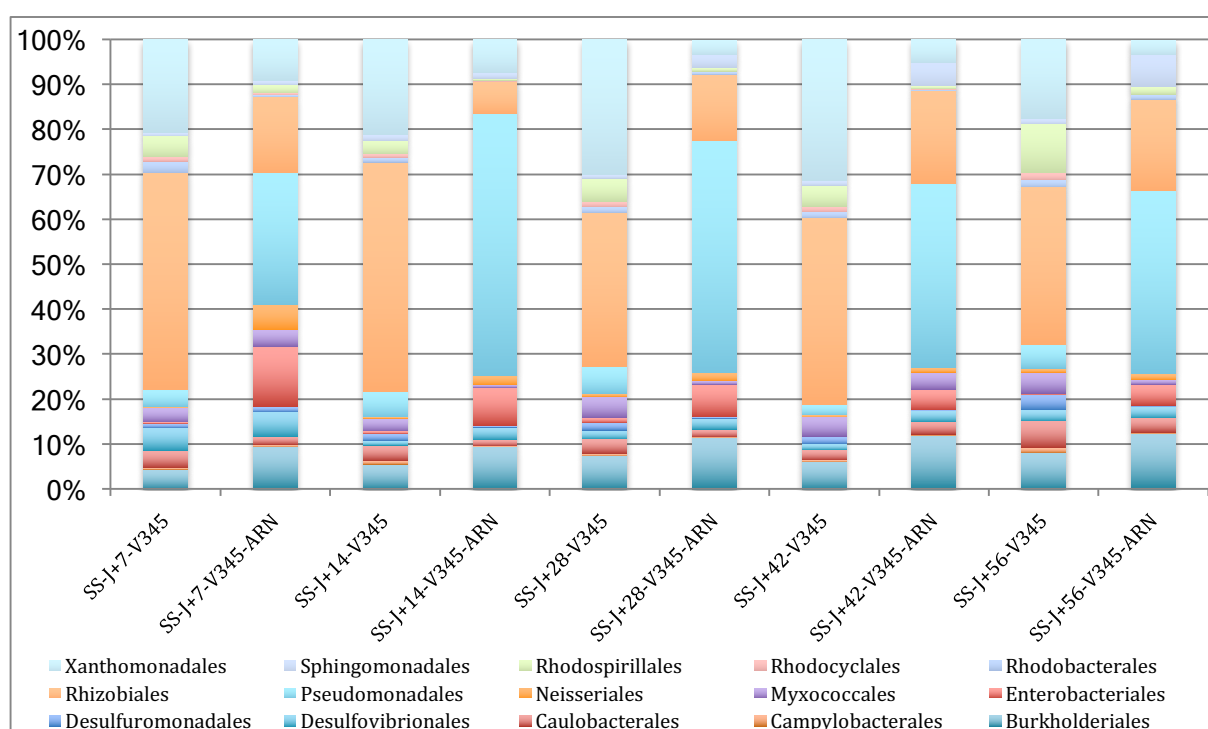


Figure 72 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols témoins. Les assignations ont été faites à partir de l'amplicon des fragments HVR 3, 4 et 5. SS : Sol Souris.

Les analyses en PCA et PLS-DA réalisées à ce niveau de taxonomie ont été représentées dans la Figure 73. De nouveau, le fragment HVR-5 semble manquer de pouvoir discriminant, les différents groupes d'échantillons manquant d'homogénéité et se recouvrant les uns les autres. Par contre, le fragment HVR-345 montre une séparation nette des deux groupes de sol, tandis que les échantillons osseux, en surface ou enfouis, ont été très comparables, particulièrement en PLS-DA. Il n'est cependant pas observé de corrélations entre analyse en composantes principales et temps de décomposition.

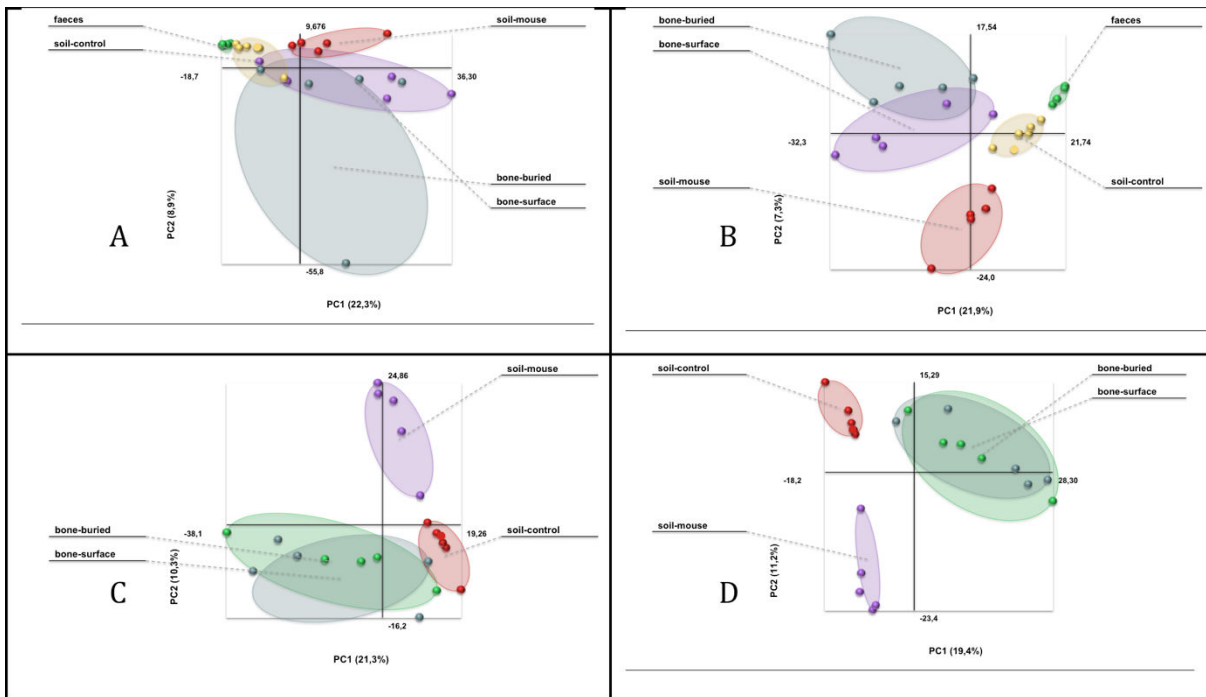


Figure 73 : A : séparation par PCA (A et C) et par PLS-DA (B et D) des différentes populations analysées, au niveau du phylum, à partir des HVR 5 (A, B) et HVR 345 (C, D). En rouge le sol contrôle, en violet le sol au contact des souris, en vert les os enfouis et en gris les os en surface.

Sur cette étude à court terme, il a été possible de démontrer un effet du mode de décomposition quant aux phénomènes taphonomiques, illustrés par les différences de profils bactériens entre souris laissées à la surface et souris enfouies. Ces phénomènes tendent, avec le temps, et relativement rapidement à l'échelle de la diagenèse (qui peut s'étendre sur plusieurs années voire siècles), à s'atténuer et il devient difficile de discriminer à partir des profils bactériens des ossements issus de dépouilles laissées à la surface et ossements issus de dépouilles enfouies. Néanmoins, ces phénomènes précoces peuvent avoir un effet à long terme sur la conservation de l'ADN endogène, par la création de microenvironnements différents à ces périodes cruciales, où l'ADN est peut-être le plus exposé à la dégradation, tant biologique par les bactéries et autres micro-organismes saprophytes, que chimiques car n'étant plus protégé à l'intérieur des noyaux cellulaires. Il a été évoqué dans la littérature l'idée de foyers favorables à la conservation de l'ADN à long terme, celui-ci par exemple en se complexant à l'hydroxy-apatite présente dans les os. La différence entre enfouissement et décomposition en surface peut modifier la nature et la quantité de ces foyers, et donc influencer sur la conservation à très long terme de l'ADN endogène. Notre étude ne permet pas d'étudier cette éventuelle hypothèse, qui nécessite des études comparatives longues et à plus grande échelle.

4. PROTOCOLE DE TAPHONOMIE A MOYEN TERME

Après ces premiers éléments issus de l'étude à court terme, nous pouvons observer les résultats de l'étude à moyen terme, portant sur 30 mois. Dans ce protocole, toutes les souris ont été enfouies, et les mesures de température n'ont été faites que dans le sol.

4.1. Suivi de la température

La Figure 74 illustre l'évolution de ces températures au cours des trente mois de l'expérimentation.

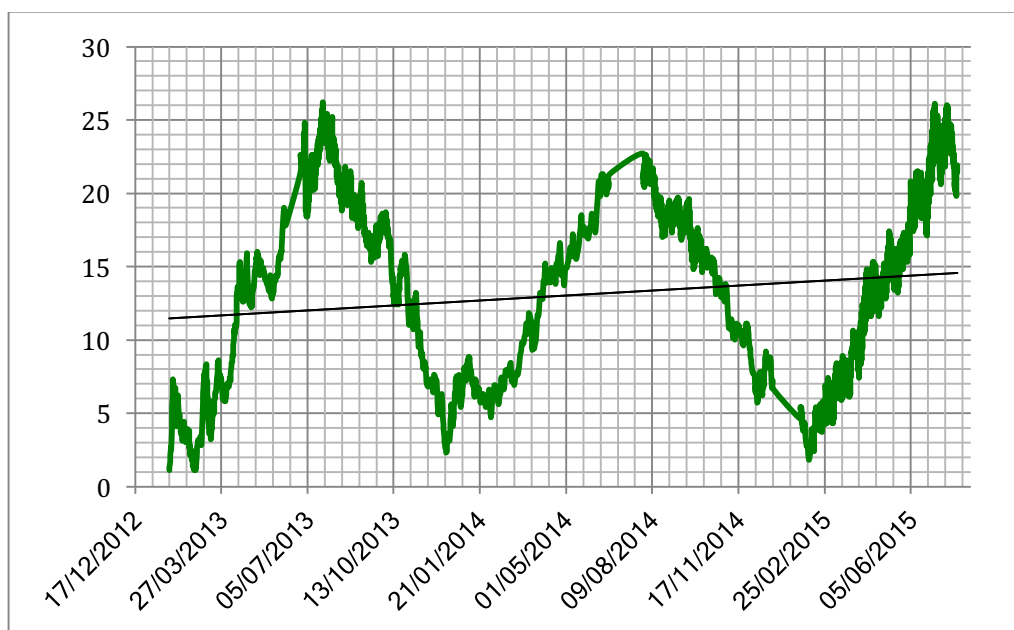


Figure 74 : évolution de la température dans le sol entre janvier 2013 et juillet 2015

La succession des saisons est clairement visible, avec un hiver plus clément en 2013-2014.

4.2. Observations macroscopiques

La Figure 75 montre l'évolution de la dégradation des dépouilles de souris au cours du temps. Comme déjà indiqué, les souris après 15 jours ne présentaient aucun signe de dégradation. Après 3 mois, elles ressemblaient aux dépouilles entre J+42 et J+56 de l'étude à court terme, qui s'était déroulée dans des conditions de température plus élevées, accélérant les processus de décomposition. A 6 mois, la dépouille était entièrement ensachée dans un réseau qui, après observation au microscope, s'est révélée être des racines (et non un mycélium).






J+15	
J+3mois	
J+6 mois	
J+12 mois	
J+18 mois	

Figure 75 : photographies des différents états de dégradation des dépouilles de souris au fur et à mesure du temps.

Afin d'identifier la plante impliquée, nous avons amplifié l'ADN extrait de ces racines avec deux gènes de ménage servant à l'identification des broméliacées, *matK* et *rbcL* [206]. Après séquençage des amplicons, il a été possible d'identifier la plante comme étant la fétuque rouge (*Festuca rubra* - Figure 76).



Figure 76 : planches de fétuque rouge (*Festuca rubra*)

Cette graminée est connue pour être très répandue dans l'hémisphère nord, avec la particularité de produire des racines plus longues que nombre de graminées, ce qui lui a permis ici d'atteindre les dépouilles, qui sont devenues une source de nutriments, et ont favorisé le développement du réseau racinaire.

A J+9 mois, le réseau racinaire était toujours présent, alors qu'il ne restait guère plus que les os des souris, l'ensemble des tissus ayant été dégradé. A partir de J+12 mois, il n'est resté que les os, encore entourés d'un réseau racinaire.

4.3. Evolution de la population bactérienne

Après ensemencement et culture, il a été possible de dénombrer les bactéries ayant poussé en aérobiose et anaérobiose. Le Tableau 13 présente les données collectées lors des 8 temps de prélèvement.

date	prélèvement		moyen	pH	numération aérobie			numération anaérobie			
					dilution	a	b	quantification	a	b	quantification
01-13	trou souris	pelle stérile	7,4	-3	86	97	9,15E+005	54	55	5,45E+005	
				-4	9	13	1,10E+006	9	8	8,50E+005	
				-5	4	1	2,50E+006			ND	
02-13	sol blanc	pelle stérile	7,32	-3	141	155	1,48E+006	64	58	6,10E+005	
				-4	11	6	8,50E+005	5	8	6,50E+005	
				-5	3	2	2,50E+006	2	0	1,00E+006	
02-13	sol souris	récolte terre dans cage souris exhumée	7,38	-3	ind	ind	ND	ind	ind	ND	
				-4	82	89	8,55E+006	51	55	5,30E+006	
				-5	12	10	1,10E+007	7	7	7,00E+006	
04-13	sol souris SS	terre associée au cadavre	8,1	-3	ind	ind	ND	ind	ind	ND	
				-4	136	119	1,87E+008	97	109	1,51E+008	
				-5	21	20	3,30E+008	21	13	2,74E+008	
	sol souris Xa	terre associée au cadavre			-3	ind	ind	ND	ind	ind	ND
					-4	139	150	2,12E+008	81	87	1,23E+008
					-5	67	57	ND	41	30	5,72E+008
	sol souris Bg	terre associée au cadavre			-3	ind	ind	ND	ind	ind	ND
					-4	86	80	1,22E+008	61	62	9,01E+007
					-5	14	12	2,09E+008	10	4	1,13E+008
	sol blanc	pelle stérile		7,81	-3	69	57	6,30E+005	ind	ind	ND
					-4	9	11	1,00E+007	3	2	2,50E+006
					-5	2	1	1,50E+007	0	0	0,00E+000
06-13	sol souris SS	terre associée au cadavre	7,32	-3	ind	ind	ND	ind	ind	ND	
				-4	124	123	1,24E+008	73	103	8,80E+007	
				-5	31	21	2,60E+008	30	23	2,65E+008	
	sol souris Xa	terre associée au cadavre		7,33	-3	ind	ind	ND	ind	ind	ND
					-4	55	116	8,55E+007	42	37	3,95E+007
					-5	8	5	6,50E+007	10	7	8,50E+007
	sol souris Bg	terre associée au cadavre		7,29	-3	ind	ind	ND	ind	ind	ND
					-4	78	111	9,45E+007	49	87	6,80E+007

date	prélèvement	moyen	pH	numération aérobie			numération anaérobie			
				dilution	a	b	quantification	a	b	quantification
	sol blanc	pelle stérile	7,38	-5	13	12	1,25E+008	6	12	9,00E+007
				-3	ind	ind	ND	ind	ind	ND
				-4	114	134	1,24E+007	40	18	2,90E+006
				-5	7	8	7,50E+006	12	12	1,20E+008
09-13	sol souris SS	terre associée au cadavre	8,29	-3	ind	ind	ND	3	3	3,00E+005
				-4	29	30	2,95E+007	0	1	5,00E+005
				-5	5	20	1,25E+008	0	0	0,00E+000
	sol souris Xa	terre associée au cadavre	8,2	-3	ind	ind	ND	27	14	2,05E+006
				-4	47	55	5,10E+007	1	0	5,00E+005
				-5	24	27	2,55E+008	0	0	0,00E+000
	sol souris Bg	terre associée au cadavre	8,22	-3	ind	ind	ND	70	64	6,70E+006
				-4	90	ind	9,00E+007	-	9	9,00E+006
				-5	28	22	2,50E+008	5	4	4,50E+007
	sol blanc	pelle stérile	8,06	-3	208	193	2,01E+006	6	19	1,25E+006
				-4	34	24	2,90E+006	0	2	1,00E+006
				-5	0	3	1,50E+006	0	0	0,00E+000
01-14	sol souris SS	terre associée au cadavre	7,7	-3	93	114	1,04E+007	nd	nd	ND
				-4	26	34	3,00E+007	0	2	1,00E+006
				-5	4	11	7,50E+007	0	0	0,00E+000
	sol souris Xa	terre associée au cadavre	7,8	-3	205	194	2,66E+007	nd	nd	ND
				-4	32	40	3,60E+007	2	0	1,00E+006
				-5	10	14	1,20E+008	0	0	0,00E+000
	sol souris Bg	terre associée au cadavre	7,6	-3	ind	ind	ND	nd	nd	ND
				-4	28	38	3,30E+007	4	1	2,50E+006
				-5	16	4	1,00E+008	0	0	0,00E+000
	sol blanc	pelle stérile	6,9	-3	76	82	7,90E+005	1	3	2,00E+005
				-4	26	24	2,50E+006	0	0	0,00E+000
				-5	10	8	9,00E+006	0	0	0,00E+000
07-14	sol souris SS	terre associée au cadavre	7,6	-3	91	93	2,38E+006	13	11	3,05E+005
				-4	13	23	4,58E+006	1	0	1,27E+005

date	prélèvement	moyen	pH	numération aérobie			numération anaérobie				
				dilution	a	b	quantification	a	b	quantification	
01-15	sol souris Xa	terre associée au cadavre		-5	1	0	1,27E+006	0	0	0,00E+000	
				-3	ind	ind	5,21E+005	19	21	1,28E+005	
				-4	85	78	1,09E+006	1	3	1,28E+005	
	sol souris Bg	terre associée au cadavre		-5	19	15	1,28E+006	0	0	0,00E+000	
				-3	ind	ind	8,56E+005	26	14	1,23E+005	
				-4	129	150	1,26E+006	2	1	9,20E+004	
	sol blanc	pelle stérile	8,32	-5	21	20	6,13E+005	0	1	3,07E+005	
				-3	ind	ind	1,55E+005	10	16	5,43E+004	
				-4	39	35	1,67E+005	0	0	0,00E+000	
	01-15	sol souris SS	terre associée au cadavre	8,22	-5	1	7	2,09E+005	0	0	0,00E+000
					-3	ind	ind	ND	8	3	9,00E+004
					-4	55	58	5,65E+006	2	0	1,00E+005
sol souris Xa		terre associée au cadavre	8,31	-5	2	5	3,50E+006	0	0	0,00E+000	
				-3	ind	ind	ND	5	7	6,00E+004	
				-4	43	35	3,90E+006	0	0	0,00E+000	
sol souris Bg		terre associée au cadavre	8,37	-5	4	3	3,50E+006	0	0	0,00E+000	
				-3	ind	ind	ND	2	5	3,50E+004	
				-4	87	61	7,40E+006	0	1	5,00E+004	
sol blanc		pelle stérile	8,27	-5	9	12	1,05E+007	0	0	0,00E+000	
				-3	146	153	1,50E+006	3	1	2,00E+004	
				-4	13	12	1,25E+006	0	0	0,00E+000	
			-5	2	1	1,50E+006	0	0	0,00E+000		

Tableau 13 : données de dénombrement sur les prélèvements issus du premier protocole. ND : non déterminé. Ind : indénombrable, le nombre de colonies étant trop élevé pour être compté.

De nouveau, nous avons pu visualiser l'évolution des populations bactériennes aérobies et anaérobies au cours du temps (Figure 77)

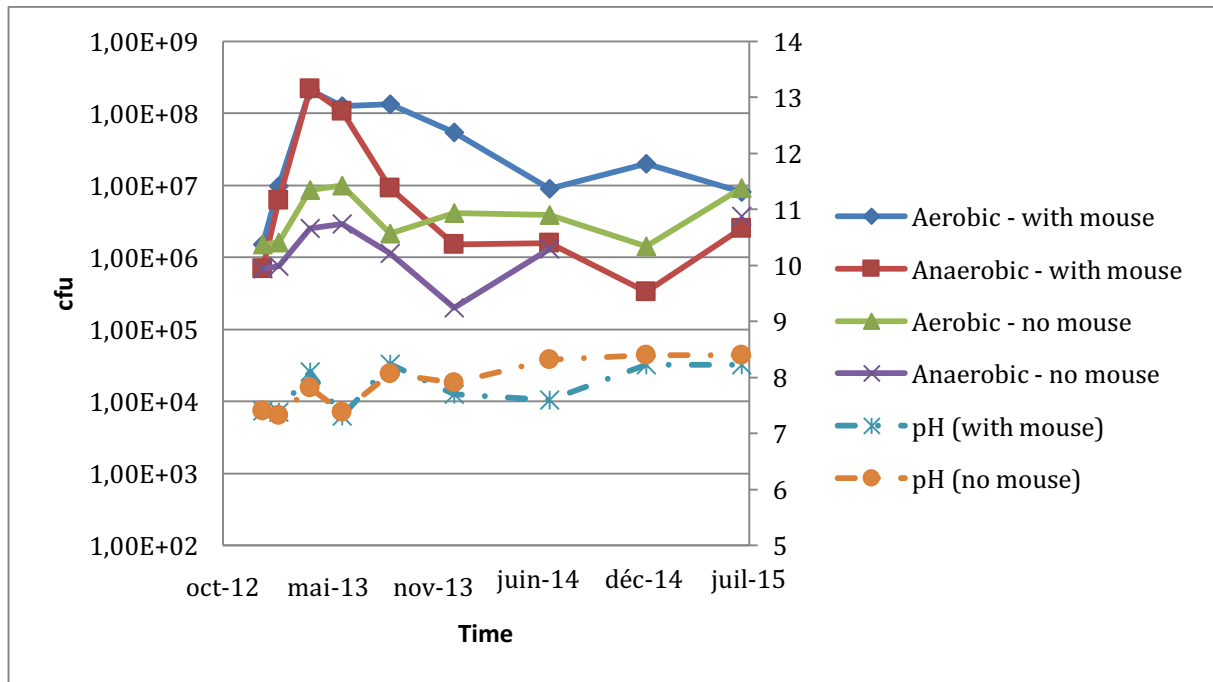


Figure 77 : évolution du pH et des populations bactériennes dans un sol témoin (courbes vert et violet) et à proximité immédiate de souris (courbes bleu et rouge). L'axe principal suit une échelle logarithmique.

Lorsque le sol témoin est analysé, il est noté une variation de la population bactérienne de l'ordre d'un logarithme décimal, entre 10^6 et 10^7 bactéries aérobies par gramme de sol et autour de 10^6 pour les bactéries anaérobies. Il n'est pas observé, à l'échelle des prélèvements de variation nette du pH, tant dans le sol témoin que dans le sol au contact des souris.

Lorsqu'est observée la courbe des bactéries aérobies dénombrées dans le sol à proximité immédiate de la souris, il est noté dès la deuxième semaine, et malgré l'apparence intacte de la dépouille, une augmentation de la population bactérienne, augmentation qui permet d'atteindre une population 100 fois plus importante que dans le sol témoin après trois mois. L'accroissement des bactéries a été similaire pour la communauté anaérobie. Après 6 mois, une décroissance a été observée. Celle-ci s'est poursuivie assez rapidement pour la communauté anaérobie, pour revenir après un an au niveau du sol témoin. Par contre, pour la communauté aérobie, la décroissance a été lente et n'a retrouvé un niveau comparable avec le sol témoin qu'après 30 mois.

Les bactéries anaérobies ont dans leur majorité une capacité saprophyte plus importante que les bactéries aérobies, et se retrouvent donc plus rapidement en manque

de nutriments, revenant plus vite à leur état pré-enrichissement, tandis que les bactéries aérobies ont pu continuer plus longtemps à exploiter la niche nutritive créée par le cadavre.

4.4. Analyse de l'ADN murin

L'ADN a été extrait d'échantillons de sol témoin et de sol au contact immédiat de souris enfouies, ainsi que les fémurs des souris enfouies après pesée et élimination (si nécessaire) des chairs adhérentes par un traitement à la trypsine. A partir des extraits d'ADN issus des os, il a été possible d'évaluer la dégradation de l'ADN en comparant les concentrations de quatre fragments de taille différente (73 – 125 – 154 – 513 bp, Figure 78).

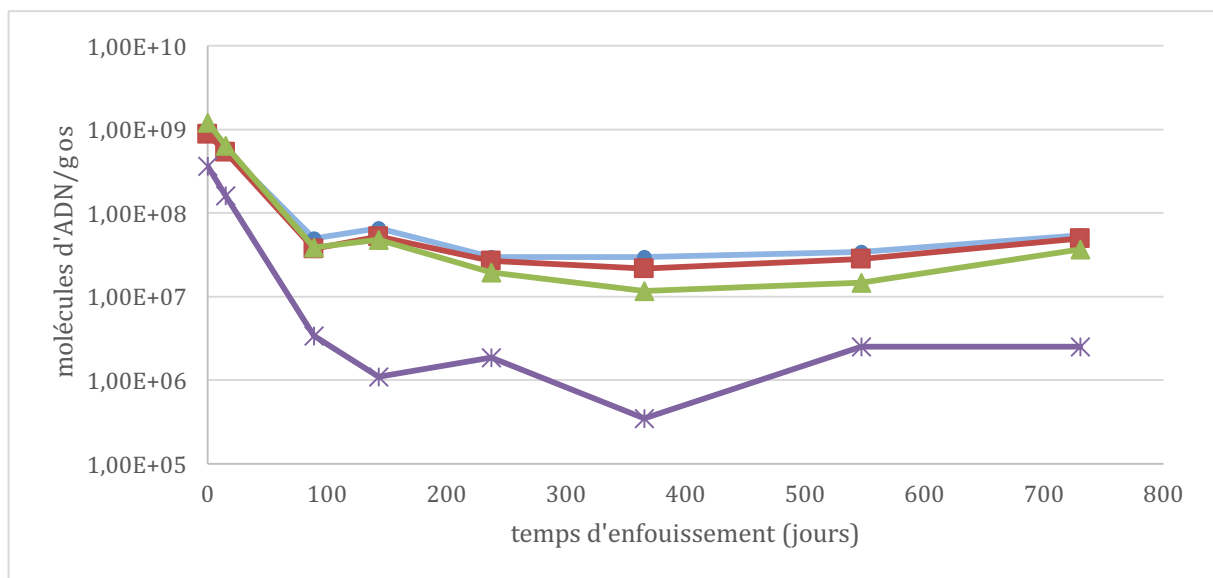


Figure 78 : Evolution du nombre de molécules d'ADN de taille variable en fonction du temps d'enfouissement. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb

Comme observé pour l'étude à court terme, le fragment de plus grande taille, 513 pb, s'est dégradé plus vite que les autres fragments, plus courts. Par contre, la différence de taille entre les fragments courts ne semble pas suffisante pour entraîner une différence observable de cinétique de dégradation.

Cette évolution a pu être confirmée en comparant le nombre de molécules par gramme d'os, en fonction du temps et de la longueur des fragments (Figure 79). Entre J0 et J+15, le nombre de molécules est resté comparable quelle que soit la longueur du fragment amplifié. Cela est rendu par la faible pente de la courbe, mesurée entre -0,002 et -0,003. Par la suite, les courbes ont présenté une pente négative (entre -0,006 et -0,01), montrant l'appauvrissement relatif en molécules de grande taille.

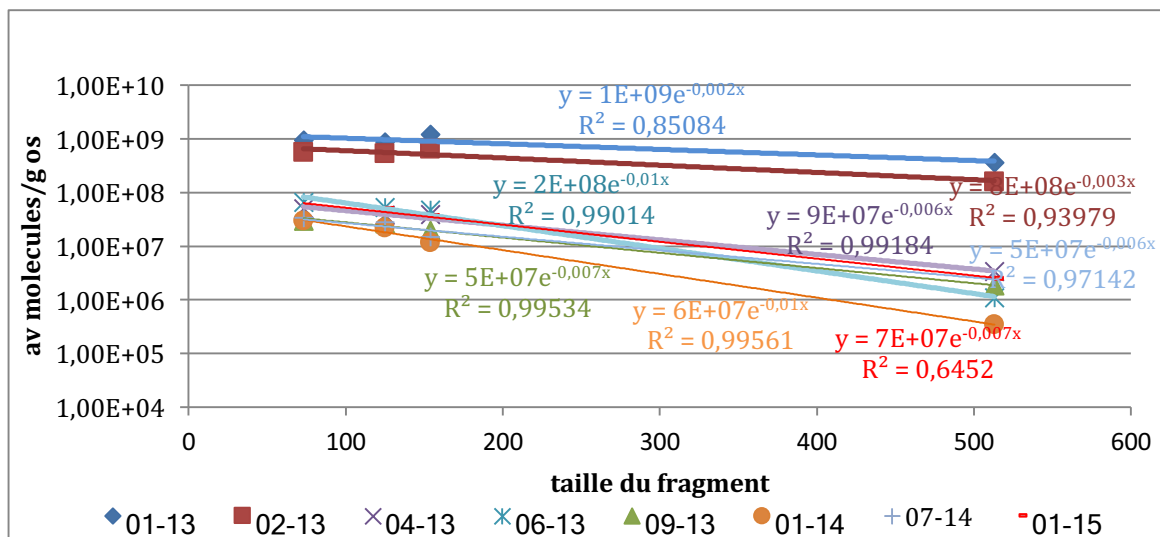


Figure 79 : évolution du nombre de molécules présentes en fonction du temps, dans les fémurs enfouis. Bleu : J0, brique : J+15 jours, violet J+3 mois, bleu turquoise: J+6 mois, vert : J+9 mois, orange : J+12 mois, bleu clair : J+18 mois, rouge : J+24 mois.

4.5. Séquençage massif

Les extraits de fémur, de sol témoin et de sol au contact des souris ont été amplifiés par un couple d'amorces amplifiant le fragment V5. Les résultats de séquençage ont été traités selon le script « usearch2taxo ». Les résultats d'amplification ont été rassemblés dans le Tableau 14. Par ailleurs, deux extraits de sol ont été séquencés en shotgun.

Ici encore, le nombre de séquences obtenu par échantillon est très variable, et il n'est pas observé de différence notable entre échantillons issus des os ou du sol.

Les banques obtenues à partir des extraits d'ARN, traités ou non à la DNase, étaient moins riches que ceux préparés à partir des extraits d'ADN. Les fragments amplifiés et séquencés ont une longueur moyenne de 140 pb, sauf pour les extraits d'ossements réalisés en juillet 2015, qui présentent une longueur entre 150 et 190 pb. Deux échantillons, réalisés en avril 2013, ont des longueurs aberrantes, ainsi qu'un échantillon en janvier 2014. Le taux de GC est en moyenne de 51%, dans les os et dans le sol. L'un des échantillons de juillet était également anormal avec un taux de 36%. Ces échantillons ont été préparés en même temps que d'autres, et il n'a pas été noté de fait particulier au cours de leur traitement. Du fait de ces valeurs aberrantes, ils ont été exclus des analyses ultérieures. Aucun sol ne présente de valeur aberrante.

Bones	raw	merged	aligned	mean length	% GC	% alignés	OTUs	Soils	raw	merged	aligned	mean length	%G C	% alignés	OTUs
CST-0114	1107526	1028579	926222	140,321	52	83,6%	25805	SS-0115	523602	484075	424309	140,793	52	81,0%	9548
CST-0413	597474	562720	526999	138,33	49	88,2%	9712	SS-0115-DNase	251141	231778	199925	140,599	51	79,6%	16160
F-BalbC	31284	30857	27389	139,349	50	87,5%	1036	SS0114-1500bp	46471	45187	39725	139,751	52	85,5%	2944
F-Bg	32889	32567	29559	139,491	51	89,9%	1178	SS0114-L3S	741381	729673	684691	139,633	52	92,4%	21948
F-cycloP	64640	63302	57894	139,22	51	89,6%	1642	SS0114-RNA	83286	81527	74458	139,96	51	89,4%	4068
F-Xa	49532	49116	45405	139,459	51	91,7%	1912	SS0413-1500bp	118773	117507	108896	138,802	50	91,7%	4407
FSB-0114	167156	154151	135898	140,16	51	81,3%	5519	SS0413-L3S	1450474	1427347	1369340	138,853	50	94,4%	25583
FSB-0115	342513	313861	280568	141,094	54	81,9%	8774	SS0413-RNA	19512	19246	17068	139,18	50	87,5%	2346
FSB-0213	281270	262910	242130	140,329	51	86,1%	5879	SS0413-RNA-DNase	57701	56235	51996	139,01	50	90,1%	1193
FSB-0413	588190	555216	518781	138,912	50	88,2%	10213	SS0713-1500bp	65363	64420	59453	139,147	51	91,0%	2468
FSB-0613	263013	244241	222365	139,035	49	84,5%	6541	SS0713-L3S	908340	891616	854250	139,731	51	94,0%	15971
FSB-0714	325274	298890	268199	141,17	52	82,5%	7758	SS0713-RNA	51105	50530	47132	139,403	50	92,2%	1723
FSB-0715	41862	40337	39397	151,194	36	94,1%	549	SS0714-1500bp	154531	153080	140001	140,074	53	90,6%	6389
FSB-0913	165887	154466	138472	139,937	51	83,5%	5313	SS0714-L3S	612604	600151	562892	139,639	53	91,9%	16839
FSB-J15	198830	184192	167161	140,062	51	84,1%	4610	SS0714-RNA	38911	38475	34511	140,391	52	88,7%	3624
FSR-0114	717561	669551	616048	139,813	51	85,9%	15085	SS0714-RNA-DNase	64609	63956	58183	140,214	52	90,1%	2577
FSR-0115	535180	494373	447447	140,427	53	83,6%	12857	SS15-1500bp	298125	294877	276159	140,406	52	92,6%	9536
FSR-0413	716398	673738	634537	139,67	50	88,6%	10571	SS15-L3S	1064898	1028790	975377	140,279	52	91,6%	19660
FSS-0715	202387	177649	143098	151,426	52	70,7%	7455	SS15-RNA	32136	31771	28806	139,946	50	89,6%	2041
FST-0114	1085989	1003986	937843	139,412	51	86,4%	14923	SS15-RNA-DNase	43103	42456	39216	139,23	51	91,0%	1648
FST-0115	565569	526712	480998	140,644	53	85,0%	13425	ST-0115	555530	520671	460671	140,691	53	82,9%	8481
FSX-0715	246968	218504	160795	190,885	54	65,1%	12490	ST-0115-DNase	210196	195378	169156	140,973	52	80,5%	17569
HST-0413	618075	593008	529367	139,876	51	85,6%	15782	ST0114-RNA	59251	58149	52796	140,256	52	89,1%	3279
MOCK-0114	267476	248867	223601	140,723	50	83,6%	6941	ST0413-1500bp	87551	86707	78532	139,811	54	89,7%	4640
MSB-0114	147697	138380	124502	139,95	52	84,3%	4458	ST0413-L3S	470952	454369	418547	140,51	54	88,9%	15765
MSB-0115	355067	329907	297752	139,36	54	83,9%	8937	ST0413-RNA	76495	75298	67926	140,56	52	88,8%	4456
MSB-0213	205587	188826	170625	140,431	52	83,0%	4406	ST0713-1500bp	90799	90029	81859	139	52	90,2%	4525
MSB-0413	522755	489202	453303	166,007	50	86,7%	10436	ST0713-L3S	172111	169309	154663	139,479	52	89,9%	7352
MSB-0613	175266	164057	149298	140,314	49	85,2%	4928	ST0713-RNA	25468	25002	21582	140,085	51	84,7%	2412
MSB-0714	253083	228569	198819	140,382	52	78,6%	7471	ST0713-RNA-DNase	34931	34415	30634	139,743	51	87,7%	1936

MSB-0715	162554	137717	108648	140,31	52	66,8%	6558	ST0714-1500bp	89199	88170	79619	140,283	54	89,3%	4793
MSB-0913	176908	165238	149678	140,464	50	84,6%	5322	ST0714-L3S	728134	719617	681541	140,115	53	93,6%	19193
MSB-J15	173072	159126	141171	138,934	52	81,6%	4264	ST0714-RNA	60215	58443	52565	140,38	52	87,3%	1576
MSR-0114	506625	476452	435079	155,209	52	85,9%	11972	ST0714-RNA-DNAse	15990	15718	13233	140,189	52	82,8%	3548
MSR-0115	395723	363432	319966	140,822	52	80,9%	12454	ST15-1500bp	185809	181376	164056	139,994	53	88,3%	8168
MSR-0413	695967	645324	600987	158,812	49	86,4%	12013	ST15-L3S	326378	320139	294467	139,533	52	90,2%	13934
MSS-0715	116394	103750	80309	140,963	52	69,0%	4938	ST15-RNA	137308	135500	124145	139,938	52	90,4%	3397
MST-0115	382588	352969	316363	145,97	54	82,7%	9629	ST15-RNA-DNAse	49254	48626	43288	139,923	52	87,9%	6838
MSX-0715	104792	98909	73641	140,696	54	70,3%	5835	STJ0-1500bp	111092	109895	99422	139,437	53	89,5%	5609
VID-1C-R3HU	242068	218184	189986	140,875	53	78,5%	7175	STJ0-L3S	278803	275344	256291	139,595	53	91,9%	11213
VID-1C-R3U1	260942	245679	220021	140,621	52	84,3%	7712	STJ0-RNA	44635	43324	37952	140,136	51	85,0%	2882
VID-2C-R3U1	258174	239696	211766	140,535	52	82,0%	8023	STJ0-RNA-DNAse	43191	42166	37289	140,03	52	86,3%	2930
OLIV-R2150	256377	238227	211536	140,294	51	82,5%	7669	Mock-sol	603586	569783	527393	140,854	50	87,4%	11524
Total	14190607	13193434	11947707	143,5	51		331115		11496919	11142128	10299931	140,0	51,8		355748
Average	354765,2	329835,9	298692,7				8277,9		249933	278553,2	257498,3				8893,7

Tableau 14 : résultats du séquençage des banques créées à partir des extraits osseux et des fèces. Le nombre d'OTUs est donné après élimination des séquences présentes en une seule copie (singleton), au niveau taxonomique du genre (niveau le plus bas obtenu avec usearch). Les valeurs aberrantes (longueur et taux de GC) ont été repérées en rouge.

A partir des assignations taxonomiques, il a été possible dans un premier temps de représenter les proportions des différents phyla et leur évolution au cours du temps. La Figure 80 donne une représentation des résultats obtenus, au niveau taxonomique du phylum, pour les sols témoins et contact.

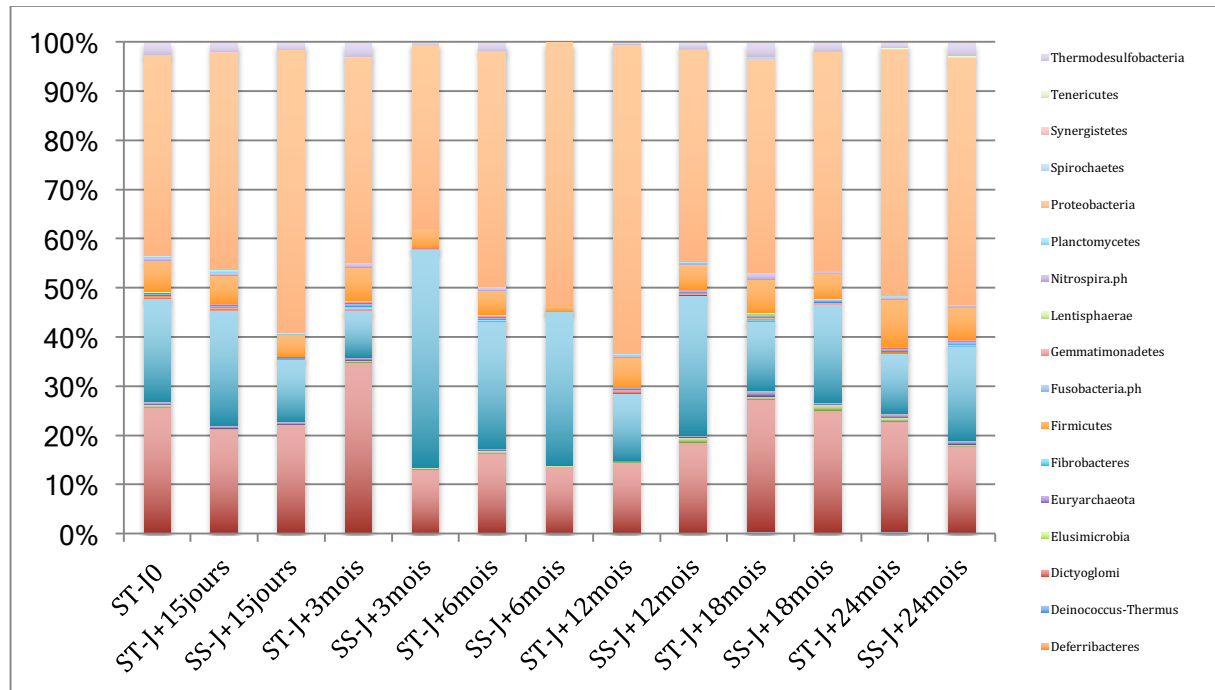


Figure 80 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau du phylum. Les extraits de sol ont été regroupés deux à deux en fonction du délai de décomposition. Les 10 phyla majoritaires ont été représentés et la mesure du sol témoin de janvier 2014 est manquante. ST : Sol Témoin ; SS : Sol Souris.

Les histogrammes alternent sols témoin et sols contact, afin de comparer à la même date les profils bactériens. La composition globale, au niveau du phylum, a assez peu varié pour le sol témoin, avec les Protéobactéries toujours en majorité, puis Actinobactéries, Bacteroidetes et Firmicutes. Les Bacteroidetes ont nettement dépassé en proportion les Actinobactéries en une seule occasion sur sept (J+6 mois). La différence avec le sol est apparue dès J+15, et cela malgré l'apparente conservation des dépouilles. Si la proportion d'Actinobactéries est restée constante, les Bacteroidetes ont diminué, les Protéobactéries ont commencé à augmenter. A trois mois les Bacteroidetes étaient majoritaires, tandis que les Firmicutes tendaient à diminuer, voire avaient presque disparu à 6 mois. Après 12 mois, où les Bacteroidetes étaient encore très représentés, les profils de sol témoin et de sol contact ont commencé à se rapprocher, ce qui est devenu plus visible pour les points de mesure à 18 et 24 mois.

La population de Bacteroidetes étant celle fluctuant le plus, il était intéressant de comparer sa composition entre sol témoin et sol contact. Mais cette analyse a tourné

court car l'ensemble était monophylétique, entièrement composé du genre *Bacteroides*. Des analyses statistiques basées sur la PCoA, déjà en cours, pourront éventuellement permettre de mettre en évidence des populations spécifiques.

Nous nous sommes aussi intéressés aux protéobactéries (Figure 81) :

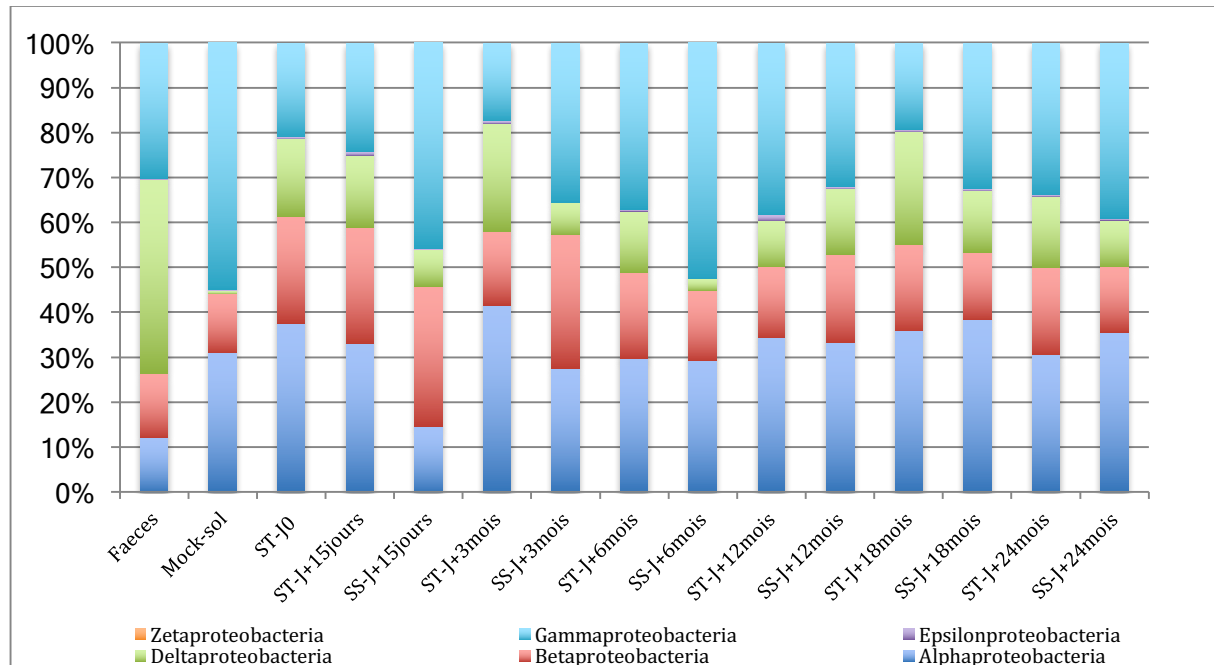


Figure 81 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau des protéobactéries. Les extraits de sol ont été regroupés deux à deux en fonction du délai de décomposition. ST : Sol Témoin ; SS : Sol Souris.

Il est apparu, dès le niveau de la classe, des différences entre sol contact et sol témoin. Pour ce dernier, il est observé en général une majorité d'Alphaprotéobactéries, puis Betaprotéobactéries et Gammaprotéobactéries, et ensuite Deltaprotéobactéries. Pour le sol contact, dès J+15, les Gammaprotéobactéries sont devenues majoritaires et les Deltaprotéobactéries sont devenues très minoritaires. A 3 mois, les Alphaprotéobactéries ont commencé à remonter, tandis que les Deltaprotéobactéries restent minoritaires. Il a fallu attendre 12 mois pour retrouver des proportions comparables de protéobactéries entre sol témoin et sol contact.

A un niveau taxonomique inférieur, au sein des Gammaprotéobactéries, les sol contact et sol témoin ont présenté des différences de populations bactériennes, comme représenté dans la Figure 82.

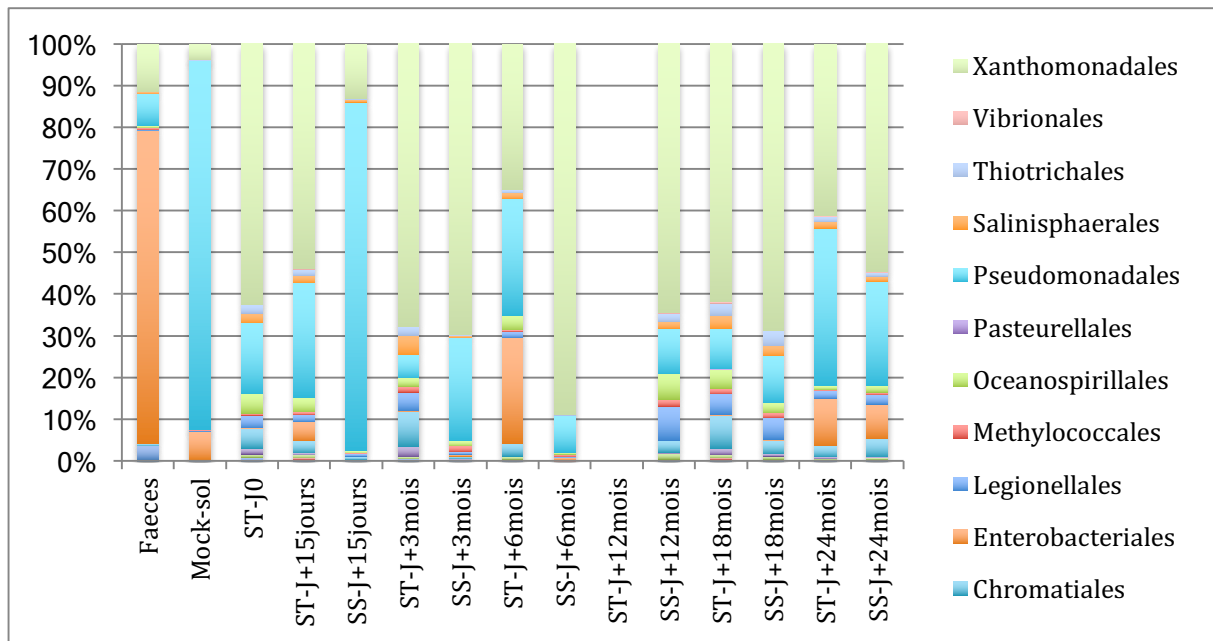


Figure 82 : représentation de la diversité bactérienne déduite des amplicons V5, au sein des Gammaprotéobactéries. ST : Sol Témoin ; SS : Sol Souris.

Le sol témoin présentait un profil plus diversifié tandis que le sol contact montrait une forte prédominance de Pseudomonadales dès 15 jours, qui ont régressé dès 3 mois, avec une très forte présence de Xanthomonadales. Au sein des Pseudomonadales, se trouvaient les genres bactériens *Acinetobacter* et *Pseudomonas*, ubiquitaires dans le sol et ayant la capacité de métaboliser de nombreux nutriments, ce qui les rend actifs dans les phénomènes de dégradation dans le sol.

Après 12 mois, les profils des sol contact et sol témoin ont commencé à se ressembler. Les expérimentations à court terme et à long terme ont été réalisées dans deux sols différents, ce qui peut expliquer des différences de profil initiales.

Lors des analyses sur échantillons anciens, nous avons procédé à l'extraction et au séquençage de l'ADN présent dans divers ossements restés enterrés plusieurs centaines, voire plusieurs milliers d'années dans le sol (Cf. 3.1). Il a ensuite été dressé les profils bactériens de ces échantillons, et il est apparu que les bactéries identifiées étaient communément trouvées dans le sol, ou étaient ubiquitaires. Nous avons également analysé les profils de dégradation des ADN amplifiés, grâce au logiciel mapDamage [50]. Lorsque nous regardons les profils des ADN endogènes, nous observons un profil de dégradation caractéristique de l'ADN ancien, avec une augmentation de la proportion d'adénine en 5' et un pourcentage élevé de cassure de l'ADN sur une guanine. La Figure 83 est une parfaite illustration de séquences authentiques issues d'un échantillon ancien, avec une fréquence élevée de mésappariements aux extrémités des fragments,

lié à la déamination de cytosines, interprétées par la Taq polymérase comme un uracile, et une fréquence élevée de guanine en position 5' immédiate du début du fragment.

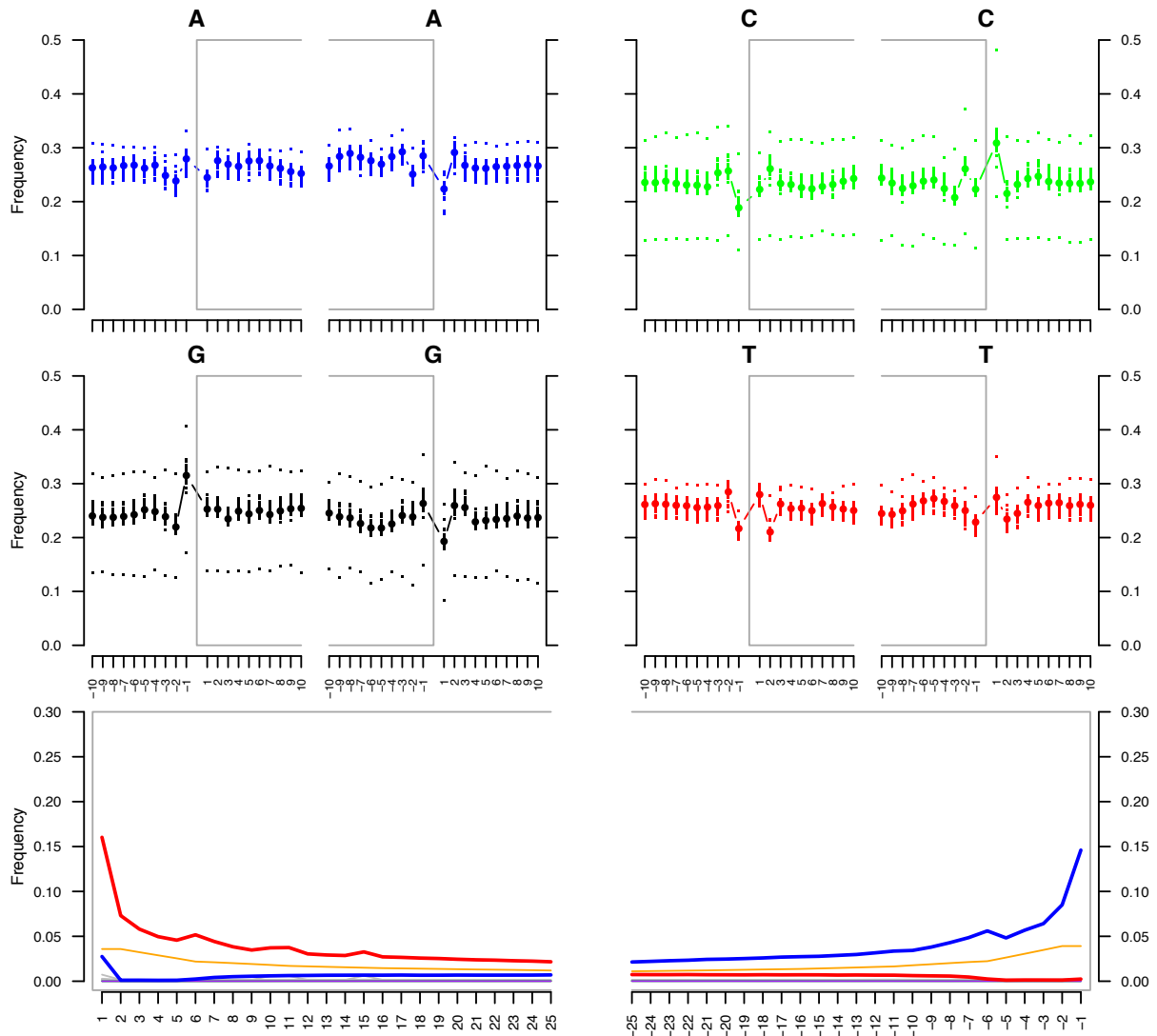


Figure 83 : représentation des résultats de l'analyse des séquences identifiées comme humaines dans un échantillon ancien (kyste de Boulogne-Sur-Mer). Les quatre premiers graphes représentent les fréquences relatives des quatre bases dans la séquence identifiée et dans son environnement dans la séquence de référence. Les deux derniers graphes représentent les fréquences de mis-appariement selon les bases par rapport aux extrémités 5' (à gauche) et 3' (à droite) des séquences amplifiées et appariées au génome de référence.

A contrario, lorsque la même analyse a été réalisée sur les séquences alignées sur la base de séquences bactériennes, les profils observés ont été très différents (Figure 84).

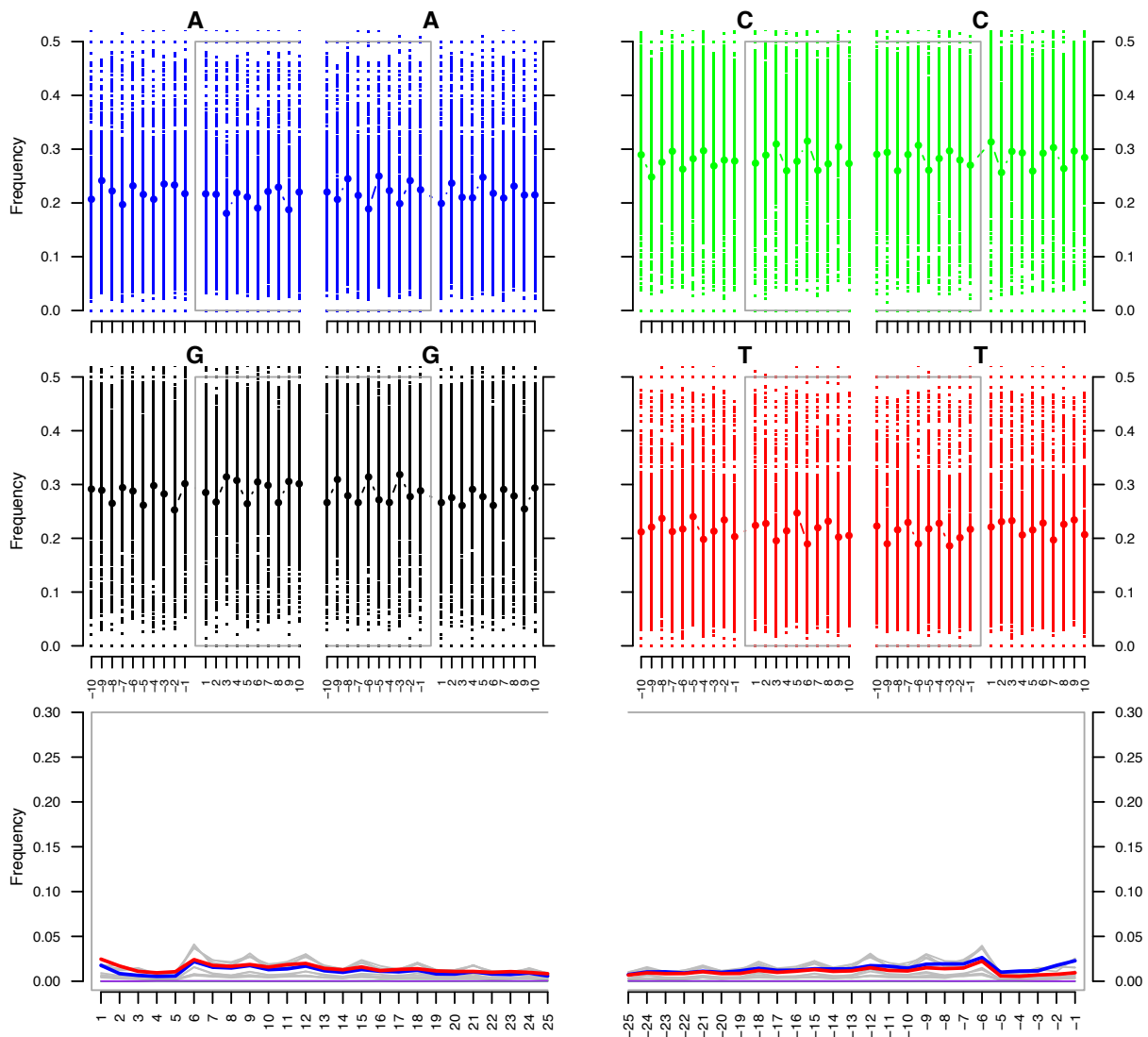


Figure 84: représentation des résultats de l'analyse des séquences identifiées comme bactériennes dans un échantillon ancien (kyste de Boulogne-Sur-Mer).

Dans ce cas, le résultat d'analyse ne présente aucune caractéristique de séquence ancienne. Les séquences amplifiées étaient donc modernes, et correspondaient à des bactéries présentes dans l'os au moment de l'extraction de l'ADN. Ces bactéries modernes utilisaient l'os comme une matrice de croissance, un support, dans le sol, mais n'avaient sans doute que très peu de points communs avec les bactéries ayant participé aux processus taphonomiques initiaux pendant les premières étapes de la dégradation post-mortem de l'organisme. Ce résultat est important car il suggère qu'*a fortiori*, il soit très peu probable de pouvoir identifier dans des os archéologiques des bactéries qui auraient été présentes dans les ossements avant le décès de l'individu étudié. Ceci est particulièrement important dans le cas des bactéries pathogènes, comme celle engendrant la lèpre, la tuberculose ou la brucellose ayant peut-être entraîné le décès de l'individu. Lors du décès, les conditions du milieu changent de manière très importante,

ce qui nuit à la viabilité des bactéries présentes dans les os. Elles vont donc soit mourir soit se mettre en dormance. En parallèle, les bactéries saprophytes colonisent les tissus, dont les os, de l'individu. Cette invasion va se traduire par une dilution des bactéries pathogènes du fait de la multiplication des bactéries invasives et de la dégradation des bactéries pathogènes et de leur ADN par ces mêmes bactéries. *In fine*, et peut-être rapidement, l'ADN des bactéries pathogènes sera presque complètement consommé et ne pourra plus être détecté, sauf cas exceptionnel. Nous avons étudié le devenir d'ADN exogène bactérien présent dans des dépouilles de souris que nous avons enfouies pour évaluer cette dégradation.

4.6. Recherche de l'ADN exogène

Les souris avaient été contaminées *ante mortem* par des bactéries non pathogènes, *Xanthomonas arboricola* comme simulant des bactéries Gram négatif et *Bacillus atrophaeus* comme simulant des bactéries Gram positif. A chaque prélèvement, il a été recherché par qPCR l'ADN de ces bactéries dans les os. La Figure 85 représente l'évolution de la quantité d'ADN de *B. atrophaeus* dans les différents tissus analysés. Il est à noter qu'à partir de 12 mois, la queue disparaît et ne peut plus être étudiée. La quantité d'ADN était faible dès J0 sauf dans la queue, qui est le site d'injection. Au cours des mois suivants, la quantité n'a plus varié dans les os, et la quantité dans la queue a atteint rapidement des concentrations comparables. *B. atrophaeus* présente la particularité de sporuler. La stabilité de concentration est peut-être due à la sporulation de ce bacille, qui s'est retrouvé dans un environnement défavorable à sa croissance. Une fois protégé dans la spore, l'ADN ne se dégrade plus et les concentrations restent stables.

	Femur	Machoire	Queue
J0	1,67 ^E -04	1,31 ^E -04	3,22
J+15jours	1,63 ^E -04	1,49 ^E -04	6,51 ^E -01
J+3mois	1,53 ^E -04	1,32 ^E -04	2,96 ^E -04
J+6mois	1,53 ^E -04	1,80 ^E -04	1,28 ^E -04
J+9mois	1,72 ^E -04	1,50 ^E -04	1,20 ^E -03
J+12mois	1,23 ^E -04	1,40 ^E -04	NA
J+18mois	1,35 ^E -04	1,31 ^E -04	NA

La quantité d'ADN est donnée en ng/ μ L d'extrait. Tous les échantillons ont été extraits dans 100 μ L de tampon.

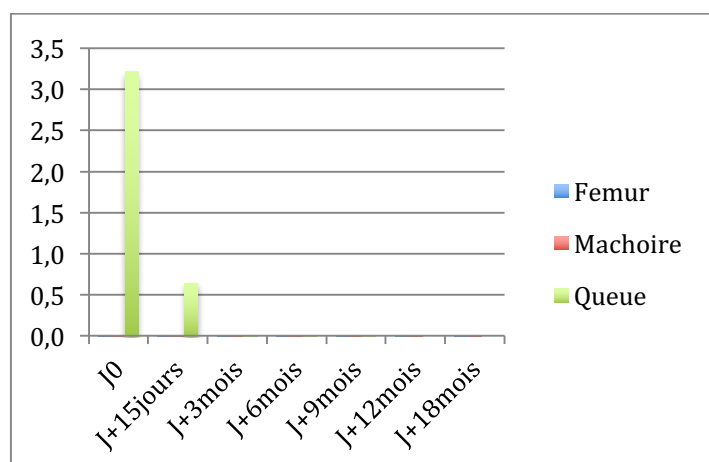


Figure 85 : évolution de la concentration en ADN de *B. atrophaeus* dans les différents tissus au cours du temps. Bleu : extrait de fémur, rouge : extrait de mâchoire, vert : extrait de queue.

Dans le cas de *X. arboricola*, qui n'a pas la capacité de sporuler, l'évolution a été comparable dans un premier temps (Figure 86). La concentration en ADN était particulièrement élevée dans la queue par rapport aux autres tissus, avec dix fois plus d'ADN de *X. arboricola* dans cet échantillon. Après 15 jours, la concentration a diminué d'un facteur 10 dans la queue, et il n'a pas été détecté d'ADN dans les os. Ensuite, ponctuellement, des traces d'ADN ont été mises en évidence dans les os, mais à des taux très faibles, qui pouvaient être liés à une amplification d'autres Xanthomonadales naturellement présentes dans le sol. Donc, il est probable que l'ADN ait entièrement disparu dans les six mois qui ont suivi l'enfouissement des souris.

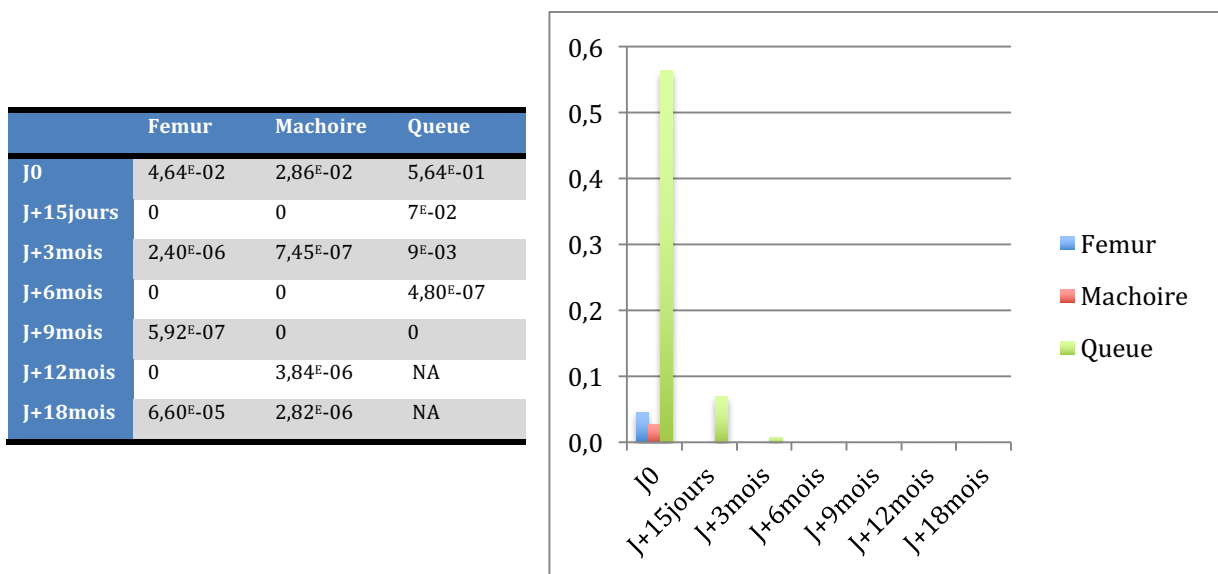


Figure 86 : évolution de la concentration en ADN de *X. arboricola* dans les différents tissus au cours du temps. Bleu : extrait de fémur, rouge : extrait de mâchoire, vert : extrait de queue.

4.7. Conclusions

Quelques conclusions peuvent être tirées des expérimentations décrites. En premier lieu, la mise en place des dépouilles dans le sol a provoqué une multiplication des bactéries cultivables, tant aérobies qu'anaérobies dans le sol au contact. Dès quinze jours d'enfouissement, la population bactérienne a crû, jusqu'à dépasser de deux ordres de grandeur la quantité du sol témoin. L'enrichissement a été transitoire pour les bactéries anaérobies, leur quantité étant redevenue comparable à celle du sol témoin après un an. Par contre, pour la communauté aérobie cultivable, il a fallu compter deux ans et demi avant de retrouver un niveau comparable. Et ce alors que les dépouilles étaient réduites après un an à l'état d'ossements. Il a alors été tenté de caractériser les

bactéries présentes, dans les os et dans le sol, afin de déterminer les mouvements de population. Néanmoins, le haut niveau de diversité des populations impliquées et le relatif manque de spécificité lié à l'amplification d'un unique marqueur hypervariable nécessitent l'emploi de tests statistiques élaborés, qui sont en cours de développement. Il a cependant été possible d'identifier de grandes tendances. Les populations sont modifiées par l'apport des dépouilles, avec une augmentation des phyla bactériens connus pour leur caractère saprophyte (Clostridiales, Burkholderiales, ...). Cette modification s'est déroulée en plusieurs étapes, avec la succession de différentes populations bactériennes, les Protéobactéries étant particulièrement actives pendant les phases d'autolyse et de putréfaction [156]. La prolifération bactérienne a commencé avant la rupture du corps et la libération de fluides riches en nutriments dans le sol. Cette prolifération a concouru aux phénomènes menant à cette rupture [155]. Cependant, les mécanismes en œuvre, pénétration transcutanée de bactéries ou échange de signaux entre communautés bactériennes endogène et environnementale, sont encore inconnus. Même après la disparition des chairs, les Protéobactéries et les Bacteroidetes sont restés présents, et il a fallu près de 18 mois pour que le sol retrouve un profil comparable au sol témoin et plus de deux ans pour que la quantité de bactéries cultivables aérobies dans le sol redescende à un niveau identique au sol témoin. Concernant la conservation de l'ADN exogène dans les os, il a été difficile de tirer des conclusions de par le faible taux de contamination initial. Ce taux était contraint par la capacité des souris à subir une injection non létale d'une suspension bactérienne, afin de permettre une fixation des bactéries dans les os et simuler une septicémie. Il en a résulté que l'ADN devient indétectable de manière fiable et univoque en moins de six mois dans les os, colonisés par les bactéries de l'environnement. Ceci explique sans doute le grand nombre d'échecs lorsque des pathogènes sont recherchés dans des ossements anciens. A l'heure actuelle, seules les dents, et en particulier la cavité pulpaire, ont donné des résultats solides, la majorité des publications ayant annoncé la détection d'ADN dans les os, essentiellement *M. tuberculosis*, ayant vu leur authenticité discutée.

Cela nous amène à la seconde partie expérimentale, portant sur les échantillons anciens.

ÉCHANTILLONS ANCIENS - RÉSULTATS ET DISCUSSION

Le projet initial portait sur la recherche de pathogènes bactériens dans des restes anciens animaux, avec en idée sous-jacente que la domestication, en favorisant les contacts étroits entre humains et animaux, avait permis le transfert de zoonoses entre ces deux groupes. En mettant en évidence des pathogènes sur des échantillons animaux anciens, il aurait été intéressant de comparer les séquences obtenues à celles trouvées chez l'humain contemporain de ces animaux, ou de comparer avec des séquences modernes.

Les résultats de l'étude taphonomique ont montré que les bactéries présentes dans les os disparaissaient très rapidement après le décès de l'animal infecté. Cette disparition peut être multifactorielle, mais il n'en demeure pas moins qu'après six mois, l'ADN devient indétectable. Les chances de trouver des traces de pathogènes dans ces échantillons animaux, sélectionnés sur des critères indirects (sépulture spécifique, absence d'équarrissage) eux-même discutés comme pertinents par les archéozoologues, étaient donc bien minces. Nous nous sommes donc concentrés pour ces échantillons sur une approche en deux étapes. Dans un premier temps, nous avons cherché à quantifier l'ADN endogène encore présent, considérant qu'il était nécessairement en quantité très supérieure à celle d'un éventuel pathogène. Lorsque l'extrait était suffisamment riche en ADN endogène pour permettre l'amplification d'une fraction de son ADN mitochondrial, nous avons séquençé l'amplicon afin de caractériser au mieux l'échantillon initial. Ensuite, toujours par qPCR, nous avons cherché à mettre en évidence l'ADN de bactéries pathogènes connues pour infecter l'espèce animale identifiée.

Dans un premier temps, nous avons exploré des voies de décontamination des réactifs qui soient complémentaires à celles décrites dans Champlot *et al.* [57] et dans Gorgé *et al.* [172]. En effet, la recherche d'ADN ancien, en particulier bactérien, impose de maîtriser au mieux les sources de contamination environnementales, les bactéries étant ubiquitaires.

1. PROTOCOLES ADDITIONNELS DE DECONTAMINATION DES REACTIFS

Pour améliorer la maîtrise de la contamination par l'ADN extérieur à nos échantillons, nous avons exploré deux pistes, l'une spécifiquement liée aux amorces, et l'autre visant tout ADN bicaténaire.

1.1. Décontamination des AMORCES

Les oligonucléotides sont parmi des réactifs les plus difficiles à décontaminer car constitués d'ADN, comme le contaminant. Au cours de ce projet, Nous avons analysé l'impact des modes de purification sur la présence d'ADN contaminants dans les amorces, tous les autres réactifs ayant été décontaminés selon les procédures les plus stringentes en vigueur au laboratoire [57]. Cette analyse a été réalisé en amplifiant le fragment V5 à l'aide de couples d'amorces ayant été purifiés par chacun des protocoles proposés par notre fournisseur, Eurofins.

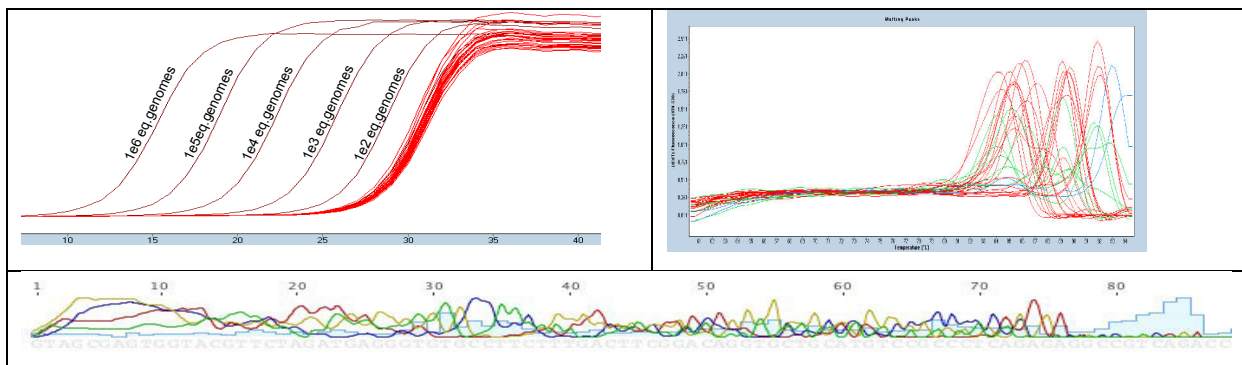


Figure 87 : courbes d'amplification et de fusion obtenues à partir d'amorces purifiées selon les différents protocoles d'Eurofins. Les produits d'amplification ont été séquencés selon la méthode Sanger.

Il est observé des produits d'amplification quel que soit le protocole de purification, et dans des quantités comparables, aux environs de 10 copies par puits de PCR, la limite de détection étant entre 10 et 100 équivalents-génomes par μL . L'analyse des courbes de fusion montre la présence de produits d'amplification présentant des températures de fusion différentes, et nous avons séquencé quelques produits de PCR. Le résultat du séquençage, présenté dans la partie inférieure de la Figure 87, montre un mélange de séquences, indiquant la présence de plusieurs fragments d'ADN bactérien contaminant dans les lots d'amorces testés. Il n'a donc pas été possible d'identifier les bactéries présentes à partir de ces mélanges de séquences.

La synthèse de ces oligonucléotides est réalisée de manière industrielle depuis des années (Figure 88). Succinctement, le premier nucléotide est lié de manière covalente à

un support solide par son extrémité 3' tandis que son extrémité 5' est protégée par un groupement chimique, le diméthoxytrityl (DMT). Ce groupement chimique est coupé par détritylation et un nouveau nucléotide (bloqué en 5' par un groupement DMT) est ajouté, afin de construire l'oligonucléotide. A la fin de la synthèse, le dernier groupement DMT est ôté et l'oligonucléotide synthétisé est détaché de son support.

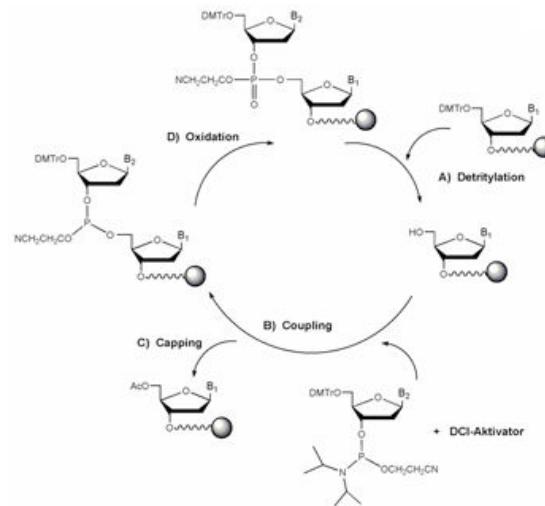


Figure 88 : procédure de synthèse des oligonucléotides selon la méthode dite Phosphoramidite [207]

Tout ce processus est automatisé, et il n'apparaît pas de source de contamination par de l'ADN bactérien au cours de ces étapes, purement chimiques. Par contre, il est possible que la contamination se produise lors des étapes de purification. Ces étapes de purification ont pour objet de filtrer les fragments d'oligonucléotides tronqués ou dépurinés. Les procédures de purification mises en œuvre par les manufacturiers sont suffisantes pour la majorité des usages mais des procédures spécifiques de purification ont été développées pour répondre à des besoins particuliers. Nous avons sélectionné les colonnes QSP clarify, de Phenomenex, pour réaliser notre purification. Ces colonnes fixent sélectivement les oligonucléotides tritylés (porteurs en 5' du groupe DMT) et éliminent les fragments non tritylés ou endommagés, ainsi que, dans notre objectif, l'ADN environnemental contaminant, qui n'est pas tritylé. Nous avons demandé à MWG Eurofins, l'un de nos manufacturiers, de nous fournir des précurseurs d'oligonucléotides, en détachant l'oligonucléotide synthétisé sans ôter le groupement DMT. Les oligonucléotides fournis ont été chargés sur les colonnes « QSP Clarify » et nous avons suivi le protocole du fabricant pour procéder à la dé-protection et à l'élution. Succinctement, après équilibrage des colonnes, les oligonucléotides ont été ajoutés à raison de 1 mL/min. Une fois la colonne vidée, nous l'avons rincée avec du tampon de

charge, afin d'assurer une élimination maximale de tout ADN non lié. La détritulation est réalisée dans la colonne par ajout d'acide dichloracétique à 2% et l'élution est réalisée par un tampon de NH_4HCO_3 20 mM dans de l'acétonitrile 25% (V/V). L'éluat est ensuite évaporé au SpeedVac/CryoVac avant de procéder à la remise en suspension des amorces dans du TE et à leur dosage par Nanodrop.

Nous avons ensuite réalisé les mêmes amplifications en utilisant ces amorces purifiées (Figure 89).

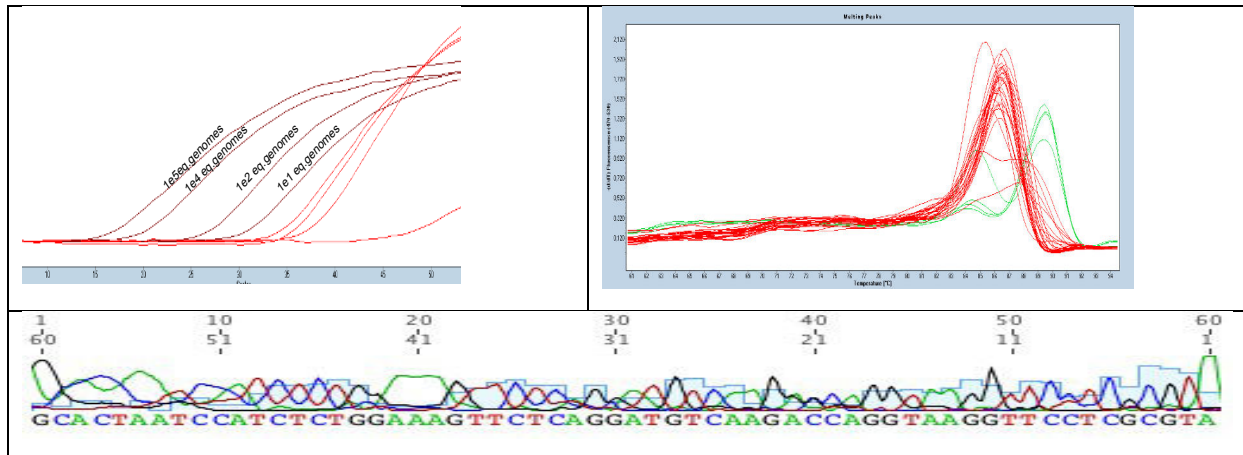


Figure 89 : courbes d'amplification et de fusion obtenues à partir du couple d'amorces purifiées ciblant le fragment V5. Le produit de séquençage obtenu est pur et correspond à *P. alcaligenes*.

L'emploi des amorces purifiées permet de faire baisser la limite de détection à moins de dix équivalents génome par μL , comme indiqué par les légendes sur les courbes d'amplification. Surtout, les courbes de diffusion présentent beaucoup moins de pics, laissant penser que les produits amplifiés sont plus homogènes. Cette impression est confirmée par le séquençage de ces produits d'amplification, avec des séquences non mélangées. Ces séquences ont été identifiées par BLAST contre la base GreenGenes [208] et les meilleurs scores sont obtenus avec *Pseudomonas spp*, plus spécifiquement *P. alcaligenes*, genre très répandu dans l'environnement, en particulier dans l'eau.

Ainsi, si nous n'avons pas pu éliminer totalement tout ADN bactérien. Hors travail en salle blanche, cela apparaît de toute façon compliqué, compte-tenu de la présence naturelle dans l'air de bactéries (environ 10^5 bactéries cultivables par m^3 [209]). Nous avons par contre réussi à montrer la présence de nombreux fragments bactériens dans les amorces commerciales, fragments apportés lors des étapes de purification. Il est possible de très fortement réduire cette contamination par l'emploi d'une méthode de décontamination relativement accessible.

1.2. Décontamination des réactifs par l'EMA et/ou le PMA

Nous utilisons au laboratoire un mastermix de PCR préparé sur place à partir de réactifs décontaminés [57]. Néanmoins, pour la préparation des banques de séquençage, nous avons dû développer de nouvelles stratégies pour éliminer autant que faire se peut les éventuelles molécules d'ADN présentes dans les différents réactifs et mélanges réactionnels fournis par les manufacturiers, essentiellement New England Biolabs. Nous avons exploré l'intérêt d'utiliser deux composés, l'EMA (Ethidium Mono Azide) et le PMA (Propidium Mono Azide). Ces deux composés, proches en structure du Bromure d'Ethidium (BET) sont des intercalants qui, une fois photoactivés, établissent des liens covalents avec la molécule d'ADN bicaténaire dans laquelle ils se sont insérés. La lumière transforme le groupe azide présent sur la molécule en un radical nitrène hautement réactif, capable de se lier à tout carbone à proximité. En conséquence, l'ADN est bloqué et ne peut plus être amplifié. La membrane cellulaire étant imperméable à ces molécules, elles ont initialement été proposés pour éliminer tout ADN libre dans un milieu avant d'extraire l'ADN cellulaire, l'amplification ne prenant en compte que les ADN contenus dans des cellules viables [210-212]. Nous nous sommes inspirés des travaux de Patel *et al.* [213] pour évaluer la capacité de ces deux molécules à bloquer les molécules d'ADN présentes dans les réactifs de qPCR et de préparation des banques de séquençage.

Ces molécules ont néanmoins l'inconvénient d'inhiber les Taq Polymérase avec un effet dose dépendant. Nous avons donc analysé les concentrations pertinentes pour bloquer les ADN contaminants sans perdre d'efficacité d'amplification.

Les premiers essais ont montré une perte complète d'efficacité du mastermix préparé au laboratoire, qui n'est pas compatible avec l'EMA et le PMA. Nous avons donc exploré les propriétés décontaminantes de ces molécules sur le mastermix commercial de Roche Diagnostics (LightCycler FastStart DNA Master^{PLUS}). Une phase de photoactivation de la molécule étant nécessaire, nous avons vérifié que les performances du mix n'étaient pas dégradées par cette étape, en variant le temps d'illumination entre 3 et 6 minutes, temps recommandés par le fournisseur. Les illuminations n'ont pas provoqué de décalage de Ct, montrant la stabilité de la Taq polymérase, mais aussi du SyBR Green à la lumière. Nous avons pu ensuite tester des concentrations croissantes d'intercalants, pour évaluer leur concentration d'emploi optimale.

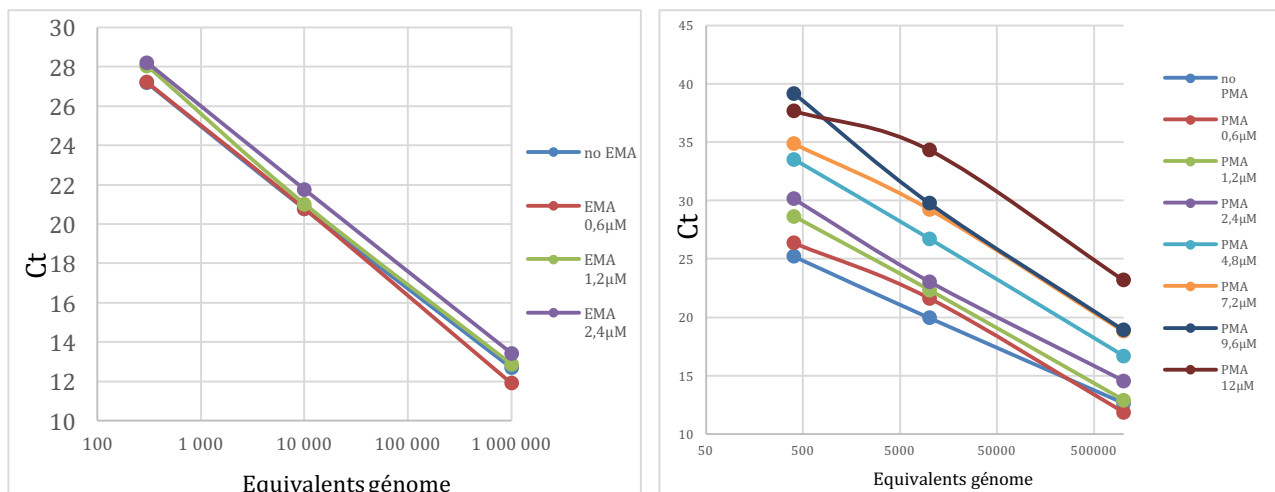


Figure 90 : effets de concentrations croissantes d'EMA (  gauche) et de PMA (  droite) sur l'efficacit  de l'amplification du fragment V5 de l'ADNr16S. Les  quivalents g nomes sont en  chelle logarithmique.

Aux concentrations  tudi es pour l'EMA, il est observ  un d calage vers des Ct plus  lev s   partir de 2,4 μM d'EMA dans le m lange r actionnel final. Ce m me d calage est observ  pour le PMA. Pour ce dernier, un d calage est apparu d s les plus faibles concentrations explor es, mais est devenu sensible au-del  de 1,8 μM. Le PMA est r put  plus actif, ce qui se confirme ici. Une fois d termin es les concentrations d'emploi, nous avons recherch  les effets de blocage des ADN contaminants dans des  chantillons t moins n gatifs Figure 91.

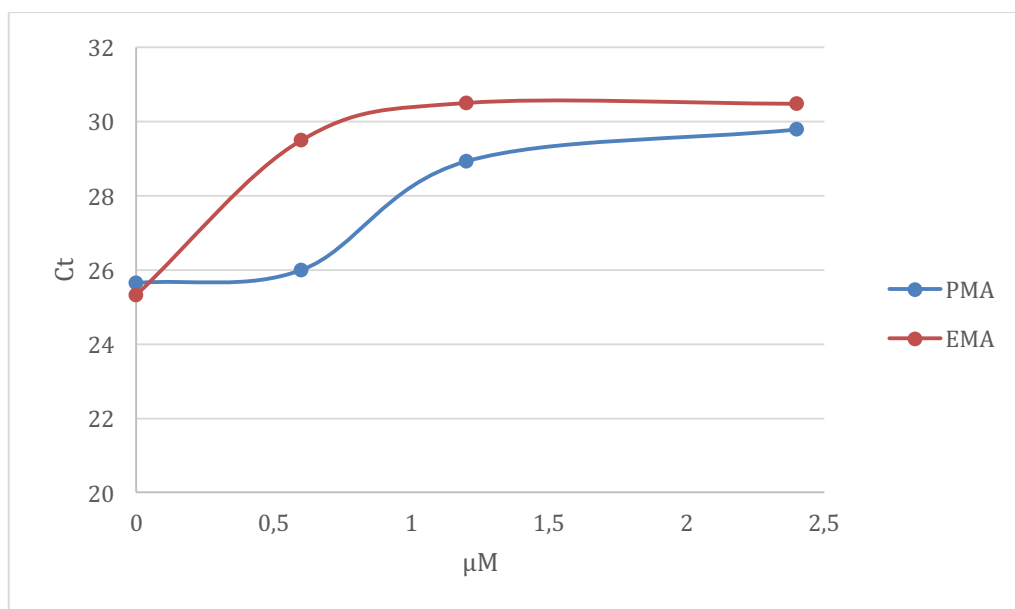


Figure 91 :  volution des Ct d'apparition des produits d'amplification sur des  chantillons n gatifs, en fonction de la concentration des deux intercalants.

Comme attendu, les t moins n gatifs pr sentent des produits d'amplification avec un d lai lorsqu'ils sont trait s par les intercalants EMA et PMA. L'EMA semble d velopper

une activité plus importante à faible concentration, tandis que le PMA devient réellement efficace à partir de 1,2 μM . L'efficacité et l'innocuité de ces intercalants à la concentration de 1,8 μM ont été validées sur les mélanges réactionnels du kit NEBNext (New England Biolabs). En conclusion, en employant ces intercalants, il est possible de faire baisser le bruit de fond de près de 6 Ct, soit une réduction de 99% de la quantité d'ADN environnemental contaminant présent. Une fois ces précautions supplémentaires mises en place, l'analyse des échantillons anciens a pu être lancée.

2. ÉCHANTILLONS ANIMAUX

2.1. Bure-Bourogne-Vellechevreux

Au vu du chantier de fouilles, l'archéozoologue a formulé l'hypothèse d'une inhumation sans équarrissage liée à une maladie infectieuse. La zone concernée est connue pour ses champs maudits, c'est-à-dire des endroits où les animaux s'infectent très régulièrement du fait de l'ingestion de spores de *Bacillus anthracis*. Il a donc été recherché dans un premier temps l'ADN endogène bovin, puis l'ADN de *B. anthracis*, puis de refaire pousser d'éventuelles spores présentes et la recherche de *M. tuberculosis*. La liste des échantillons et le résultat du séquençage (si positif) d'un fragment informatif de l'ADN mitochondrial sont indiqués dans le Tableau 15.

Deux échantillons, issus du site de Bure, n'ont pas donné de résultats en qPCR, et il a donc été impossible de déterminer l'espèce présente. La recherche de pathogènes n'a pas été effectuée non plus, considérant le ratio désavantageux entre ADN pathogène et ADN endogène.

La recherche d'ADN de *B. anthracis* s'est faite par qPCR, en utilisant des amorces et des sondes Taqman ciblant des facteurs de virulence (*capA* et *pagA*) portés par les plasmides pXO1 et pXO2, développées à DGA Maîtrise NRBC. Malgré leur haute sensibilité, il n'a pas été possible de mettre en évidence l'ADN de *B. anthracis*.

Site	Echantillon	Couple d'amorces (Ct)		sequence
		BB3r4m	BB3s4r	
Bourogne	0164	37,38		<i>Bos taurus</i>
	0168	35,88		<i>Bos taurus</i>
	0200	negative		
	0200 2 nd os	negative		
Bure	0890-1	37,91		<i>Bos taurus</i>
	0890-2	33,62		<i>Bos taurus</i>
	0890-3	35,16		<i>Bos taurus</i>
	1272	34,55		<i>Bos taurus</i>
Vellechevreux	26-001	negative	36,51	<i>Bos taurus</i>
	26-002	negative	42,47	<i>Bos taurus</i>
	26-003	37,21		<i>Bos taurus</i>
	130-010	46,95		<i>Bos taurus</i>
	130-011	négative	39,98	<i>Bos taurus</i>
	130-012	34,2		<i>Bos taurus</i>
	114-007	négative	35,56 / 36,1	<i>Bos taurus</i>
	114-008	38,24		<i>Bos taurus</i>
	114-009	négative	négative	
Birgels- gaerten	3325-394B	44,68		<i>Bison bonasus</i>
	3325-1C	négative	négative	
	3219-3	35,5		<i>Bison bonasus</i>
	3325-2D	33,2		<i>Bos taurus</i>
Les Sablons	549-1	négative	négative	
	549-3	39		<i>Bos taurus</i>
	650-2	39		<i>Bos taurus</i>
	709-1	55		<i>Bos taurus</i>
	888-2	36		<i>Bos taurus</i>
	888-3	38,5		<i>Bos taurus</i>
	914-1	39,5		<i>Bos taurus</i>
	917-2	48		<i>Bos taurus</i>
	946-1	36,5		<i>Bos taurus</i>
	1555-2	40,5		<i>Bos taurus</i>
	1555-3	négative	négative	
	1576-1	35		<i>Bos taurus</i>
	1576-3	négative	négative	
	1578-1	46		<i>Bos taurus</i>
1578-2	33,5		<i>Bos taurus</i>	
Mutterholtz	1019-7	négative		
	1156-39	négative		

Tableau 15 : liste des échantillons issus de Bure, Bourogne, Vellechevreux et Birgelsgaerten, avec le résultat de l'amplification par qPCR d'un fragment de l'ADN mitochondrial et l'assignation d'espèce déduite de cette séquence,

Les échantillons de sol et des écouvillons secs et ou humides (Cf. protocole en annexe 2) ont été mis à pousser sur gélose et en bouillon BHI (Figure 92) pour une remise en culture d'éventuelles spores encore présentes de *B. anthracis*.

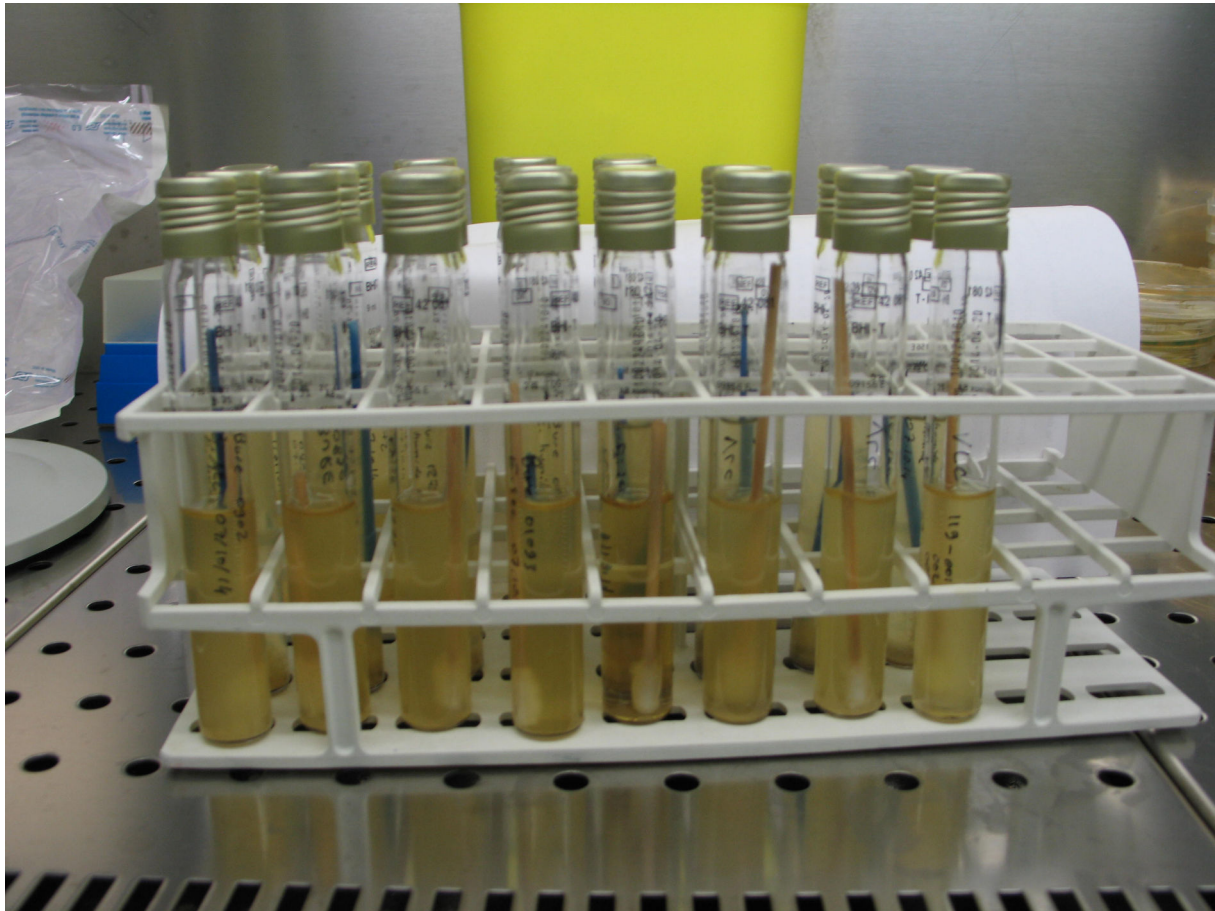


Figure 92 : écouillons secs et prélèvements de sédiments mis en culture au LSB3 de DGA Maîtrise NRBC.

Après une première étape de mise en culture sur milieu générique (TSA), une seconde étape de culture a été réalisée sur milieu sélectif Cereus Ident Agar. Seules les espèces du groupe *Bacillus cereus* poussent sur ce milieu. La majorité des espèces donne des colonies bleues mais *B. anthracis* donne des colonies blanches.

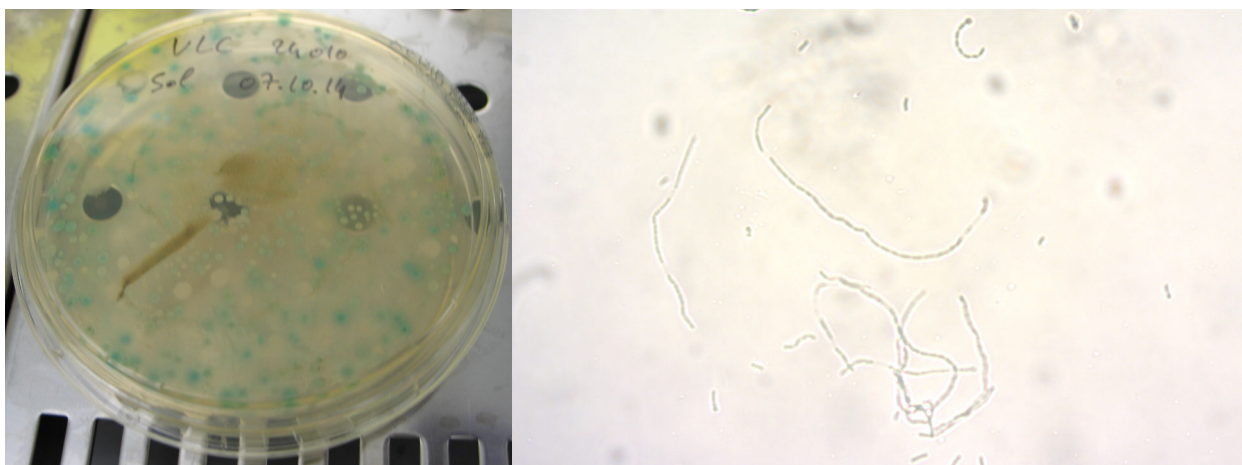


Figure 93 : sur le milieu sélectif (à gauche), il a été possible d'observer quelques colonies blanches sur cette culture issue d'un sédiment. L'examen à l'état frais d'une de ces colonies (à droite) a révélé des structures en chaînette.

Il a été observé dans plusieurs boîtes, tant issues d'écouvillons que de sédiments, des colonies blanches, susceptibles de correspondre à des colonies de *B. anthracis* (Figure 93). Un examen à l'état frais a été réalisé sur chacune de ces colonies et a révélé des structures en chaînettes, compatibles avec le groupe *B. cereus*, mais morphologiquement éloignées de celles formées par *B. anthracis*. Néanmoins, il a été procédé à la recherche par qPCR de *B. anthracis* par amplification des gènes de virulence décrits ci-dessus. Le résultat a été à chaque fois négatif, confirmant l'examen à l'état frais.

2.2. Ostheim Birgelsgaerten. Muttersholtz et Les Sablons

Parmi les ossements issus du site d'Ostheim, la discrimination entre bison et aurochs était difficile sur des critères morphologiques. Nous avons par conséquent procédé à une analyse génétique, visant une zone spécifique de l'ADN mitochondrial permettant de discriminer l'aurochs du bison.

Un prélèvement d'environ 0,2g a été réalisé sur chaque os et réduit en poudre selon le protocole standard du laboratoire, en maîtrisant les risques de contamination. L'ADN a ensuite été extrait et purifié. Une fois purifié, les échantillons ont été soumis à amplification, afin de chercher la présence de fragments d'ADN mitochondrial pouvant être amplifiés. Deux échantillons sur quatre ont donné un signal en qPCR. Après purification, les produits d'amplification ont été envoyés à une compagnie de séquençage. Le séquençage Sanger ayant réussi, les fragments ont été assemblés (logiciel Geneious Version 6) et les séquences ainsi obtenues comparées aux séquences de bison et de bœuf. Les séquences obtenues étaient comparables à la séquence publiée de *Bison bonasus* (6 SNPs, pas d'indels) et, pour le fragment amplifié, rigoureusement identiques entre les deux échantillons.

Les ossements de Mutterholtz ont été prélevés dans un ancien chenal, avec les sédiments environnants. Le long séjour dans cet environnement humide a modifié la structure osseuse, devenue spongieuse et marron, avec impossibilité de mettre en évidence un ADN endogène. L'analyse n'a pas été poussée plus loin pour ce site.

Les ossements des Sablons ont été identifiés par séquençage Sanger comme étant issus de *Bos taurus*, excepté trois ossements, correspondant à deux squelettes, qui n'ont pas donné de signal en qPCR. Pour les Sablons comme pour Ostheim, aucun pathogène n'a pu être détecté par qPCR.

2.3. Bernolsheim – Horbourg-Wihr

Après extraction et purification de l'ADN, les produits obtenus ont été soumis à amplification pour détecter l'ADN de cheval, en visant à amplifier un fragment de 107 paires de base. Des produits d'amplification ont été obtenus pour 5 extraits. Ces produits ont été séquencés et, pour trois d'entre eux, ont montré la présence d'ADN de cheval (*Equus caballus*) : 12040-1 ; 12040-2 et 12041-2. Ces trois extraits correspondaient à un squelette médiéval. Pour deux, les produits de séquençage ont montré une mauvaise qualité, empêchant une identification correcte. Une seconde amplification a été menée en cherchant à mettre en évidence une séquence plus courte (83bp). Des produits ont été obtenus sur les mêmes extraits que précédemment, mais en permettant de confirmer que les deux extraits douteux étaient bien du cheval : 12039-1 et 12041-1. Il n'a pas pu être mis en évidence d'ADN caractéristique pour les autres extraits, les produits d'amplification montrant la présence de multiples produits d'amplification, sans qu'il soit possible d'isoler les produits spécifiques. Ainsi, il est possible de conclure qu'au moins trois des animaux sont des chevaux. Les ossements issus des animaux identifiées 12043, 12080 et 6013.05 n'ont pas permis la mise en évidence d'ADN du genre *Equus* (Tableau 16).

site	échantillon	séquence
Bernolsheim	12080-1	
Bernolsheim	12080-2	
Horbourg-Wihr	6013.05-01	
Horbourg-Wihr	6013.05-02	
Bernolsheim	12040-1	<i>Equus caballus</i>
Bernolsheim	12040-2	<i>Equus caballus</i>
Bernolsheim	12043.2-1	
Bernolsheim	12043.2-2	
Bernolsheim	12043.2-3	
Bernolsheim	12041-1	<i>Equus caballus</i>
Bernolsheim	12041-2	<i>Equus caballus</i>
Bernolsheim	12041-3	
Bernolsheim	12039.1-1	<i>Equus caballus</i>
Bernolsheim	12039.1-2	

Tableau 16 : liste des échantillons analysés pour les sites de Horbourg et Bernolsheim. Un squelette peut être analysé à plusieurs reprises avec des os différents, voire des extraits différents.

Dans le but d'identifier la cause de la mort de ces animaux, une recherche préliminaire d'ADN de pathogènes a été menée. Ici encore, les produits d'amplification ont été multiples, et l'isolement des produits compatibles avec les fragments attendus n'a pas permis de mettre en évidence d'ADN de *M. tuberculosis*.

3. ECHANTILLONS HUMAINS

Devant les échecs de mise en évidence de pathogènes sur nos échantillons animaux, et le peu d'éléments indiquant une pathologie éventuelle, nous avons décidé de nous orienter vers des échantillons humains, pour lesquels nous disposons de plus d'éléments pouvant être associés à une pathologie, et pour lesquels le séquençage massif serait utile non seulement pour la recherche de pathogènes, mais également pour d'autres membres de l'équipe travaillant sur l'humain. Cette première série d'échantillons humains est principalement issue de quatre sites de fouilles, deux en Alsace, un dans les Pyrénées-Orientales et le dernier dans le Pas-de-Calais (Figure 94). Une seconde série, restreinte au tartre dentaire, est issue de marins du Mary-Rose.



Figure 94 : localisation des échantillons humains utilisés dans cette étude. Les restes humains anciens ont été excavés de quatre sites, Artzenheim (1) et Kolbsheim (2) sont en Alsace, Boulogne-sur-Mer (3) dans le Pas-de-Calais et Saint-Laurent de la Cabrerisse (4) dans les Pyrénées-Orientales. Carte dessinée avec mapcustomizer (mapcustomizer.com).

3.1. Analyse des kystes et restes associés

Les kystes et les ossements associés ont été comparés afin de rechercher leur teneur respective en ADN humain, et rechercher dans les kystes les éventuelles causes pathologiques de leur formation. Des banques de séquençage ont été préparées à partir de 5 individus issus de 4 sites de fouille en France. Après séquençage sur Illumina Miseq, les séquences obtenues ont été filtrées selon leur qualité et les adaptateurs ôtés, et les séquences restantes ont été analysées et comparées à la séquence humaine de référence (hg19) et à la séquence mitochondriale (rCRS).

Dans les os et le tartre, la proportion de séquences d'origine humaine était très faible, inférieure à 1%, ce qui est généralement décrit dans la littérature [214]. Pour les os pétreux, les dents, et surtout pour les kystes, la proportion d'ADN endogène a été élevée, voire très élevée, comme représenté sur la Figure 95. Il n'y a pas de corrélation entre taux dans les os et dans les os pétreux et les kystes.

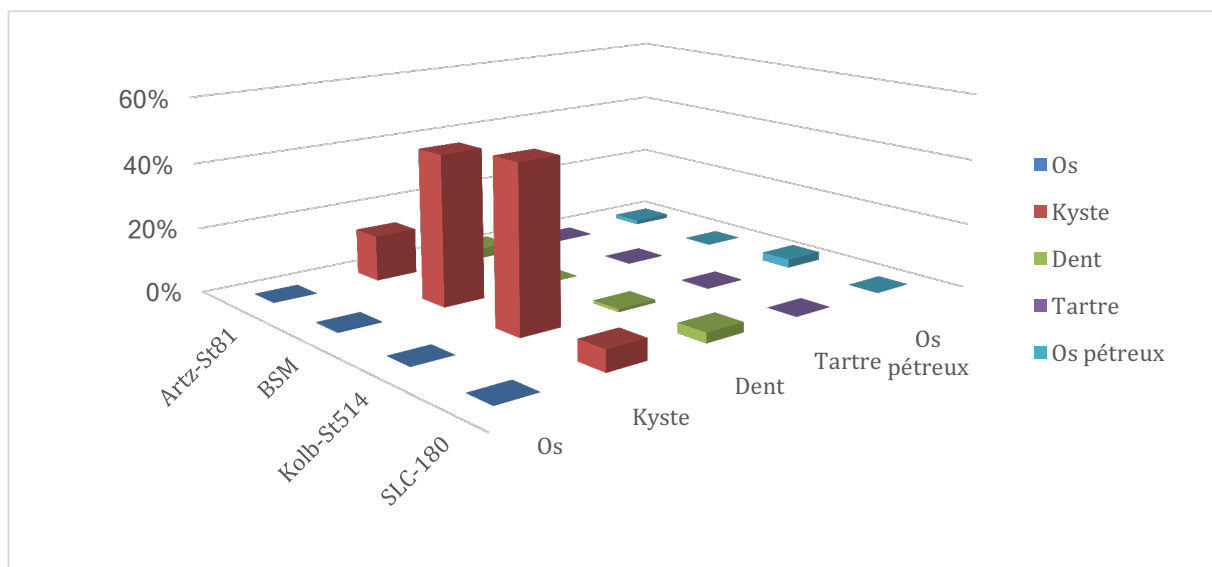


Figure 95 : distribution de la proportion de séquences identifiées comme d'origine humaine, selon l'origine tissulaire de la banque dans chaque échantillon. Les os pétreux étaient absents des assemblages pour BSM et SLC-180.

Une publication récente [214] a étudié la proportion d'ADN endogène dans différents os et a proposé l'os pétreux comme l'os avec le meilleur taux d'ADN endogène. Dans notre étude, il a également, pour un des deux échantillons pour lesquels nous avons l'os pétreux, présenté une teneur en ADN endogène plus riche que la dent et que l'os, et pour le second intermédiaire entre l'os (pauvre) et la dent (plus riche), ce qui confortait les résultats de l'article, mais l'os pétreux est apparu bien moins riche que les kystes, pour tous les échantillons analysés. Retrouver plus de 10% d'ADN endogène dans des os

mérovingiens est tout à fait exceptionnel. Retrouver un tel taux sur quatre échantillons différents, répartis sur quatre sites répartis sur la France est un bon indicateur de la valeur de ce type d'échantillons lorsqu'on recherche de l'ADN endogène.

Du fait du nombre élevé de séquences identifiées comme d'origine humaine, il a été possible de déterminer le sexe des individus en comparant les fréquences relatives des chromosomes sexuels et des autosomes. Quel que soit l'échantillon en effet, il y a en permanence quelques séquences qui s'alignent avec le chromosome Y, soit du fait d'une contamination à une étape du traitement de l'échantillon (depuis l'excavation jusqu'à la préparation de banques). La présence de séquences pour le chromosome Y n'est donc pas indicative par elle-même. Par contre, la proportion de séquences alignées, rapportées à la longueur de chaque chromosome, doit être comparable pour les autosomes. En cas de sexe féminin, le nombre de séquences s'alignant sur le chromosome X suit une proportion comparable à celle des autosomes, tandis que dans le cas du sexe masculin, les proportions sont réduites de moitié [215].

Cette interprétation est entrée en conflit dans un cas avec l'interprétation de l'archéologue, qui avait identifié l'individu de Boulogne-sur-Mer comme étant de sexe masculin. Dans le cas présent, l'archéologue a employé la méthode morphoscopique fondée sur la conformation de diverses régions d'intérêt de l'os coxal ainsi qu'une méthode morphoscopique permettant de calculer la probabilité à posteriori d'appartenance à l'un ou l'autre sexe. Les deux méthodes ont une fiabilité supérieure à 95% hors l'échantillon de référence ayant servi à leur élaboration. L'attribution du sexe proposé ici par l'archéologue est moins assurée car elle reposait en premier lieu sur la méthode morphoscopique appliquée lors du prélèvement des ossements sur le terrain. L'application ultérieure de la méthode morphométrique a donné comme résultat une probabilité *a posteriori* d'être un homme de 93%, probabilité forte mais inférieure au seuil généralement considéré comme suffisant pour retenir une attribution définitive (communication personnelle de Sacha Kacki).

Sample

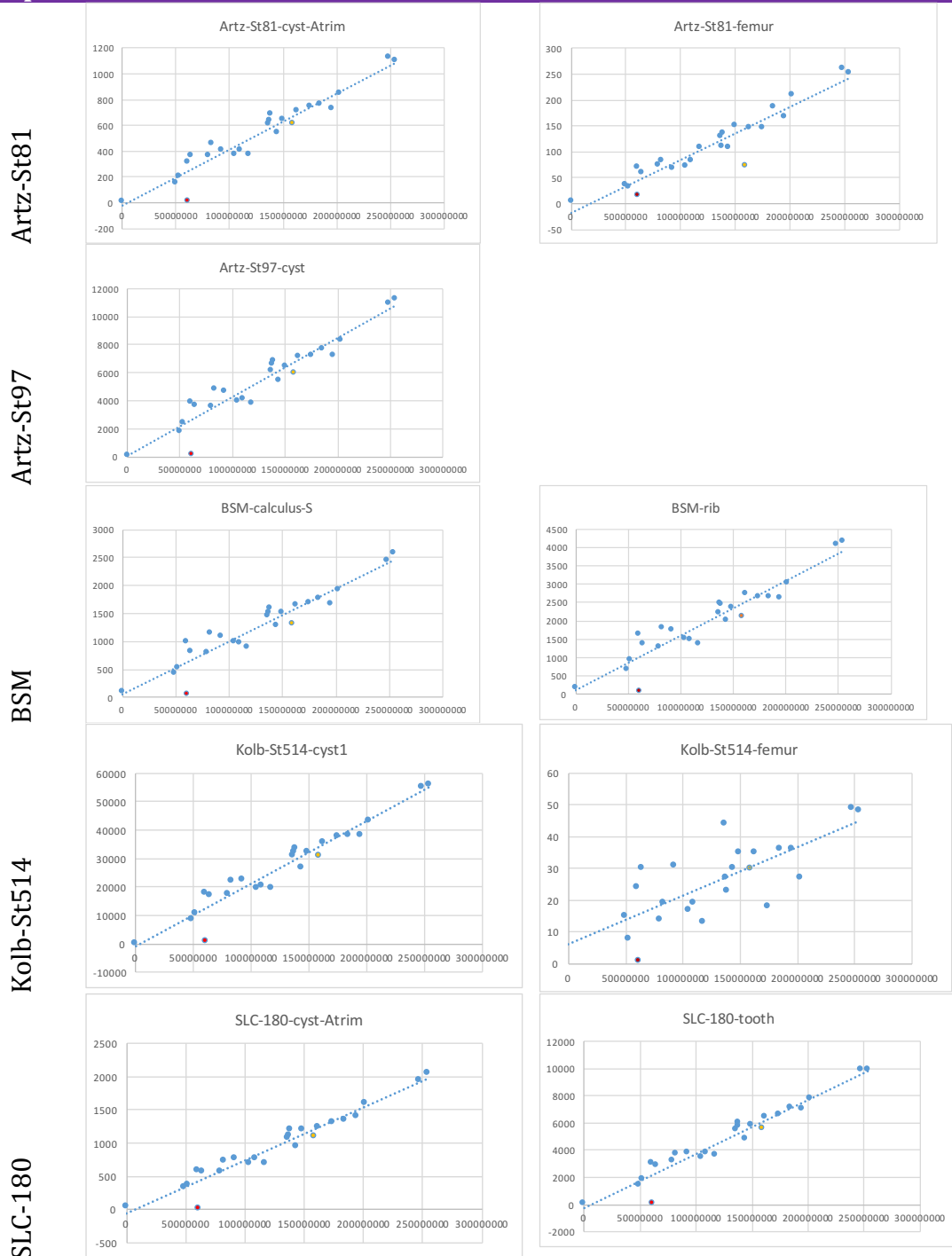


Figure 96: distribution des quantités de séquences alignées sur les chromosomes selon la longueur de ceux-ci en paires de base. La ligne en pointillé indique la corrélation linéaire de ces points (droite de tendance). Le ratio pour le chromosome X a été indiqué en orange et celui du chromosome Y en rouge. Pour tous les échantillons sauf Artz-ST81, le ratio du chromosome X était proche ou très proche de la droite de tendance, indiquant la présence de deux chromosomes X et donc que l'individu était de sexe féminin. Pour Artz-ST81, le ratio est deux fois inférieur à la moyenne, pour le chromosome X et pour le chromosome Y, indiquant la présence d'un seul exemplaire de chaque chromosome, et donc que l'individu était de sexe mâle. On note que pour tous les autres échantillons, le ratio pour le chromosome Y était très proche de 0, indicatif d'un sexe féminin.

Il a été également possible de reconstruire le génome mitochondrial dans trois cas sur cinq, avec des couvertures allant de 96x pour BSM (haplogroupe U4a1a) à 10x pour Kolb-St514 (T2c1d2) et 0,5x pour SLC-180 (HV9). Le premier haplogroupe est surtout retrouvé en Scandinavie, en Allemagne et en Pologne. Actuellement, la fréquence de cet haplogroupe dans la région de Boulogne-sur-Mer est de 2 à 3% (http://www.eupedia.com/europe/Haplogroup_U4_mtDNA.shtml) tandis que l'haplogroupe T2c1d2 est plus fréquemment retrouvé en Italie, en Espagne et en Iran (http://www.eupedia.com/europe/Haplogroup_T_mtDNA.shtml).

Nous avons ensuite analysé par *MetaPhlan* [199] les populations bactériennes présentes dans les différents échantillons (Figure 97).

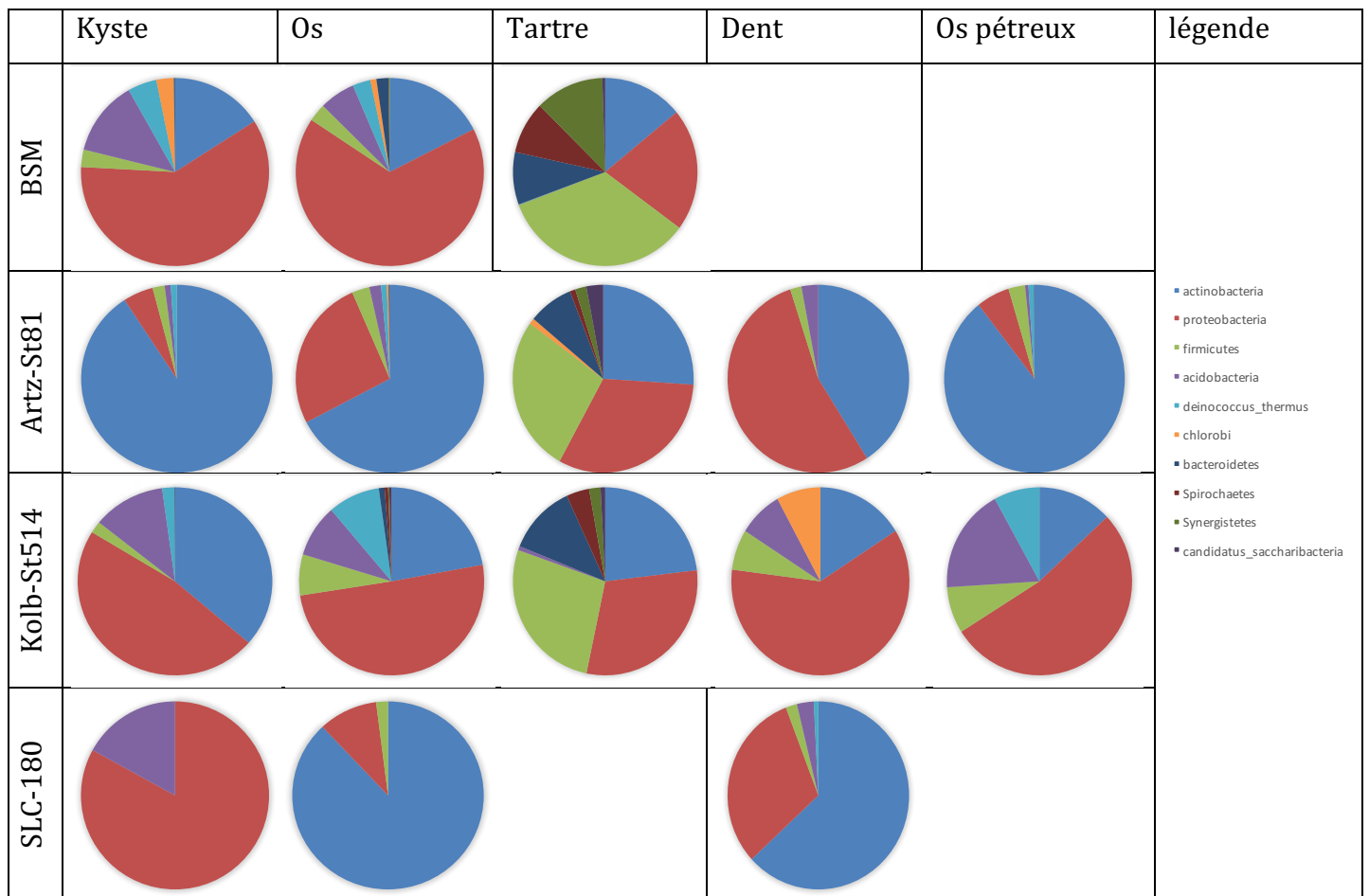


Figure 97: Distribution des phyla bactériens dans les différentes parties des échantillons soumises à analyse, et visualisée par l'algorithme Krona tool [201]. La légende est identique pour l'ensemble des graphiques, permettant une comparaison entre les différents échantillons. Nous ne disposons pas d'os pétreux pour les échantillons de SLC et de BSM, et seulement du tartre pour ce dernier. La dent de l'échantillon SLC-180 avait été analysée avant notre intérêt pour le tartre. Il n'a pu être analysé rétrospectivement.

Il est apparu avec cette représentation que les populations bactériennes étaient différentes selon les différentes parties d'un même échantillon. Pour les kystes, du fait de leur faible épaisseur, leur surface n'avait pas été nettoyée avant extraction de l'ADN, et les profils bactériens obtenus ont reflété le sol dans lequel ils étaient enfouis. Il en est de même pour les os, qui, comme démontré dans la partie consacrée à la taphonomie, ont un profil bactérien identique au sol dans lequel ils étaient enfouis. Par contre, le profil bactérien obtenu à partir des tartres a été comparable pour les trois échantillons, avec une distribution par phylum plus régulière, quatre phyla (Actinobactéries, Protéobactéries, Firmicutes et Deinococcus) comptant pour au moins 80% des séquences bactériennes identifiées. Cette similarité de distribution sur des échantillons divers par ailleurs mérite d'être explorée plus avant. Il est possible avec *MetaPhlan* de descendre à des niveaux taxonomiques plus profonds, jusqu'au niveau de l'espèce. Nous avons ici analysé la diversité bactérienne au niveau du genre bactérien, en représentant les résultats sous forme de carte de chaleur (*heatmap*) (Figure 98).

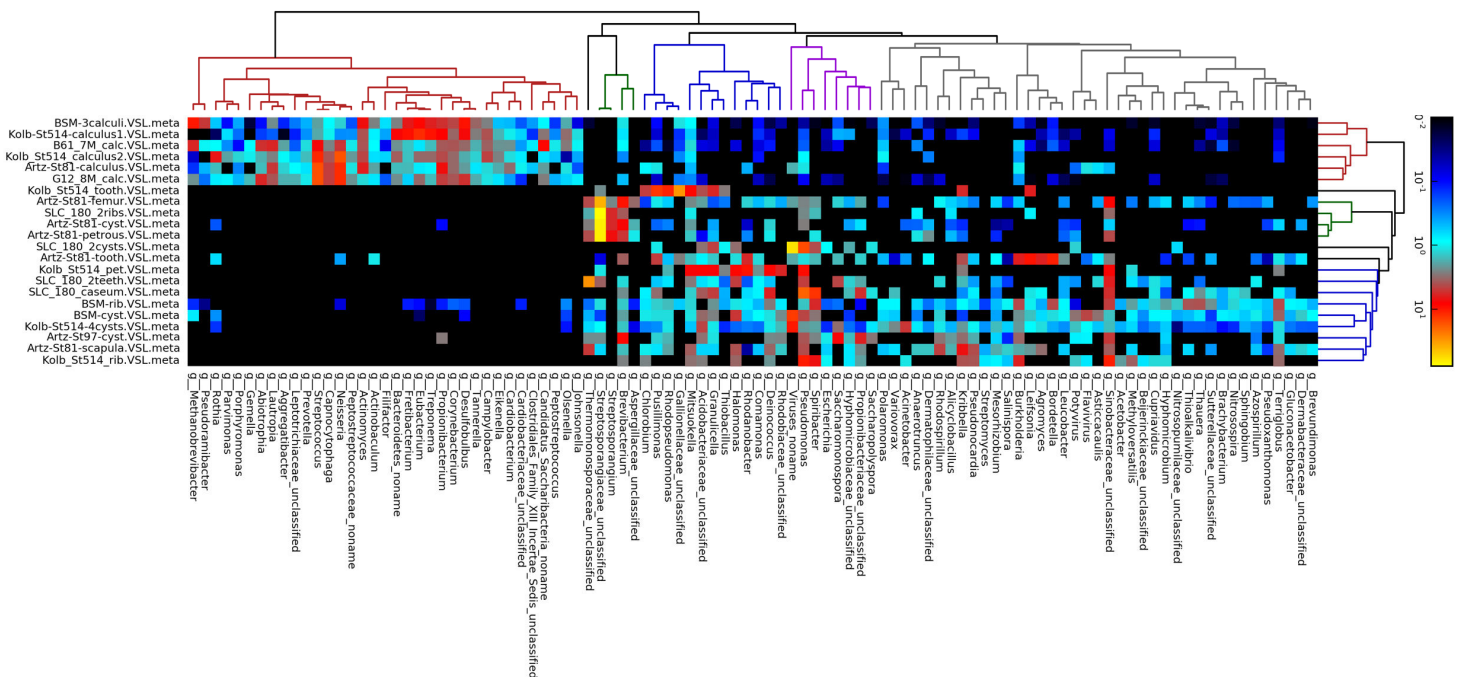


Figure 98: Heatmap obtenue grâce à *MetaPhlan* (mode Very Sensitive Local pour Bowtie2) [199] représentant les 100 genres bactériens les plus représentés au sein des échantillons analysés. Dans la partie supérieure gauche de la *heatmap*, il est observé un cluster de genres bactériens communément présents dans la communauté bactérienne buccale, en bonne santé ou à l'hygiène déficiente. Ce cluster correspondait aux échantillons de tartre analysés ici. Les kystes et les os présentaient les mêmes genres bactériens, beaucoup plus divers, liés à la grande diversité de bactéries présentes dans les sols.

Il est apparu dans la Figure 98 que nos échantillons se répartissaient en deux grands clusters. Le premier, présent dans la partie haute de la carte, et identifié par la partie

rouge du dendrogramme, regroupait tous les tartres analysés, quel que soit l'échantillon de départ. Cette population était homogène, et regroupait des genres bactériens absents des autres échantillons. On a vu lors de l'analyse taphonomique que les os tendaient à présenter un profil de population bactérienne comparable au sol dans lesquels ils ont été enfouis. Les os et les kystes analysés ici présentaient des profils variés et riches en genres telluriques ou ubiquitaires. Les profils des tartres avaient très peu de genres bactériens communs avec les sols et les kystes issus des mêmes individus. Par contre, ils étaient riches en genres bactériens naturellement présents dans la communauté bactérienne de la bouche. Il existait cependant des différences de richesse en représentation de 105 genres bactériens, pouvant être liées à l'état sanitaire de la bouche.

Il est ressorti de ces analyses que les tissus calcifiés avaient une teneur élevée en ADN endogène. Les kystes, formés par la minéralisation de tissus fibrosés, étaient riches en ADN humain, ayant préservé celui-ci de la digestion microbienne qui intervient dans tout tissu poreux, dont l'os à l'exception partielle de l'os pétreux, beaucoup plus dense et moins perméable. Le tartre présentait une teneur élevée en ADN bactérien spécifique de la communauté bactérienne buccale. La calcification de la plaque dentaire en tartre a provoqué une fossilisation des bactéries participant à l'établissement de la plaque dentaire et a protégé celles-ci de la dégradation biologique.

Les os communément analysés, côtes ou os longs, avaient majoritairement une teneur très pauvre en ADN endogène et étaient riches en ADN bactérien dont le profil ressemble à la communauté tellurique, les os servant de matrice ou de support à des communautés bactériennes proliférant naturellement dans le sol. Nous avons déjà discuté de l'origine moderne des bactéries présentes dans les ossements anciens en 4.5.

3.2. Analyse des Tartres des marins du Mary-Rose

Du fait de leur très petite taille, les 28 tartres n'ont pu être pesés, mais simplement estimés à quelques milligrammes. Ils ont été dissous dans l'EDTA 0,5M additionnée de NLS 1% (P/V), de β mercapto-éthanol 1% et de protéinase K à 200 μ g/ml. Les banques ont été préparées selon le protocole décrit dans Gorgé *et al.* [172] en modifiant la première étape de purification pour utiliser la purification sur plaque du kit Nucleospin96 (Macherey-Nagel).

A l'issue des préparations de banques, le séquençage a été réalisé sur Illumina MiSeq pour MR2017 et MR2087 (cartouche V3 2x150bp) et sur NextSeq pour l'ensemble des banques, en utilisant une cartouche HT 2x150bp. Le Tableau 17 indique les résultats obtenus.

samples	raw-reads		cutadapt			leehom (merging of paired reads)			
	reads	GC	reads	GC	% kept	reads	GC	mean length	% kept
MOCKbank	685	45,41%	629	46,04%	91,8%	434	47,46%	80,9654	63,36%
MOCKpurif	688	45,04%	552	44,39%	80,2%	393	43,43%	90,3588	57,12%
MR2003	14025622	41,49%	13493710	41,36%	96,2%	11467540	41,62%	88,1849	81,76%
MR2004	9599853	42,41%	9460786	42,43%	98,6%	7735926	41,97%	91,9229	80,58%
MR2006	12374503	42,54%	10308200	41,22%	83,3%	9095350	40,86%	79,9045	73,50%
MR2013	7334761	38,97%	7250832	38,89%	98,9%	5432600	38,52%	97,5498	74,07%
MR2015	10436046	39,78%	9939615	39,00%	95,2%	8592459	38,62%	86,8072	82,33%
MR2017	3914744	55,28%	3559383	57,27%	90,9%	3483818	57,71%	51,688	88,99%
MR2020	11030989	39,13%	10460231	38,65%	94,8%	8335379	38,46%	94,2108	75,56%
MR2023	11098767	39,15%	10894363	38,85%	98,2%	9143285	38,79%	89,9251	82,38%
MR2024	8987893	37,87%	8859059	37,55%	98,6%	7004271	37,19%	92,3926	77,93%
MR2025	8081289	39,34%	7919680	39,04%	98,0%	6080972	38,58%	94,8713	75,25%
MR2028	11039924	38,60%	10995201	38,32%	99,6%	9033109	38,03%	91,018	81,82%
MR2029	15332480	37,92%	14904779	37,56%	97,2%	8825141	39,28%	90,6806	57,56%
MR2037	1303843	35,75%	1295931	35,74%	99,4%	561231	36,55%	104,177	43,04%
MR2040	6706843	44,18%	4610136	43,50%	68,7%	3883861	43,57%	84,8819	57,91%
MR2042	10788057	46,08%	6576009	46,37%	61,0%	5797281	46,49%	78,6311	53,74%
MR2045	12570288	44,10%	8388979	43,34%	66,7%	7196732	43,50%	85,4538	57,25%
MR2046	9980786	47,57%	1736298	46,43%	17,4%	950801	44,67%	74,752	9,53%
MR2049	12813082	41,64%	12603160	41,26%	98,4%	11335450	41,30%	84,0558	88,47%
MR2051	10283897	42,51%	7401292	41,17%	72,0%	6028076	41,17%	90,4885	58,62%
MR2057	9888195	43,32%	8821250	43,19%	89,2%	6841326	43,20%	92,7917	69,19%
MR2068	22057167	41,73%	21272889	41,47%	96,4%	16936423	41,44%	91,2407	76,78%
MR2074	7668718	42,57%	6188301	41,21%	80,7%	5074055	41,12%	89,2145	66,17%
MR2079	25429807	41,61%	24568837	41,15%	96,6%	20005113	41,12%	88,7616	78,67%
MR2087	5785473	52,70%	5294791	54,08%	91,5%	5142546	54,67%	56,1763	88,89%
MR2107	21385020	43,65%	20898490	43,66%	97,7%	16960522	43,68%	89,4577	79,31%
MR2108	903495	45,47%	873778	45,58%	96,7%	660580	44,88%	92,5699	73,11%
MR2111	13314089	45,76%	10578301	46,12%	79,5%	8607212	46,53%	89,5673	64,65%
MR2116	12772288	51,96%	11489104	53,52%	90,0%	10706252	53,76%	70,481	83,82%
MRMock	25580	37,51%	7543	53,87%	29,5%	5501	60,15%	56,2903	21,51%

Tableau 17 : synthèse des données obtenues lors du séquençage des tartres du Mary-Rose, et effet des processus de filtrage des données sur la quantité finale de séquences analysées,

Trois blancs ont été réalisés lors des deux séquençages ayant permis de produire des séquences présentées dans ce tableau. Le nombre de séquences observées dans ces

échantillons a été très faible, bien inférieur au nombre de séquences produites par l'échantillon le plus pauvre en ADN. Cela montrait en première analyse que la contamination était maîtrisée et que les index ont bien été lus lors du démultiplexage. Concernant les échantillons vrais, le nombre de séquences allait de 903.495 à 25.429.807, avec une moyenne de 10.960.997 séquences ($\pm 5.498.653$). Les différentes étapes de filtrage des séquences ont entraîné une perte moyenne de 30% des séquences, avec un maximum de 91% de séquences filtrées et un minimum de 11%. Il n'a pas pu être étudié de corrélation entre masse initiale de l'échantillon, quantité d'ADN extraite et quantité de séquences obtenues par échantillons.

Après filtrage des séquences, celles-ci ont été comparées au génome humain, pour rechercher les fragments pouvant être identifiés. Le tartre est relativement pauvre en ADN humain, mais il a déjà été décrit qu'il était possible de déterminer des haplogroupes par capture [216]. Il a aussi été possible de calculer le nombre de séquences identifiables par chromosome et, à partir du tableau obtenu, de vérifier si tous les marins étaient de sexe masculin, ce qui est attendu pour l'époque sur un navire de guerre engagé dans une bataille. Ainsi, les tartres contenaient de l'ADN humain, avec entre 0,05% et 4,44% des séquences obtenues d'origine humaine (moyenne 0,74%) et en calculant les ratios nombre de séquences/taille du chromosome, il a été montré que les chromosomes X et Y ont un ratio toujours inférieur aux chromosomes, indiquant le sexe masculin des individus.

samples	Human mapped reads		Nb reads per chromosome																									
	qty	%	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	19	22	21		Y	M	
MR2003	22866	0,20%	1450	1533	1286	1246	1216	1069	1027	956	729	917	835	817	651	584	508	483	403	539	377	321	189	240	532	83	164	
MR2004	26822	0,35%	1737	1908	1600	1507	1342	1363	1166	1187	898	1151	1024	1046	748	661	653	582	530	586	463	333	225	271	626	90	13	
MR2006	404111	4,44%	26674	28346	23412	22439	21196	20128	17687	17314	12968	16310	15567	15245	11057	10236	9316	8580	7741	9040	6893	5064	3133	4136	9104	1398	36	
MR2013	56933	1,05%	3751	4245	3527	3545	3228	3048	2690	2581	1883	2225	2184	2245	1657	1546	1283	1110	1008	1359	861	613	347	634	1291	208	15	
MR2015	79859	0,93%	5157	5690	4912	4853	4479	4223	3659	3451	2581	3147	3014	3062	2273	2010	1736	1559	1414	1888	1138	916	492	846	1798	247	117	
MR2017	1695	0,05%	253	250	179	155	164	155	161	143	134	142	142	114	88	90	98	105	111	72	68	91	55	26	65	8	1	
MR2020	157150	1,89%	10691	11576	9738	9612	8663	8143	7247	6936	5079	6414	5967	6234	4685	4194	3612	3026	2954	3702	2487	1801	1071	1647	3674	544	22	
MR2023	35959	0,39%	2443	2559	2193	2073	1928	1885	1724	1565	1149	1425	1418	1461	1007	922	814	723	669	812	551	428	238	384	878	110	10	
MR2024	16666	0,24%	1072	1139	1017	1064	905	862	754	749	523	680	637	640	487	454	351	323	283	399	288	170	136	161	406	47	11	
MR2025	29174	0,48%	1941	2005	1728	1719	1588	1479	1301	1295	904	1211	1170	1104	846	700	630	548	510	723	454	296	193	303	767	86	16	
MR2028	19159	0,21%	1270	1392	1171	1152	1028	1024	878	818	576	738	717	704	521	505	429	349	366	423	308	172	126	205	485	67	34	
MR2029	28653	0,32%	1959	1990	1611	1659	1468	1451	1324	1143	860	1125	1106	1101	759	695	664	653	592	612	463	382	229	277	634	94	8	
MR2037	289	0,05%	16	22	19	23	18	19	6	7	4	10	9	18	8	7	7	5	4	7	5	4	0	4	5		0	
MR2040	9719	0,25%	679	683	523	510	491	454	462	403	330	410	371	370	243	225	215	232	214	202	215	155	91	99	230	32	10	
MR2042	59548	1,03%	4226	4052	3355	2983	2969	2874	2721	2416	1961	2436	2384	2241	1472	1535	1460	1509	1426	1361	1203	1070	720	641	1236	216	19	
MR2045	21898	0,30%	1455	1475	1278	1212	1076	1062	971	896	745	889	857	782	628	608	549	488	486	450	436	339	224	214	526	55	12	
MR2046	41829	4,40%	2864	2881	2301	2281	2132	2071	1776	1641	1297	1680	1588	1590	1064	966	970	928	856	927	711	576	340	414	951	166	34	
MR2049	5878	0,05%	374	373	277	297	292	258	236	270	166	231	195	226	143	129	124	118	93	112	91	93	42	68	162	14	6	
MR2051	12430	0,21%	799	896	688	669	665	642	550	500	401	537	463	462	323	328	262	265	272	290	220	157	117	147	311	42	14	
MR2057	13613	0,20%	955	912	713	776	697	705	644	541	449	548	502	499	379	314	304	302	279	302	253	186	92	138	345	49	12	
MR2068	119102	0,70%	7801	8158	6929	6450	6074	5758	5278	5036	3774	4629	4550	4544	3248	2937	2730	2623	2426	2732	1934	1679	956	1238	2672	381	32	
MR2074	8133	0,16%	543	505	435	430	405	396	350	318	229	339	333	291	234	156	212	179	148	181	149	116	75	82	179	25	18	
MR2079	31150	0,16%	1944	2115	1719	1761	1574	1493	1322	1336	955	1228	1165	1141	825	774	736	674	620	692	527	425	228	320	686	89	37	
MR2087	39992	0,78%	2795	2729	2215	1972	2011	1814	1777	1710	1322	1656	1584	1525	983	1030	984	1084	1142	867	861	847	536	462	835	110	5	
MR2107	117878	0,70%	8080	8111	6742	6376	6030	5700	5303	4989	3877	4784	4655	4534	3222	3006	2749	2839	2675	2626	2206	1905	1215	1214	2459	398	24	
MR2108	1767	0,27%	133	124	114	96	87	78	91	87	61	66	69	75	47	51	39	43	27	41	26	24	16	18	23	5	2	
MR2111	21467	0,25%	1478	1439	1133	1021	1037	978	1003	849	704	867	859	789	567	545	547	556	565	452	439	408	248	193	451	60	18	
MR2116	67671	0,63%	4796	4436	3511	2883	3072	2947	2906	2601	2226	2776	2675	2514	1561	1812	1681	1942	1999	1387	1469	1553	997	657	1329	178	31	

Tableau 18 : synthèse des valeurs obtenues à partir de l’alignement des séquences sur le génome humain de référence. Trois échantillons présentent un très faible nombre de séquences humaines, et un pourcentage très bas (MR2017, MR2037 et MR2049). Le grand nombre de séquences obtenues compense cette rareté, sauf pour MR2037 pour lequel il n’est obtenu que 289 séquences humaines. Le plus fort taux est de 4,44% et le taux moyen de séquences humaines 0,74%).

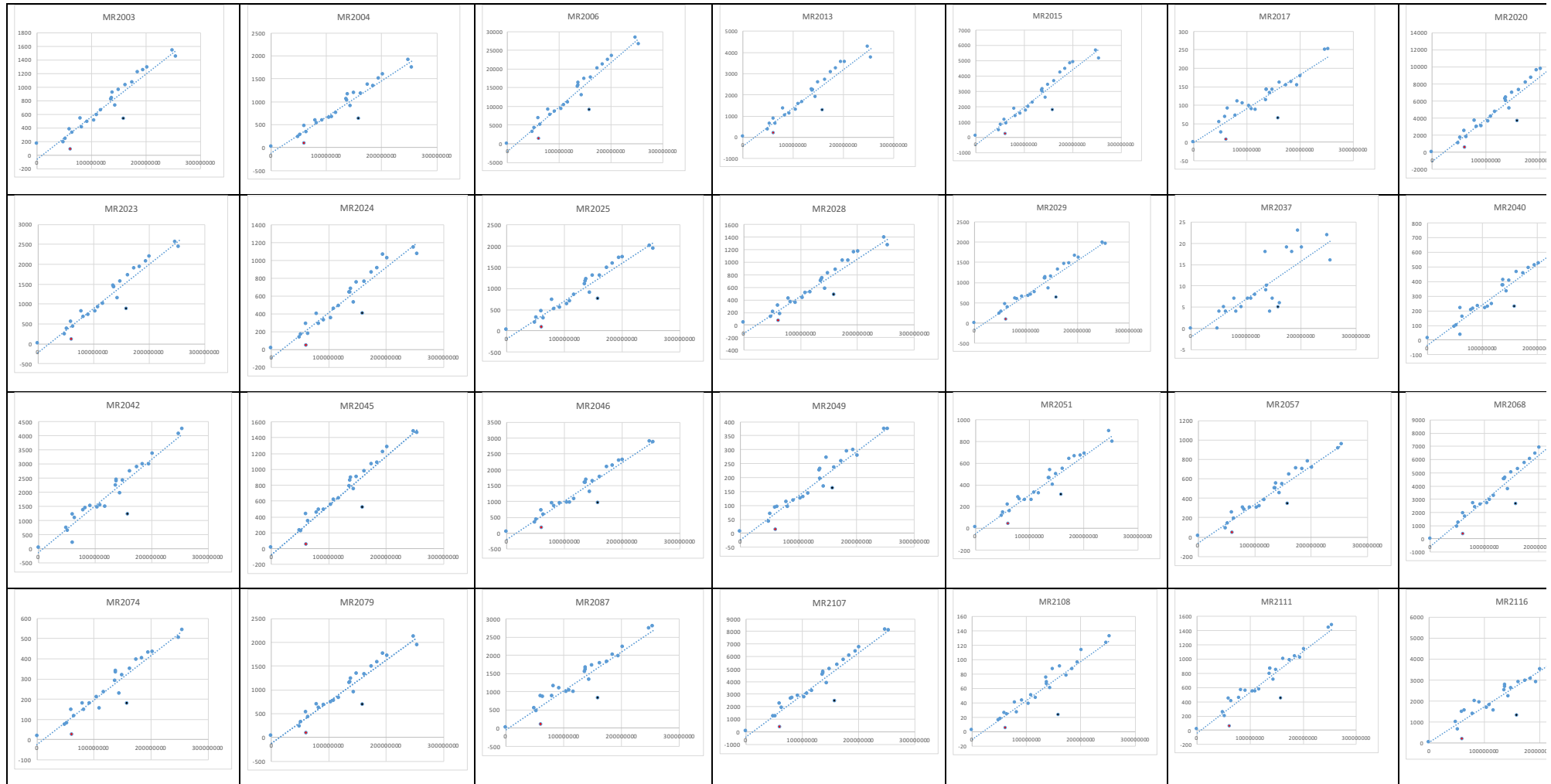


Figure 99 : distribution des ratios nombre de séquences/longueur de chromosome pour chacun des échantillons analysés. Le ratio pour le chromosome X est représenté en noir et le ratio pour le chromosome Y en rouge. Il est observé une bonne corrélation pour tous les autosomes, quel que soit l'échantillon, même avec un nombre de séquences restreint. Seul MR2037, avec seulement 289 séquences humaines, présente une mauvaise distribution et des autosomes qui ont un ratio comparable aux chromosomes sexuels.

Une telle découverte n'est bien sûr pas surprenante, considérant que nous sommes en présence de restes issus d'un navire de la marine royale britannique. Néanmoins, cela a permis de renforcer la confiance en nos échantillons. Mais la détermination du sexe n'était pas la motivation principale de cette étude, mais plutôt l'analyse de la composition bactérienne des tartres.

Les échantillons se sont répartis en deux grandes populations, correspondant apparemment à des communautés bactériennes buccales saines d'une part, pathologiques d'autre part, avec un continuum entre ces deux populations au fur et à mesure de la dégradation de l'état de la bouche. Cette répartition a pu être validée par la mise en évidence de représentants du « complexe rouge », c'est-à-dire *Treponema denticola*, *Porphyromonas gingivalis* et *Tannerella forsythia*. Ces trois bactéries sont particulièrement représentées dans certains extraits de tartre, tandis qu'ils sont peu voire pas retrouvés dans d'autres extraits. Afin de renforcer les distributions en clades, nous avons associé à nos échantillons du Mary-Rose des échantillons de tartre mérovingiens décrits en 3.1 (Kolb-St514 et BSM), des échantillons de tartre néolithiques et antiques analysés au laboratoire, ainsi que des banques de séquençage publiées [103, 217]. Nous avons également ajouté une banque issue d'un extrait de côte, provenant d'un échantillon mérovingien (Kolb-St514) pour servir de représentant externe (*outlier*).

Lorsque nous avons réalisé la comparaison des banques grâce à *MetaPhlAn*, nous avons obtenu une distribution de notre population en trois groupes (et un *outlier*), représenté en Figure 100. Le premier groupe, à droite, rassemblait des échantillons modernes sains (H1 et H2) et pathologique P1 [217], des échantillons anciens (B61 et G12) publiés par Warinner [103] et cinq de nos échantillons anciens, l'un de Kolbsheim et quatre du Mary Rose (MR2004, MR2017, MR2029 et MR2037). Ce groupe était caractérisé par une faible représentation de bactéries du complexe rouge et du complexe orange, tandis que les genres *Bacteroidetes*, *Streptococcus* et *Corynebacterium*, indicateurs d'une bouche saine [217], étaient mieux représentés que dans les autres groupes. *Capnocytophaga*, *Rothia* sont également mieux représentés [218]. Le second groupe, dans la partie gauche de la représentation, présentait encore une bonne présence de *Bacteroidetes*, mais la proportion d'espèces des genres *Treponema*, *Tannerella* et *Desulfobulbus* s'accroissait. Ce dernier genre est lui-aussi un marqueur de périodontite [218]. Ce second groupe

comportait trois échantillons du Mary Rose, les échantillons du second site mérovingien (Boulogne-Sur-Mer) et les tartres néolithiques et antiques.

Enfin, le troisième groupe, central, présentant les plus fortes proportions de représentants du complexe rouge, comportait un échantillon pathologique moderne et la grande majorité des échantillons du Mary Rose (21 échantillons, soit 75%).

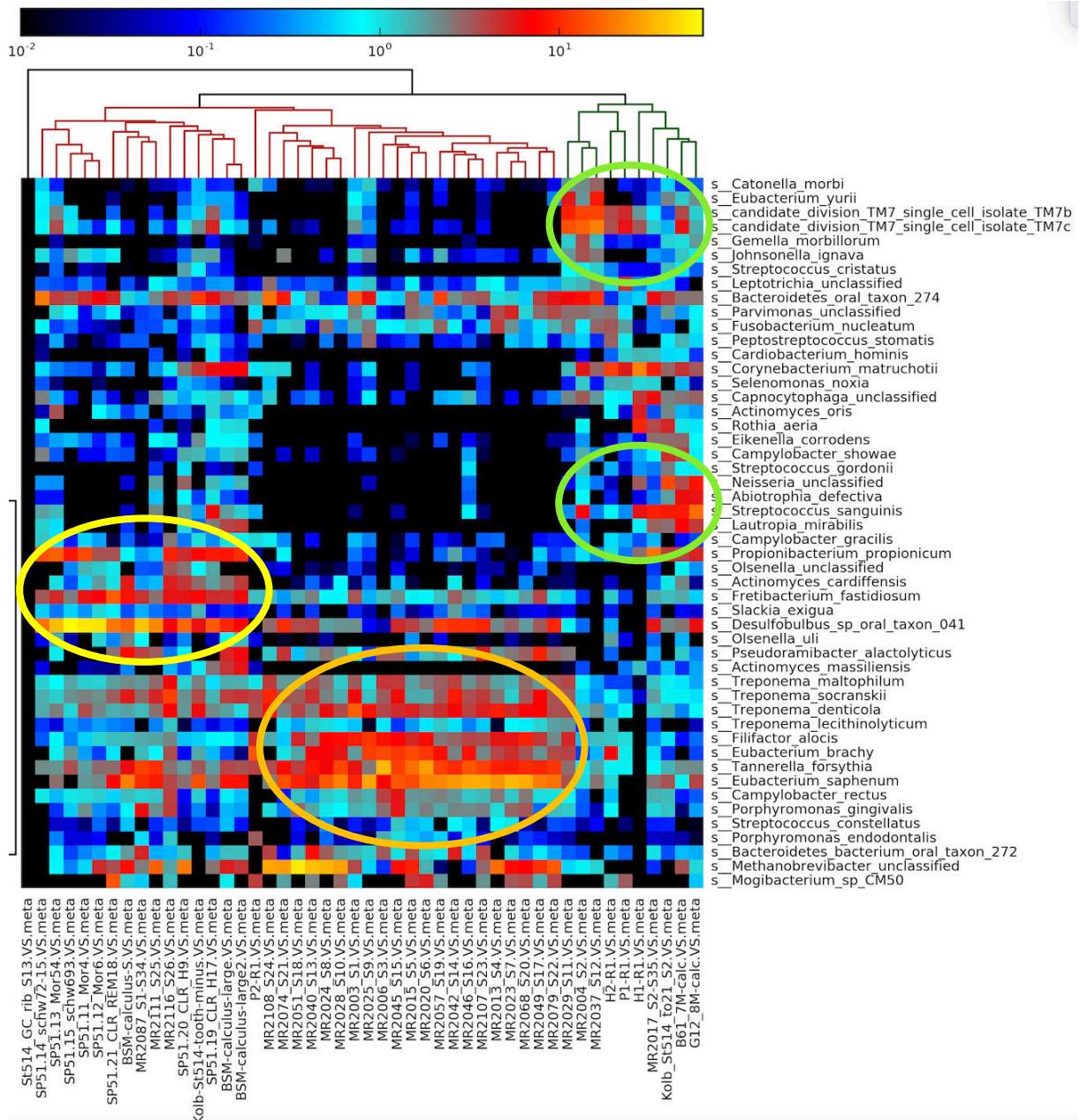


Figure 100 : représentation sous forme de *heatmap* des 50 espèces bactériennes majoritaires identifiées dans les différents échantillons de tartre analysés, comparés entre elles et à un échantillon externe (os costal, à gauche). Le groupe sain est repéré par les orbes vertes, le groupe intermédiaire par l'orbe jaune et le groupe pathologique par l'orbe orange.

Cette distribution semble indiquer un très mauvais état sanitaire de la majorité de l'équipage, confirmant l'image d'Epinal de marins édentés (Figure 101). A l'époque,

l'hygiène buccale était inexistante dans la grande majorité de la population. Les marins étaient particulièrement exposés de par les carences en vitamines liées aux navigations au long cours, le scorbut étant la maladie la plus connue.



Figure 101 : dentiste en uniforme arrachant une dent à un marin à la mâchoire déjà dégarinée. Lithographie colorée, Wellcome Library (18^{ème} siècle (?)).

Nous nous sommes attachés à analyser la répartition des séquences obtenues sur les principaux pathogènes oraux. Pour ce faire, nous avons cartographié la distribution des séquences obtenues lors du séquençage sur les principaux génomes de pathogènes. Cela nous a permis de dresser le Tableau 19.

	total reads (post Quality trimming/merging)	human		Treponema denticola			Tannerella forsythia				Porphyromonas gingivalis				
		Sequences mapped	% mapped vs total	Sequences m	Mean Coverage	% Of Ref Seq	% mapped vs total	Sequences m	Mean Coverage	% Of Ref Seq	% mapped vs total	Sequences m	Mean Coverage	Ref-Seq	% mapped vs total
B61_7M-calc	7500000	666	0,0%	4935	0,14	11,8%	0,1%	34325	0,78	40,5%	0,5%	1449	0,04	3,3%	0,0%
G12_8M-calc	10000000	13885	0,1%	24120	0,72	47,0%	0,2%	78444	1,83	74,3%	0,8%	12158	0,39	28,5%	0,1%
H1-R1	5124862	176240	3,4%	117	0,00	0,2%	0,0%	251	0,00	0,3%	0,0%	94	0,00	0,2%	0,0%
H2-R1	7381107	78205	1,1%	16383	0,42	30,7%	0,2%	39508	0,85	51,7%	0,5%	24814	0,78	49,5%	0,3%
P1-R1	9827681	760703	7,7%	18211	0,44	32,8%	0,2%	21471	0,44	32,6%	0,2%	945	0,03	1,1%	0,0%
P2-R1	8552661	816263	9,5%	36381	0,90	50,4%	0,4%	45610	0,96	55,4%	0,5%	2133	0,06	2,7%	0,0%
BSM-calculus-L	14646084	27417	0,2%	32058	1,16	59,9%	0,2%	91355	2,69	81,2%	0,6%	14477	0,57	28,4%	0,1%
BSM-calculus-L2	6858321	15647	0,2%	9784	0,36	27,0%	0,1%	35616	1,02	54,8%	0,5%	5911	0,22	13,4%	0,1%
BSM-calculus-S	13037267	31248	0,2%	26513	0,88	49,6%	0,2%	102597	2,77	81,5%	0,8%	18437	0,64	27,3%	0,1%
Kolb-St514-to21	1139697	3408	0,3%	3800	0,12	10,7%	0,3%	4723	0,12	10,7%	0,4%	3644	0,13	11,3%	0,3%
Kolb-St514-toot-minu	4564081	1639	0,0%	24360	0,64	41,9%	0,5%	21649	0,59	34,5%	0,5%	6365	0,23	10,9%	0,1%
MR2003_S1	11467540	18246	0,2%	186501	5,89	91,0%	1,6%	110049	2,78	72,8%	1,0%	30941	1,12	44,5%	0,3%
MR2004_S2	7735926	21823	0,3%	40089	1,29	63,0%	0,5%	32993	0,87	44,9%	0,4%	7721	0,28	20,2%	0,1%
MR2006_S3	9095350	324464	3,6%	68378	1,93	65,4%	0,8%	696097	16,71	88,8%	7,7%	79415	2,63	59,7%	0,9%
MR2013_S4	5432600	47252	0,9%	59706	2,03	71,3%	1,1%	83124	2,36	65,8%	1,5%	11580	0,45	24,6%	0,2%
MR2015_S5	8592459	64893	0,8%	81535	2,51	76,8%	0,9%	335911	8,77	87,5%	3,9%	43106	1,54	52,2%	0,5%
MR2017_S2-S35	4551298	7035	0,2%	9773	0,21	17,7%	0,2%	9639	0,16	13,5%	0,2%	7984	0,17	14,5%	0,2%
MR2020_S6	8335379	130272	1,6%	82761	2,72	73,4%	1,0%	677622	19,05	89,1%	8,1%	60200	2,17	58,5%	0,7%
MR2023_S7	9143285	29501	0,3%	54164	1,69	68,1%	0,6%	128407	3,39	71,7%	1,4%	34694	1,25	50,7%	0,4%
MR2024_S8	7004271	13612	0,2%	41331	1,34	59,6%	0,6%	61794	1,65	57,8%	0,9%	14253	0,55	25,6%	0,2%
MR2025_S9	6080972	23615	0,4%	126138	4,23	83,9%	2,1%	83120	2,25	49,7%	1,4%	30129	1,14	44,0%	0,5%
MR2028_S10	9033109	15523	0,2%	39855	1,28	57,2%	0,4%	183089	4,98	77,9%	2,0%	13610	0,51	26,8%	0,2%
MR2029_S11	8825141	22982	0,3%	75376	2,35	79,0%	0,9%	79923	2,10	65,6%	0,9%	25110	0,79	42,8%	0,3%
MR2037_S12	561231	227	0,0%	4178	0,15	12,8%	0,7%	5690	0,17	12,2%	1,0%	654	0,03	2,3%	0,1%
MR2040_S13	3883861	7901	0,2%	13327	0,41	29,8%	0,3%	51307	1,29	62,7%	1,3%	10067	0,34	24,0%	0,3%
MR2042_S14	5797281	48757	0,8%	44423	1,30	59,5%	0,8%	93903	2,17	73,9%	1,6%	18655	0,58	37,2%	0,3%
MR2045_S15	7196732	17816	0,2%	40636	1,26	59,8%	0,6%	132125	3,32	82,2%	1,8%	62535	2,15	67,8%	0,9%
MR2046_S16	950801	33186	3,5%	10607	0,30	23,6%	1,1%	47902	1,10	56,3%	5,0%	6363	0,20	15,7%	0,7%
MR2049_S17	11335450	4417	0,0%	42469	1,26	62,1%	0,4%	66515	1,63	65,0%	0,6%	4553	0,16	8,3%	0,0%
MR2051_S18	6028076	10070	0,2%	16987	0,54	36,1%	0,3%	26291	0,67	34,4%	0,4%	13292	0,46	29,2%	0,2%
MR2057_S19	6841326	10944	0,2%	77774	2,62	80,3%	1,1%	236070	6,61	85,2%	3,5%	44088	1,72	54,1%	0,6%
MR2068_S20	16936423	95054	0,6%	125395	4,11	83,2%	0,7%	251676	6,76	87,2%	1,5%	73374	2,70	72,5%	0,4%
MR2074_S21	5074055	6384	0,1%	16280	0,52	35,8%	0,3%	73486	1,94	52,6%	1,4%	5854	0,20	10,1%	0,1%
MR2079_S22	20005113	24510	0,1%	90647	2,88	81,4%	0,5%	314980	8,39	87,8%	1,6%	41804	1,52	56,8%	0,2%
MR2087_S1-S34	6528012	42316	0,6%	21004	0,49	35,4%	0,3%	195097	3,57	84,7%	3,0%	78455	1,81	69,8%	1,2%
MR2107_S23	16960522	96185	0,6%	212108	6,85	88,5%	1,3%	575801	15,51	89,2%	3,4%	116283	4,31	80,2%	0,7%
MR2108_S24	660580	1451	0,2%	5050	0,17	14,8%	0,8%	3908	0,11	9,2%	0,6%	1328	0,05	4,4%	0,2%
MR2111_S25	8607212	17392	0,2%	32618	1,06	55,0%	0,4%	142087	3,88	85,0%	1,7%	32223	1,21	52,8%	0,4%
MR2116_S26	10706252	54316	0,5%	65577	1,81	66,0%	0,6%	40783	0,92	42,6%	0,4%	48362	1,49	57,6%	0,5%
SP51.11_Mor4	9375918	1218	0,0%	23539	0,48	34,5%	0,3%	18405	0,26	20,7%	0,2%	15528	0,34	26,3%	0,2%
SP51.12_Mor6	7459126	2805	0,0%	15849	0,30	24,1%	0,2%	25824	0,36	27,2%	0,3%	5737	0,12	7,6%	0,1%
SP51.13_Mor54	14621380	5058	0,0%	46514	0,92	53,8%	0,3%	47168	0,67	42,4%	0,3%	6704	0,14	9,1%	0,0%
SP51.14_schw72-15	12466665	6277	0,1%	50159	1,06	56,9%	0,4%	84172	1,40	48,6%	0,7%	2131	0,05	1,8%	0,0%
SP51.15_schw693	5214812	8592	0,2%	9071	0,15	13,3%	0,2%	32579	0,41	31,9%	0,6%	6399	0,12	8,9%	0,1%
SP51.19_CLR_H17	14268823	15800	0,1%	30079	0,63	42,4%	0,2%	33341	0,52	31,6%	0,2%	3094	0,07	5,0%	0,0%
SP51.20_CLR_H9	12052150	672	0,0%	20059	0,41	31,1%	0,2%	26297	0,40	26,5%	0,2%	2057	0,04	3,0%	0,0%
SP51.21_CLR_REM18	5306398	2772	0,1%	6592	0,12	10,9%	0,1%	13442	0,17	15,2%	0,3%	4363	0,09	5,4%	0,1%

Tableau 19 : cartographie des séquences des différents échantillons sur les génomes humain et des trois bactéries du complexe rouge. La longueur des barres dans les cellules est proportionnelle à la quantité de séquences rapportées aux échantillons entre eux. Les échantillons modernes présentent la proportion majoritaire de séquences cartographiées sur le génome humain, car prélevés sur des patients vivants. Les couleurs des échantillons sont corrélées à leur regroupement dans la représentation en heatmap.

Comme attendu, et décrit dans l'article de Liu *et al.* [217], les échantillons modernes étaient riches en ADN humain, car prélevés sur des patients. Pour les échantillons anciens, trois échantillons contenaient plus de 1% des séquences qui correspondent à de l'ADN humain, MR2006, MR2020 et MR2046. Afin d'isoler les séquences issues de *T. denticola*, nous avons utilisé *bwa*, en prenant comme séquence de référence la souche ATCC 35405 [219]. En recherchant les séquences de *T. denticola*, certains échantillons issus de tartres de marins du Mary Rose ont présenté des taux de couverture et des profondeurs de séquençage élevés pour des échantillons anciens, avec jusqu'à 91% de couverture et une profondeur de 7x. Une telle couverture a permis de reconstruire le génome de *T. denticola* pour ces échantillons et de les comparer à des échantillons modernes. Nous avons sélectionné les dix échantillons pour lesquels nous avons une profondeur de couverture supérieure ou égale à trois (MR2003, 2006, 2013, 2015, 2020, 2025, 2029, 2057, 2068, 2079, 2107) et nous avons importé les cinq génomes disponibles (outre la souche de référence) de *T. denticola* (NZ_CM00194-195-196-197-198) depuis le site www.patricbrc.org. Plusieurs autres souches sont en cours d'assemblage, et nous avons importé depuis le site NCBI les données de séquençage disponibles. Nous avons également analysé la diversité de l'espèce *T. denticola* à partir des données de séquençage disponibles sur le site NCBI (Tableau 20).

Strain	Level	Size	Scaffolds	Symmetric Identity	Gapped Identity
ATCC 33521	Scaffold	2,76211	3	98,0369	99,978
ATCC 35404	Chromosome	2,77151	-	98,0346	99,9837
AL-2	Chromosome	2,84309	-	86,2191	97,4617
H-22	Chromosome	2,75786	-	85,7465	97,4724
US-Trep	Scaffold	2,76573	2	85,4225	97,2102
ASLM	Scaffold	2,76824	2	85,3331	97,213
ATCC 33520	Scaffold	2,88689	2	85,0458	97,2624
SP37	Scaffold	2,82377	2	84,9725	96,956
SP44	Scaffold	2,82474	8	84,941	96,9556
SP23	Scaffold	2,826	3	84,8726	96,9526
F0402	Scaffold	2,74263	12	84,6504	97,4625
H1-T	Chromosome	2,9297	-	84,0196	97,2187
MYR-T	Scaffold	2,93585	2	83,991	97,2147
OTK	Chromosome	2,85753	-	83,8355	97,0416
SP33	Scaffold	2,93349	4	73,082	94,1834
SP32	Scaffold	2,99028	6	72,6693	94,171

Tableau 20 : rapport d'analyse de similarité entre les séquences disponibles sur le site www.ncbi.nlm.nih.gov consulté le 28/08/2016, comparées à la souche de référence ATCC 35405.

Cette diversité est importante, le Tableau 20 montrant les homologies de séquences entre les différentes souches disponibles dans la base de données NCBI, comparées à la souche de référence ATCC 35405. Il est possible d'observer une identité de séquence de seulement 72,7% pour la souche la plus éloignée, qui est la souche SP32 soulignée en rouge dans la Figure 102.

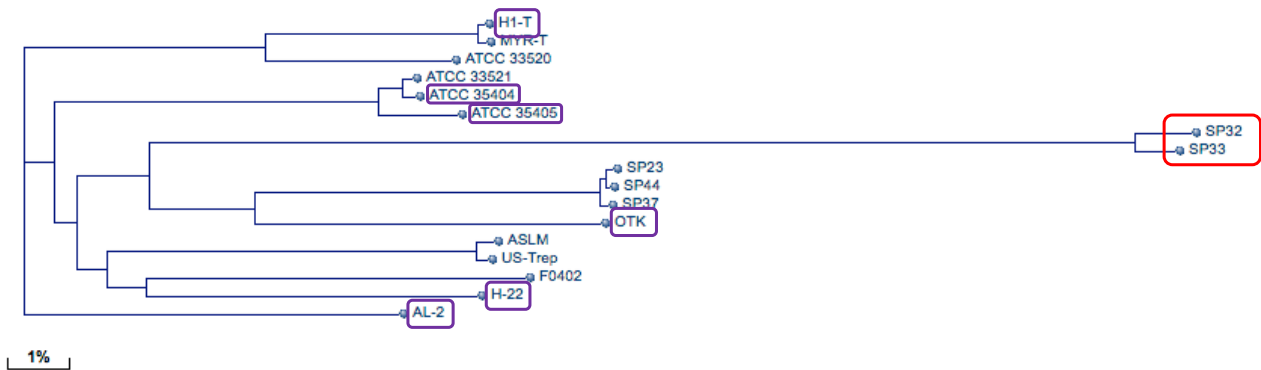


Figure 102 : dendrogramme de l'espèce *T. denticola*, calculé à partir des données de BLAST sur le site NCBI, consulté le 28/08/2016. Les génomes assemblés sont entourés en violet, tandis que les souches les plus éloignées sont identifiées par un cadre rouge.

La comparaison des séquences obtenues de nos échantillons et des séquences publiées a été réalisée à l'aide du logiciel Bionumerics, en utilisant des scripts et pipelines développés par Y. Blouin [220].

Du fait de la couverture incomplète de chaque échantillon, et de la diversité des zones non couvertes en fonction des échantillons et des séquences publiées, il est resté un *core-génome* de 352.772 pb. Cela peut sembler peu en première analyse (le génome de *T. denticola* ATCC 35405 comprend 2.843.201 pb), mais ce *core-génome* est bien supérieur aux quelques kilobases utilisées par le MLST [221] et reste tout de même conséquent en se rappelant que nous partons de données métagénomiques.

Ces 352.772 pb de *core-génome* ont permis de définir 24.452 SNPs séparant entre elles les 28 séquences analysées. A partir de ces SNPs, nous avons pu rechercher les relations phylogénétiques entre les souches, visualisées sur l'arbre représenté en Figure 103.

Cet arbre nous a fourni des données très intéressantes quant à la population analysée. Nos souches anciennes se sont connectées sur la ligne reliant *T. putidum*, *outgroup*, à l'ensemble des souches modernes. De plus, il est arrivé en trois occurrences (lignes en rouge), que deux souches anciennes aient une branche commune à partir de cette tronc reliant *T. putidum* aux souches modernes de *T. denticola*. Cette distribution le long de cette ligne peut indiquer l'existence de lignées anciennes de *T. denticola*, disparues depuis l'époque du Mary-Rose ou non encore séquencées. Les souches séquencées ont

été isolées majoritairement sur des individus nord-américains, au Canada ou aux Etats-Unis d'Amérique [222]. Il est probable que la diversité mondiale soit bien plus élevée. De plus, nos souches prédatent l'expansion de la population européenne aux Amériques. C'est que ce que pourrait indiquer le placement de la souche MR2029, qui partage 220 SNPs avec trois souches modernes, très proches entre elles, ATCC 35404, ATCC 35405 et ATCC 33521.

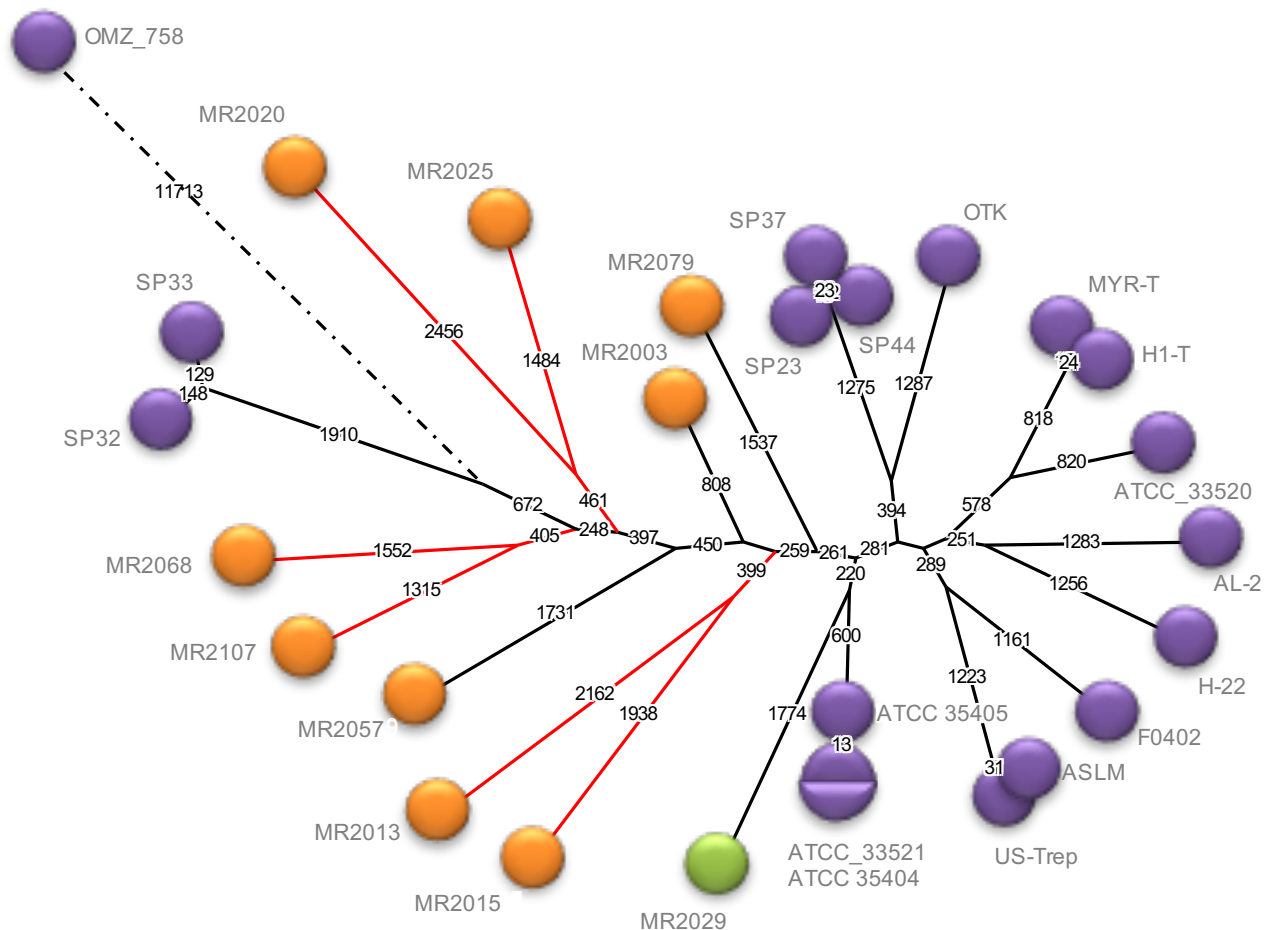


Figure 103 : représentation graphique sous forme d'arbre couvrant minimal (MST, Minimum Spanning Tree) des relations phylogénétiques inférées des SNPs identifiés entre les différentes souches et séquences analysées de *Treponema denticola*. Les souches publiées sont représentées en violet. La souche OMZ_758 appartient à l'espèce *T. putidum*, espèce la plus proche de *T. denticola* [223]. La couleur des autres souches est corrélée à leur positionnement dans les clusters hautement pathologique (orange) ou sain (vert). Le nombre de SNPs séparant les souches entre elle est indiqué sur chaque branche. La ligne en pointillés n'est pas proportionnelle au nombre de SNPs.

Cette souche pourrait être proche des souches qui ont été apportées dans le Nouveau-Monde, et qui se sont diversifiées avec la multiplication de la population d'origine européenne. Par ailleurs, les deux souches séquencées SP32 et SP33, présentant la plus faible similarité avec la souche de référence (Tableau 20), sont dans notre arbre très

éloignées des autres souches séquencées. La topologie générale de notre arbre est comparable, pour les souches séquencées, au dendrogramme obtenu à partir de NCBI (Figure 102). Ces deux éléments montrent la robustesse de notre analyse, malgré le fort taux de bases non prises en compte (2.490.429 bases masquées dans l'analyse sur 2.843.201 bases dans la séquence de référence, soit 87%).

Dans le même ordre d'idée, nous avons procédé à la recherche de SNPs pour *T. forsythia*, en retenant les 14 échantillons présentant une profondeur de couverture supérieure ou égale à trois, auxquels nous avons ajouté le concaténat des séquences obtenues à partir des tartres de BSM (nous ne l'avons pas fait pour *T. denticola* car la profondeur de séquençage était insuffisante). Nous avons utilisé comme séquence de référence la souche 92A2 [224] et téléchargé depuis le site www.patricbrc.org les trois autres génomes disponibles, correspondant aux souches ATCC 4307, KS16 et 3313. La souche ATCC 4307 s'est révélée identique à la souche de référence 92A2 et n'a donc pas été utilisée ultérieurement. Nous avons tenté d'ajouter les cinq génomes de *Tannerella sp.* disponibles à l'état de *contigs* ou de *scaffolds* sur le site NCBI mais seule la souche ASM100648 présentait suffisamment d'homologie avec la souche de référence pour être ajoutée. Nous avons obtenu pour cette espèce un *core-génome* de 125.694 pb, pour un génome dans la souche de référence comprenant 3,405,521 pb. Ce *core-génome* a permis de déterminer 1703 SNPs, séparant les 19 séquences analysées.

A partir de ces SNPs, nous avons pu rechercher les relations phylogénétiques entre les souches, visualisées sur l'arbre représenté en Figure 104. La topologie de cet arbre est différente de celle obtenue avec *T. denticola*. Nous n'avons pu inclure de représentant d'une espèce proche, *T. forsythia* étant la seule espèce dans son genre. De plus, des échantillons sont en commun entre les deux arbres, mais il n'y a pas d'identité parfaite. Les échantillons qui partageaient une branche commune dans l'arbre de *T. denticola* n'ont pas été regroupés dans cet arbre et vice-versa. On peut remarquer que la longueur des branches est moindre, pouvant indiquer une diversité moindre des souches de *T. forsythia* comparée à *T. denticola*. Il n'a pas non plus été noté de regroupement évident selon l'état pathologique de la bouche. Nous n'avons pas dans cet arbre de représentant isolé de bouche supposée saine, le nombre de séquences obtenues étant insuffisant pour permettre une reconstruction phylogénétique. Nous avons observé cependant des groupes de souches et surtout, l'ensemble des souches se branchent sur l'axe reliant les quatre souches séquencées, ce qui pourrait indiquer que la diversité des souches

circulant en Angleterre au XVIème siècle était inférieure ou égale à la diversité moderne, les populations de ce pathogène oral ayant continué à se diversifier depuis cette époque et n'ayant pas subi de goulot d'étranglement ou d'expansion d'un clone ayant acquis une virulence particulière.

Il n'a pas été possible de suivre cette stratégie pour *P. gingivalis*, les profondeurs de séquençage et couvertures étant insuffisantes.

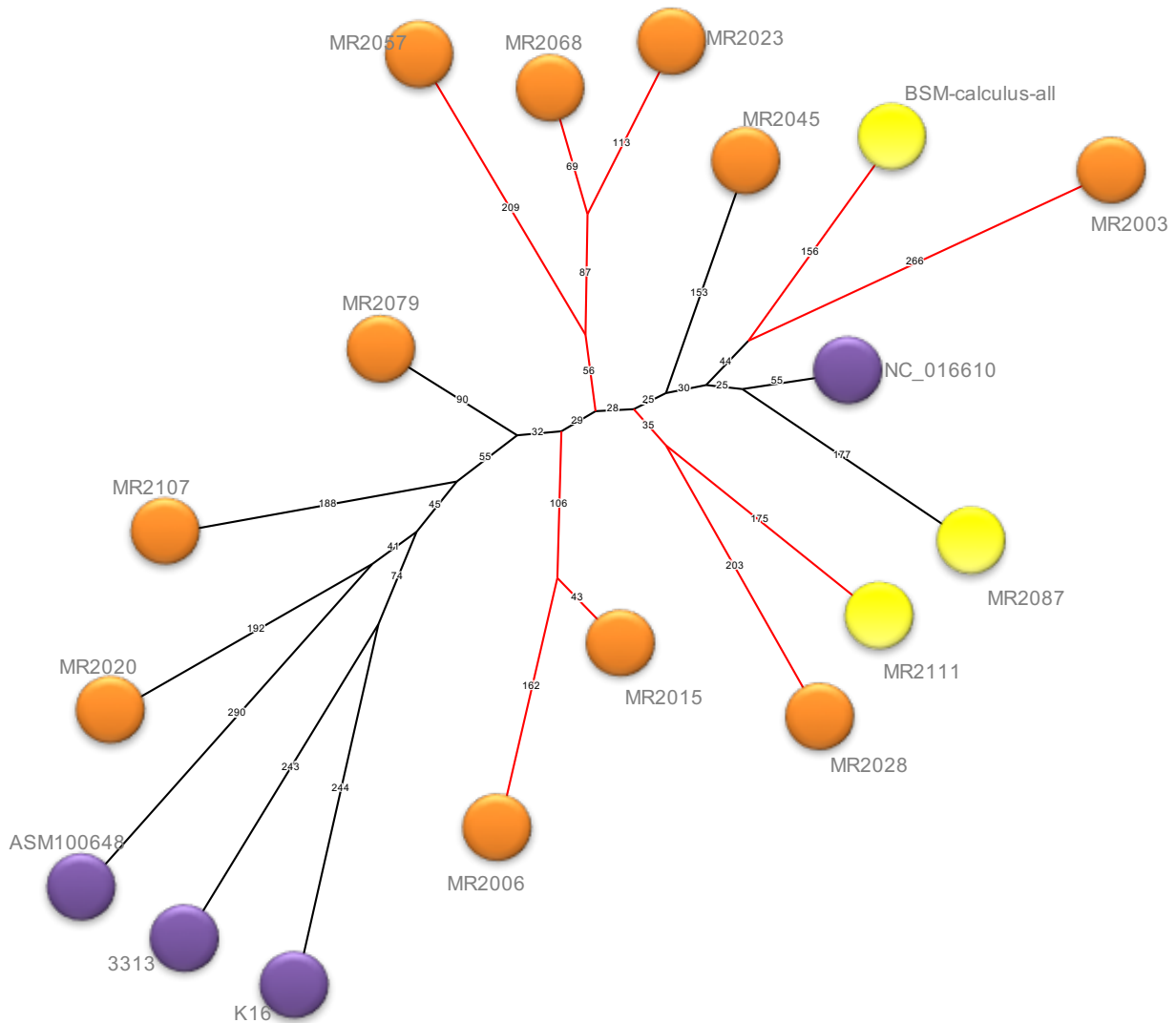


Figure 104 : représentation graphique sous forme d'arbre couvrant minimal (MST, Minimum Spanning Tree) des relations phylogénétiques inférées des SNPs identifiés entre les différentes souches et séquences analysées de *Tannerella forsythia*. Les souches publiées sont représentées en violet, La couleur des autres souches est corrélée à leur positionnement dans les clusters pathologique (jaune) et hautement pathologique (orange). Le nombre de SNPs séparant les souches entre elle est indiqué sur chaque branche.

Ces analyses ont montré la pertinence d'analyser cet échantillon particulier qu'est le tartre dentaire. Il recèle une quantité d'informations extrêmement importante, rapportée à son volume. Il nous a permis de faire un état sanitaire de l'équipage du Mary-Rose 500 ans après son naufrage, en montrant que 85% des marins avaient des problèmes dentaires, avec les $\frac{3}{4}$ présentant des périodontites plus ou moins avancées. Il a été possible de reconstruire le génome de deux des principaux pathogènes bactériens responsables de ces infections dentaires, sans qu'il soit nécessaire de réaliser un séquençage ultra-profond. Un séquençage sur NextSeq a suffi, en multiplexant les 28 échantillons à analyser et en utilisant un kit 2x150 pb. Cette analyse est en partie possible grâce au bruit de fond environnemental faible, le tartre étant très peu perméable aux bactéries de l'environnement et permettant la conservation de l'ADN des bactéries ayant contribué à la fabrication de la plaque dentaire. Le tartre est donc un matériel de choix pour analyser l'état sanitaire de populations anciennes, humaines ou animales. Il est également possible de chercher à mettre en évidence des pathogènes systémiques susceptibles de se retrouver dans la salive et de chercher à caractériser le régime alimentaire en analysant les séquences mitochondriales non endogènes et les séquences chloroplastiques.

CONCLUSION ET PERSPECTIVES

Dans ce travail, nous avons cherché à caractériser l'évolution des populations bactériennes lors de la décomposition du corps après la mort d'un individu et à suivre en parallèle la dégradation de l'ADN pour corrélérer ces deux phénomènes. Cette étude devait aussi servir à aider à la recherche de traces d'ADN de pathogènes bactériens dans des os et dents anciens. Or, il est vite apparu, avec les bactéries Gram négatif, que l'ADN bactérien présent au moment de la mort de l'individu, au même titre que l'ADN endogène, était très rapidement dégradé par les micro-organismes. Ce phénomène n'avait pas encore été décrit mais explique grandement la difficulté à mettre en évidence un agent pathogène par son ADN dans des restes anciens. Cette dégradation rapide de l'ADN endogène est liée à l'activité bactérienne de la communauté environnementale, qui colonise rapidement les tissus, dont les os, et qui utilise ceux-ci d'abord comme substrat, puis comme matrice.

Au cours de notre projet, en parallèle de notre étude taphonomique, l'équipe de Metcalf et de Knight a publié trois articles d'intérêt sur l'évolution des populations bactériennes dans des cadavres [155, 225, 226]. La première, publiée en 2013, avait pour objectif de déterminer s'il était possible d'utiliser l'évolution des populations microbiennes dans des carcasses de souris [223] pour déterminer l'intervalle de temps depuis le décès de l'organisme. Les auteurs se sont clairement positionnés dans une approche forensique, afin de tenter de mettre en place une sorte d'horloge microbienne. Leurs analyses leur ont permis de dégager des tendances fortes, et de proposer un modèle de datation des dépouilles à trois jours près, sur un intervalle allant jusqu'à 48 jours après la mort. Les échantillons étaient obtenus à partir du sol, de la peau et de la cavité abdominale des souris, et les auteurs n'ont pas du tout analysé l'évolution des populations bactériennes dans les os. La meilleure précision leur était offerte par la combinaison des caractérisations des populations bactériennes et microbiennes eucaryotes, en particulier mesurées à partir des sols en contact direct avec les dépouilles. Le second article, en 2014, cherchait à explorer l'action des microbes du sol dans la décomposition de carcasses de souris laissées en surface [155]. La complexité des phénomènes en jeu n'a pas permis aux auteurs de déterminer l'origine, endogène ou environnementale, des microorganismes responsables des phénomènes de dégradation de la dépouille. Nous

avons rencontré les mêmes difficultés dans notre étude, les marqueurs étudiés ne permettant pas l'identification fine des espèces bactériennes et de leur origine, tellurique ou intestinale. Il a été néanmoins possible aux auteurs de [155] de démontrer que la population microbienne du sol sur lequel est déposé la dépouille influe sur les phénomènes de dégradation, celle-ci étant retardée lorsque le sol est stérilisé. Enfin, leur dernier article, publié en 2016 [224], tente de faire la synthèse des connaissances sur les phénomènes de dégradation post-mortem, en utilisant des corps humains se décomposant à l'air libre dans une installation dédiée (Installation de criminalistique appliquée ou *Body Farm*) et en utilisant des souris mises à se décomposer sur trois types de sols différents dans des conditions de température et d'humidité contrôlées. Grâce à ces essais, ils ont développé un modèle prédictif des phénomènes se déroulant au cours de la décomposition des corps, et ont trouvé pour les souris, quel que soit le sol employé, une succession prédictible de communautés bactériennes, présentes à l'état latent dans les trois sols et se développant de manière comparable au cours des différentes phases taphonomiques précoces. Il ne leur a pas été possible de déterminer les espèces bactériennes précisément actives, et ont considéré que ces populations ne pouvaient se caractériser au niveau de l'espèce, mais à des niveaux taxinomiques plus élevés. Ils en concluent néanmoins que l'analyse des populations microbiennes, bactériennes et eucaryotes, permet de déterminer l'intervalle post-mortem de manière plus précise que les méthodes communément employées lorsque l'intervalle devient important. Ils montrent aussi, comme nous, que les communautés bactériennes tendent à se ressembler entre tissus et sols. De même, ils ont vérifié que les communautés microbiennes étaient modifiées dans les sols au contact, et que cette modification perdurait au moins 30 jours après le retrait des corps. Cette découverte est d'intérêt pour la mise en évidence de tombes clandestines ou de sites où des cadavres auraient été temporairement laissés. La durée d'étude ne leur a pas permis de montrer le retour à l'état initial des sols en contact avec les dépouilles. Nos deux approches sont donc complémentaires et permettent de mieux décrire l'ensemble des phénomènes en œuvre au niveau microbiologique lors de la décomposition des vertébrés dans l'environnement, avec pour l'équipe de Metcalf et Knight une vision pratique criminalistique et pour nous une vision historique et archéologique.

Dans notre étude, nous montrons que dans les 18 mois qui suivent l'enfouissement des souris, les os présentent un profil métagénomique comparable au sol environnant. Nous

avons retrouvé ce phénomène avec les échantillons osseux anciens, pour lesquels les profils métagénomiques ressemblaient à ceux de sols. En règle générale, les ossements analysés avaient été nettoyés après leur découverte, et les archéologues n'avaient pas réalisé de prélèvement de sol sur le site de leur découverte. Il n'est donc pas possible d'établir une comparaison précise entre les os analysés et les sols environnants, mais les bactéries majoritaires identifiées dans les os sont des bactéries telluriques ou ubiquitaires. Cela accroît la complexité des analyses visant à identifier un pathogène dans les ossements anciens, la population bactérienne d'un sol étant très diverse, potentiellement spécifique de ce sol et, par conséquent, nécessairement méconnue dans sa diversité. Nous avons décrit dans la partie bibliographique la majorité des publications ayant mis en évidence, ou pensant avoir mis en évidence des traces d'ADN pathogènes dans des restes anciens. Si l'on fait abstraction des nombreux articles sur *M. tuberculosis*, dont la crédibilité a été fortement entamée depuis la démonstration de la présence de marqueurs identiques à ceux réputés spécifiques dans d'autres mycobactéries du sol, il ne reste que de rares publications, ayant pour principal agent *Yersinia pestis*. Il a été possible d'identifier des os portant l'ADN de ce pathogène en criblant les échantillons de fosses communes qui ont été remplies lors des épidémies de peste. Le cas de figure est relativement exceptionnel, mais plusieurs fosses communes ont déjà été explorées en Europe, l'exemple le plus célèbre étant le cimetière d'East Smithfield à Londres, qui a livré les restes ayant permis le séquençage complet de la souche de *Y. pestis* responsable de la peste noire.

Il n'existe pas d'autres pathogènes bactériens qui provoquent une telle mortalité en un temps aussi restreint, et il n'y aura donc pas de fosses communes permettant un recueil de nombreux restes pour analyse. Cependant, certaines maladies entraînaient un enfermement des malades. Le meilleur exemple est la lèpre, qui se traduisait pour les malades par un rejet de la société, une stigmatisation (la crécelle) et un confinement dans des léproseries ou lazarets. Les ossements retrouvés dans ces établissements ont donc une plus grande probabilité d'être issus de malades, et donc la chance de mettre en évidence des traces d'agent pathogène est plus élevée que pour un ossement, présentant les mêmes signes, mais issu d'un site isolé. Cela a permis de séquencer le génome de *M. leprae* circulant en Europe il y a mille ans [22]. Enfin le séquençage d'une souche de *Mycobacterium tuberculosis* du Nouveau-Monde, prédatant l'arrivée de Christophe

Colomb a été réalisé à partir de seulement trois échantillons après analyse de 68 restes humains portant des lésions indicatives de la tuberculose [40].

Ces exemples sont exceptionnels et correspondent à de riches lots d'échantillonnage, susceptibles de compenser la pauvreté des échantillons en ADN d'intérêt. Notre étude de taphonomie montre que l'ADN bactérien, sauf exception, en particulier les mycobactéries de par la composition particulière de leur paroi, n'est ni plus ni moins stable que l'ADN endogène mais qu'étant beaucoup moins bien représenté, il sera beaucoup plus difficile à mettre en évidence. L'identification bactérienne par qPCR et par séquençage repose sur la mise en évidence de séquences spécifiques à chaque espèce bactérienne. Ces séquences ne représentent qu'une petite fraction du génome par rapport aux gènes communs à toutes les bactéries, comme le gène de l'ADNr 16S ou les gènes de ménage. Les bactéries pathogènes meurent ou entrent en dormance rapidement après la mort de leur hôte, et leur ADN est dégradé par des bactéries, endogènes et environnementales. Ces bactéries à leur tour vont mourir pour laisser la place à d'autres populations, adaptées au sol dans lequel est enfoui l'hôte. Elles seront donc dégradées à leur tour. Les gènes spécifiques des bactéries pathogènes, ne représentant qu'une fraction du génome, seront dilués dans le pool de gènes bactériens apportés par les différentes espèces bactériennes spécialisées dans la dégradation des organismes morts dans le sol de l'ADN, ce qui rend encore plus difficile la mise en évidence d'un agent pathogène bactérien dans un échantillon ancien, la majorité des gènes de ce pathogène n'étant pas suffisamment différents des gènes des bactéries environnementales surtout quand des petits fragments d'ADN sont comparés. La recherche de ces pathogènes dans des échantillons anciens représente donc encore un défi technologique, mais surtout la nécessité de sélectionner et de trier les échantillons à analyser pour augmenter les chances de pouvoir mettre en évidence un signal.

Dans cette étude, il n'a pas été possible de suivre la décroissance de l'ADN bactérien exogène sur le long terme, la quantité initiale étant faible du fait du modèle employé. Il serait intéressant de chercher à améliorer ce modèle, en augmentant la quantité d'ADN bactérien présent dans les os au moment du sacrifice des animaux modèles. Ceci pourrait être fait en identifiant une bactérie capable de développer des affections chroniques dans les os de souris, et donc coloniser de manière plus systématique et à plus forte concentration les cadavres. Une alternative serait de développer un modèle animal capable de recevoir des injections plus concentrées en bactéries, ou plus

susceptibles à des bactéries modèles, capables de mimer plus fidèlement une septicémie que le protocole que nous avons exploité ici. Enfin, un séquençage à plus grande profondeur, s'affranchissant de l'amplification de l'ADNr 16S, sera peut-être plus à même d'identifier les différentes populations bactériennes responsables de la dégradation des tissus et de la colonisation des os lors de l'enfouissement des cadavres. Pour étudier l'effet de différents sols sur les phénomènes taphonomiques, et comprendre les lois qui régissent la décomposition des corps, il est nécessaire d'élargir le spectre des sols analysés, en particulier afin de mieux appréhender les phénomènes complexes influant sur la préservation à long terme de l'ADN endogène.

Ayant cherché à mettre en évidence d'autres bactéries plus difficiles à détecter dans les squelettes animaux et humains du passé, ce volet de notre recherche est resté sans résultats, en particulier pour les échantillons animaux, qui n'ont pas livré de séquences de pathogènes bactériens, malgré la recherche de plusieurs d'entre eux.

Pour les échantillons humains (hors tartres des marins du Mary-Rose), nous nous sommes attachés à analyser des ensembles porteurs de signes pertinents d'infection, en se focalisant sur les kystes calcifiés, au sein d'une collaboration avec une doctorante de l'équipe travaillant sur les parasites. Notre postulat était que si le kyste avait entouré un corps étranger (parasite) ou confiné une zone infectieuse (bactérie), nous pourrions retrouver dans le kyste des traces d'ADN de l'agent étranger. Ce postulat s'est révélé faux dans les cas analysés. En effet, dans le cas de kystes calcifiés, cette minéralisation a lieu du vivant de l'hôte, et transforme une matrice de fibroblastes en une coque minérale qui entoure l'intrus. A la mort de l'individu, le contenu du kyste se dégrade au même titre que le reste des tissus de l'organisme. On peut imaginer que l'intérieur du kyste reste à l'abri des bactéries environnementales tant que le kyste reste intact et s'il est complet. Néanmoins, au cours du temps, les mouvements dans le sol entraînent le plus souvent une fragmentation du kyste, permettant aux bactéries environnementales de coloniser l'intérieur du kyste et donc de poursuivre la dégradation et la dilution de l'ADN éventuellement présent. Seul reste protégé sur le long terme l'ADN prisonnier du tissu calcifié de la coque. Il est possiblement moins accessible aux bactéries environnementales, il ne peut être déplacé de la matrice minérale lors de ruissellements ou autres phénomènes liés à l'eau. Il se dégrade à cause de l'hydrolyse, mais *in fine*, lors de l'analyse des données de séquençage, il comporte encore une proportion élevée

d'ADN endogène. Notre étude comparative sur trois échantillons a confirmé cette hypothèse, en permettant d'obtenir des taux d'ADN endogène dépassant dans certains cas 50% de l'ADN total. C'est d'ailleurs à notre connaissance le reste potentiellement le plus riche en ADN endogène, devant l'os pétreux et la pulpe dentaire.

Dans le cas du Mary-Rose, nous avons obtenu une grande richesse d'informations à partir des quelques milligrammes des échantillons de départ, qui ont permis d'estimer l'état sanitaire des marins de ce navire 500 ans après son naufrage, et démontrer que plus de $\frac{3}{4}$ des marins présentaient des atteintes pathologiques de la bouche, avec des périodontites plus ou moins avancées. Avec ces quelques milligrammes, il a également été possible de reconstituer dans quelques cas le génome de pathogènes membres du complexe rouge, *Tannerella forsythia* et *Treponema denticola*. La quantité d'informations contenues dans ces quelques milligrammes a dépassé toutes nos attentes, et des séquençages plus profonds seront susceptibles de rendre plus robustes et de compléter les arbres phylogénétiques proposés dans ce travail, voire réussir à détecter d'autres pathogènes que ceux spécifiques de la bouche. D'autres types d'analyses sont envisageables sur les tartres dentaires, telles que la protéomique. Il a déjà été mis en évidence des protéines alimentaires dans ce matériau, mais d'autres protéines pourraient être identifiées, comme des anticorps, ou d'autres marqueurs immunologiques, permettant d'étudier l'évolution du système immunitaire, et indirectement l'environnement pathogène, des individus dont le tartre est issu. La recherche de protéines bactériennes, mais également virales, est également envisageable, ce qui ouvrirait une fenêtre sur les pathologies virales, plus difficiles à mettre en évidence car moins souvent responsables de modifications morphologiques. Et nombre de virus infectant l'homme sont des virus à ARN, plus fragile que l'ADN et donc plus difficile à mettre en évidence sur des échantillons anciens. Un important travail méthodologique sera sans doute également nécessaire pour permettre l'exploitation de l'ARN ancien, en adaptant les techniques de RT-PCR. Quelques travaux initiaux sont prometteurs, en analysant l'ARN contenu dans des graines, mais les mécanismes de préservation de l'ARN sont particuliers dans ce type d'échantillons. Avec la découverte de l'immense diversité des virus dans l'environnement, et leur caractérisation, il sera possible de réduire la proportion de séquences non attribuées dans les séquençages *shotgun*, et de suivre l'évolution des virus et des bactéries sur des périodes de temps étendues, en limitant l'influence de l'environnement, les tartres, et

plus largement les tissus calcifiés, offrant une sorte d'instantané de la diversité microbienne à un moment donné.

Enfin, pour en revenir à la problématique initiale, qui était la mise en évidence d'activités passées ou masquées de prolifération biologique, les éléments de taphonomie et d'analyse d'échantillons anciens montrent qu'il sera très difficile de mener à bien une telle recherche. L'environnement recycle très rapidement les cellules qui ne sont pas adaptées à la survie à l'extérieur d'un hôte, à l'exception des bactéries dotées de capacités particulières de résistance comme la sporulation. C'est pourquoi l'île de Gruinard est restée contaminée plusieurs dizaines d'années après les essais d'épandage de spores de *Bacillus anthracis* menés par les Britanniques pendant la seconde guerre mondiale et pourquoi il a été possible de caractériser la souche de *Bacillus anthracis* vectorisée par l'Iraq pendant le régime baassiste de Saddam Hussein, malgré la destruction des munitions avant les inspections de l'UNSCOM et de l'UNMOVIC. L'épidémie de charbon qui a sévi en Sibérie à l'été 2016 serait également liée à la reviviscence de spores de *Bacillus anthracis* jusqu'alors préservées dans le permafrost, et libérées dans l'environnement du fait du réchauffement climatique. La grande majorité des bactéries n'ont pas ces propriétés de résistance, et il sera très difficile de les détecter dans l'environnement, et surtout de certifier que l'ADN mis en évidence est bien issu d'une souche pathogène manipulée en laboratoire, et non d'une souche sauvage naturellement présente. Des études sont encore à mener pour identifier des marqueurs de manipulation de souches dans une optique de programme proliférant de développement d'armes biologiques, et être capable de discriminer entre souches sauvages, naturelles, et souches cultivées en laboratoire. S'il est identifié des marqueurs particuliers, indiquant une telle mise en culture, il sera alors envisageable de les rechercher sur des échantillons environnementaux, pour mettre en évidence des bactéries pathogènes modifiées.

INDEX DES FIGURES

- Figure 1 : nombre d'articles portant sur l'ADN ancien entre 1984 et 2015. Source Pubmed, avec une requête simple « *ancien*[All Fields] AND "*dna*"[MeSH Terms] OR *adna*[All Fields] ». 17
- Figure 2 : distribution de la taille des fragments amplifiés à partir d'un os de mammouth. Les banques ont été préparées à partir de la technique double-brin (en bleu) et simple-brin (en rouge). D'après Bennett et al. [48]. 20
- Figure 3 : distribution de la fréquence des substitutions de base dans un ADN ancien (ici Neandertal). D'après Briggs et al. [47]. 20
- Figure 4 : cycle de vie de *Bacillus anthracis*. Les spores sont ingérées par l'animal alors qu'il broute. Les spores se transforment alors en formes végétatives qui causent une infection fulminante. Après la mort de l'hôte, la décomposition et l'action des charognards, entraîne le relargage de sang infecté dans l'environnement. A l'extérieur de l'organisme, les bactéries sporulent dans l'attente d'un nouvel hôte. Adapté de Schuch et al. [64]. 23
- Figure 5 : carte de diffusion de la peste noire entre 1347 et 1351 en Europe. D'après [68] 25
- Figure 6 : dissémination de la maladie lors de la troisième pandémie 26
- Figure 7 : localisation des sites ayant permis la mise en évidence de traces de *Y. pestis* en Europe. Les sites en orange correspondent à la peste justinienne, les sites en jaune à la peste noire, les sites en bleu à la peste noire ou ses résurgences. 27
- Figure 8 : représentation de la phylogénie de *Y. pestis* basée sur Achtman *et al.* [73]. L'arbre est enraciné dans *Y. pseudotuberculosis* et les souches analysées par Haensch sont représentées par des rectangles verts [76]. 29
- Figure 9 : représentation de la phylogénie de *Y. pestis* basée sur 133 SNPs, Les quatre lignées majeures sont représentées et l'arbre est enraciné dans *Y. pseudotuberculosis*. Les triangles orange près du nœud N07 identifient les souches séquencées par Bos *et al.* D'après Cui *et al.* [11] 29
- Figure 10 : représentation de la phylogénie de *Y. pestis*, basée sur Cui et al. [11]. Les quatre lignées majeures sont représentées et l'arbre est enraciné dans *Y. pseudotuberculosis*. En rouge sont localisées les souches responsables de la peste justinienne, en orange celles de la peste noire et en violet celles de la pandémie contemporaine. D'après Harbeck *et al.* [23]. 30
- Figure 11 : scénario de dispersion des souches de peste responsables des trois pandémies. D'après Wagner *et al.* [78]. 30

Figure 12 : prévalence de la tuberculose (phtisie) aux Etats-Unis en 1874. L'intensité de la coloration rouge reflète la morbidité [79].	31
Figure 13 : maladie de Pott, touchant la colonne vertébrale. D'après Evinger <i>et al.</i> [81]	32
Figure 14 : lésions typiques de la lèpre au niveau de la tête : résorption des os du nez et de la partie antérieure de la mâchoire avec perte des dents antérieures. D'après Taylor <i>et al.</i> , 2000 [Taylor].	35
Figure 15 : à gauche, cette gravure sur bois (1496) est une des premières représentations des signes de la syphilis (Wellcome Library). A droite, cette peinture sur bois, représentant un syphilitique, est attribuée à Albrecht Dürer (1471-1528).	36
Figure 16 : vertèbres d'un squelette du 10 ^{ème} – 13 ^{ème} siècle, exhumé en Albanie (d'après Mutolo <i>et al.</i>).	37
Figure 17 : analyse phylogénétique du concaténat des trois gènes amplifiés dans la momie sibérienne (astérisque rouge). Ce concaténat apparaît à la racine de la diversité des souches modernes de variole. D'après [25].	40
Figure 18 : représentation sous forme d'arbre de la diversité microbienne connue à partir du séquençage en profondeur de l'ADNr 16S. Cette représentation illustre la grande diversité observée au sein des différents phyla bactériens et la mise en évidence de nouveaux phyla (SAR406 (Marine Group A), OP3, OP8, WS1, WS3, BRC1, CD12, EM19, EM3, NKB19 and Oct-Spa1-106) non encore isolés et cultivés. [144]	48
Figure 19 : représentation du cycle taphonomique, montrant les différents cheminements des organismes morts ou des rebus organiques (fèces par exemple) selon qu'ils soient recyclés (et donc détruits) ou soient conservés et entrent dans les étapes ultérieures de la transition entre biosphère et lithosphère. Les filtres sur les flèches indiquent les processus qui modifient les informations biologiques et écologiques entre les différents stades (d'après [154]).	50
Figure 20 : coléoptères fréquemment associés aux cadavres	52
Figure 21 : localisation et composition des trois sites ayant livré des ossements bovins analysés dans ce projet. D'après Putelat [167]	56
Figure 22 : photographies des kystes trouvés dans les tombes 81, 92 et 97 d'Artzenheim [168].	59
Figure 23 : photographies des quatre kystes retrouvés dans la tombe 514 de Kolbsheim [170]	60
Figure 24 : calcification de Boulogne-sur-Mer <i>in situ</i> (O. Blamangin et S. Kacki)	61

Figure 25 : Calcifications découvertes dans une tombe de Boulogne-sur-Mer (O. Blamangin et S. Kacki).	61
Figure 26 : Sépulture S. 180 du site de Saint-Laurent-de-la-Cabrerisse (A. Gaillard et S. Kacki).	62
Figure 27 : Calcification biologique retrouvée au niveau du thorax de l'individu de la sépulture S. 180 (A : vue externe / B : vue interne) (A. Gaillard et S. Kacki).	63
Figure 28 : photographie en gros plan de la zone calcifiée blanchâtre	63
Figure 29 : représentation du Mary-Rose dans le rôle d'Anthony, registre répertoriant les navires de la marine d'Angleterre au 16 ^{ème} siècle. Cette image est la seule du Mary-Rose après son recarénage et montre les sabords ouverts proches de la ligne d'eau (adapté de [171]).	65
Figure 30 : formule chimique du cyclophosphamide.....	68
Figure 31 : exemple de cages métalliques préparées pour l'expérimentation et disposition de ces cages	68
Figure 32 : plan du site d'expérimentation, avec focus sur la disposition dans une fosse et indication de la localisation du capteur de température. A gauche, disposition des cages dans les fosses d'inhumation et en bas, zone d'expérimentation.....	69
Figure 33 : Corrélation entre taille du fragment d'ADN et quantité initiale d'ADN dans des os ayant subi des traitements taphonomiques différents (résultats non publiés)	71
Figure 34 : milieu sélectif Cereus-Ident-Agarensemencé avec <i>B. cereus</i> [183]	74
Figure 35 : stratégie de prélèvement sur échantillon sec (<i>i.e.</i> os) et de sédiments accompagnant ces os, afin de chercher à mettre en évidence des spores de <i>B. anthracis</i>	75
Figure 36 : structure secondaire de l'ARNr 16S	76
Figure 37 : séquence consensus du gène de l'ADNr 16S de <i>E. coli</i> , établi à partir des 7 opérons 16S présents dans la souche K12 - MG1655. Les 9 zones hypervariables sont de couleur rouge, les amorces de couleur verte et les amplicons de couleur violette (V5) et gris (V3-4-5).	77
Figure 38 : détection du plasmide pXO1 dans le fémur et un muscle de la cuisse d'une souris exposée au bacille du charbon (données non publiées).	85
Figure 39 : évolution de la température dans le sol au cours des quinze premières semaines de la première expérimentation.	87

- Figure 40 : aspect des souris après 15 jours d'inhumation dans un sol froid. A gauche deux souris dans la cage de contention, à droite une souris après ouverture de la cage. 88
- Figure 41 : évolution de la température dans le sol et à la surface au cours de la seconde expérimentation. Les fortes variations de températures pour l'air (courbe rouge en pointillé) étaient liées à la fréquence de prélèvement, biquotidienne (6H-18H)..... 89
- Figure 42 : dispositif expérimental et évolution de l'aspect des souris au cours des 56 jours de l'expérimentation. Les souris laissées à la surface sont à gauche et les souris enfouies à droite..... 92
- Figure 43 : le cliché A montre la section de l'humérus, où l'on voit les chairs entourant encore l'os. Le cliché B montre la surface de l'os, avec un canal haversien (flèche rouge) et des ostéoblastes (clichés : Yolanda Fernandez-Jalvo)..... 94
- Figure 44 : mandibule analysée après six mois d'enfouissement. Les flèches rouges identifient des zones où les attaques bactériennes étaient visibles. Le cliché B, au grossissement x12.000, met en évidence des bactéries à la surface des zones érodées (clichés : Yolanda Fernandez-Jalvo)..... 94
- Figure 45 : ossements analysés après douze mois d'enfouissement. Le cliché A montre la corrosion des articulations et le cliché B l'érosion observée sur les crêtes osseuses. A un plus fort grossissement, on peut observer sur le cliché C la présence de racines à l'intérieur des os, os qui étaient couverts de bactéries et de spores fongiques (cliché D) (clichés : Yolanda Fernandez-Jalvo). 95
- Figure 46 : ossements analysés après dix-huit mois d'enfouissement. Le cliché A montre la corrosion d'un os pelvien et le cliché B la prolifération de microbes à la surface de cet os. A un plus fort grossissement (flèches noires sur le cliché C), on peut observer la colonisation des canaux osseux par des bactéries, plus visibles sur le cliché D (agrandissement de la zone indiquée par l'une des flèches) (clichés : Yolanda Fernandez-Jalvo)..... 96
- Figure 47 : clichés de l'invasion bactérienne après 24 mois (clichés A et B) et 30 mois (clichés C et D) d'enfouissement (clichés : Yolanda Fernandez-Jalvo)..... 97
- Figure 48 : évolution de la quantité d'ADN bactérien dans les os en ng/ μ L calculé par gramme d'os. La figure A correspond au fragment V5 et la figure B au fragment V345. L'axe secondaire indique les valeurs de température, rendues sous forme d'une courbe rouge. 98
- Figure 49 : évolution du pH et des populations bactériennes dans un sol témoin (courbes rouges) et à proximité immédiate de souris (courbes bleues). L'axe principal suit une échelle logarithmique.....100

Figure 50 : évolution de la quantité brute d'ADN extrait à partir d'échantillons de sol au contact de la souris (courbe rouge) et de sol témoin (courbe bleue).....	101
Figure 51 : courbes d'amplification d'échantillons murins avec le couple d'amorces Rd73.....	102
Figure 52 : courbes d'amplification d'échantillons murins avec le couple d'amorces V5	102
Figure 53 : Evolution du nombre de molécules d'ADN murin de taille variable en fonction du temps d'enfouissement. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb	102
Figure 54 : Evolution du nombre de molécules d'ADN murin de taille variable en fonction du temps laissé à la surface. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb	102
Figure 55 : évolution du nombre de molécules d'ADN présentes en fonction du temps, dans les fémurs enfouis. Bleu : J0, rouge J+7, Bleu clair : J+28, vert : J+42 et orange : J+56. Les équations des courbes ont été données dans la même couleur.....	103
Figure 56 : évolution du nombre de molécules d'ADN présentes en fonction du temps, dans les fémurs des souris restées en surface. . Bleu : J0, rouge J+7, Bleu clair : J+28, vert : J+42 et orange : J+56. Les équations des courbes ont été données dans la même couleur.....	103
Figure 57 : évolution de la concentration d'ADN murin dans le sol au contact des cadavres de souris au cours du temps. La quantité d'ADN est mesurée par deux marqueurs, Rd73 (courbe bleue) amplifiant un fragment de 73 bp, et Rd125 (courbe rouge) amplifiant un fragment de 125 bp.	104
Figure 58 : Résultats du séquençage des banques créées à partir des extraits osseux et des fèces. Le nombre d'OTUs est donné après élimination des séquences présentes en une seule copie (singleton), au niveau taxonomique du genre (niveau le plus bas obtenu avec usearch). FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; FT : Fémur Témoin. F-xxx : banques obtenues à partir de fèces.....	106
Figure 59 : Représentation en PCA des deux populations amplifiées, par V5 (points verts) et par V345 (points gris).....	107
Figure 60 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau du phylum. Les extraits de fémur ont été regroupés deux à deux en fonction du délai de décomposition. Un échantillon de sol témoin est représenté à droite et de fèces à gauche. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin.....	108

- Figure 61 : séparation par PCA (à gauche) et par PLS-DA (à droite) des différentes populations analysées, au niveau du phylum, à partir des amplicons HVR5. En jaune le sol contrôle, en rouge le sol au contact des souris, en vert les fèces, en gris-bleu les os enfouis et en violet les os en surface. (1) : J7 ; (2) : J42 ; (3) : J42-J56.....109
- Figure 62 : représentation de la diversité bactérienne déduite des amplicons V345, au niveau du phylum. Les extraits de fémur ont été regroupés deux à deux en fonction du délai de décomposition. Un échantillon de sol témoin est représenté à droite. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin.....110
- Figure 63 : séparation par PCA (à gauche) et par PLS-DA (à droite) des différentes populations analysées, au niveau du phylum, à partir des HVR 3, 4 et 5. En rouge le sol contrôle, en violet le sol au contact des souris, en vert les os enfouis et en gris les os en surface.....111
- Figure 64 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification du fragment hypervariable V5, au sein des populations en surface et enterrées. Un échantillon de sol témoin est représenté à droite et de fèces à gauche. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin.....112
- Figure 65 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification d'un fragment englobant les régions hypervariables 3, 4 et 5, au sein des populations en surface et enterrées. Seules les 15 classes majoritaires ont été représentées. Un échantillon de sol témoin est représenté à droite. FSE : Fémur Souris Enfouie ; FSS : Fémur Souris Surface ; ST : Sol Témoin 112
- Figure 66 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification du fragment hypervariable 5, au sein des sols témoin et contact. Seules les 15 classes majoritaires ont été représentées. Un échantillon de faeces témoin est représenté à droite. ST : Sol Témoin ; SS : Sol Souris.113
- Figure 67 : évolution des populations bactériennes au cours du temps *post-mortem*, inférées par l'amplification d'un fragment englobant les régions hypervariables 3, 4 et 5, au sein des sols témoin et contact. Seules les 15 classes majoritaires ont été représentées. ST : Sol Témoin ; SS : Sol Souris.114
- Figure 68 : évolution des populations bactériennes dans le sol témoin, à partir des amplicons du fragment HVR5. Les 15 phyla majoritaires ont été représentés. ST : Sol Témoin ; SS : Sol Souris.....115
- Figure 69 : évolution des populations bactériennes dans le sol au contact des souris, à partir des amplicons du fragment HVR5. Les 15 phyla majoritaires ont été représentés. ST : Sol Témoin ; SS : Sol Souris.116

Figure 70 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols témoins. Les assignations ont été faites à partir des amplicons du fragment HVR5. ST : Sol Témoin.....	116
Figure 71 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols contact. SS : Sol Souris.....	117
Figure 72 : pourcentage des séquences assignées au niveau du genre, au sein de l'ordre des pseudomonadales, à partir des amplicons du fragment HVR5. Les ordres sélectionnés représentent 85% des séquences assignées pour le sol témoin et le sol contact. Les valeurs en rouge indiquent les populations actives, c'est-à-dire présentant un ration ARN/ADN supérieur à 1.....	118
Figure 73 : composition du phylum des protéobactéries au niveau taxonomique de l'ordre, en conservant les 15 principaux ordres présents dans les sols témoins. Les assignations ont été faites à partir de l'amplicon des fragments HVR 3, 4 et 5. SS : Sol Souris.....	119
Figure 74 : A : séparation par PCA (A et C) et par PLS-DA (B et D) des différentes populations analysées, au niveau du phylum, à partir des HVR 5 (A, B) et HVR 345 (C, D). En rouge le sol contrôle, en violet le sol au contact des souris, en vert les os enfouis et en gris les os en surface.....	120
Figure 75 : évolution de la température dans le sol entre janvier 2013 et juillet 2015 .	121
Figure 76 : photographies des différents états de dégradation des dépouilles de souris au fur et à mesure du temps.....	122
Figure 77 : planches de fétuque rouge (<i>Festuca rubra</i>)	123
Figure 78 : évolution du pH et des populations bactériennes dans un sol témoin (courbes vert et violet) et à proximité immédiate de souris (courbes bleu et rouge). L'axe principal suit une échelle logarithmique.....	127
Figure 79 : Evolution du nombre de molécules d'ADN de taille variable en fonction du temps d'enfouissement. Vert : 73pb, Bleu : 125pb, rouge : 154pb et violet : 513pb	128
Figure 80 : évolution du nombre de molécules présentes en fonction du temps, dans les fémurs enfouis. Bleu : J0, brique : J+15 jours, violet J+3 mois, bleu turquoise: J+6 mois, vert : J+9 mois, orange : J+12 mois, bleu clair : J+18 mois, rouge : J+24 mois.....	129
Figure 81 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau du phylum. Les extraits de sol ont été regroupés deux à deux en fonction du	

délai de décomposition. Les 10 phyla majoritaires ont été représentés et la mesure du sol témoin de janvier 2014 est manquante. ST : Sol Témoin ; SS : Sol Souris.	132
Figure 82 : représentation de la diversité bactérienne déduite des amplicons V5, au niveau des protéobactéries. Les extraits de sol ont été regroupés deux à deux en fonction du délai de décomposition. ST : Sol Témoin ; SS : Sol Souris.	133
Figure 83 : représentation de la diversité bactérienne déduite des amplicons V5, au sein des Gammaprotéobactéries. ST : Sol Témoin ; SS : Sol Souris.	134
Figure 84 : représentation des résultats de l'analyse des séquences identifiées comme humaines dans un échantillon ancien (kyste de Boulogne-Sur-Mer). Les quatre premiers graphes représentent les fréquences relatives des quatre bases dans la séquence identifiée et dans son environnement dans la séquence de référence. Les deux derniers graphes représentent les fréquences de mis-appariement selon les bases par rapport aux extrémités 5' (à gauche) et 3' (à droite) des séquences amplifiées et appariées au génome de référence.	135
Figure 85 : représentation des résultats de l'analyse des séquences identifiées comme bactériennes dans un échantillon ancien (kyste de Boulogne-Sur-Mer).....	136
Figure 86 : évolution de la concentration en ADN de <i>B. atrophaeus</i> dans les différents tissus au cours du temps. Bleu : extrait de fémur, rouge : extrait de mâchoire, vert : extrait de queue.....	137
Figure 87 : évolution de la concentration en ADN de <i>X. arboricola</i> dans les différents tissus au cours du temps. Bleu : extrait de fémur, rouge : extrait de mâchoire, vert : extrait de queue.....	138
Figure 88 : courbes d'amplification et de fusion obtenues à partir d'amorces purifiées selon les différents protocoles d'Eurofins. Les produits d'amplification ont été séquencés selon la méthode Sanger.	141
Figure 89 : procédure de synthèse des oligonucléotides selon la méthode dite Phosphoramidite [207]	142
Figure 90 : courbes d'amplification et de fusion obtenues à partir du couple d'amorces purifiées ciblant le fragment V5. Le produit de séquençage obtenu est pur et correspond à <i>P. alcaligenes</i>	143
Figure 91 : effets de concentrations croissantes d'EMA (à gauche) et de PMA (à droite) sur l'efficacité de l'amplification du fragment V5 de l'ADNr16S. Les équivalents génomes sont en échelle logarithmique.	145
Figure 92 : évolution des Ct d'apparition des produits d'amplification sur des échantillons négatifs, en fonction de la concentration des deux intercalants.....	145

- Figure 93 : écouvillons secs et prélèvements de sédiments mis en culture au LSB3 de DGA Maîtrise NRBC.....148
- Figure 94 : sur le milieu sélectif (à gauche), il a été possible d'observer quelques colonies blanches sur cette culture issue d'un sédiment. L'examen à l'état frais d'une de ces colonies (à droite) a révélé des structures en chaînette.....148
- Figure 95 : localisation des échantillons humains utilisés dans cette étude. Les restes humains anciens ont été excavés de quatre sites, Artzenheim (1) et Kolbsheim (2) sont en Alsace, Boulogne-sur-Mer (3) dans le Pas-de-Calais et Saint-Laurent de la Cabrerisse (4) dans les Pyrénées-Orientales. Carte dessinée avec mapcustomizer (mapcustomizer.com).....151
- Figure 96 : distribution de la proportion de séquences identifiées comme d'origine humaine, selon l'origine tissulaire de la banque dans chaque échantillon. Les os pétreux étaient absents des assemblages pour BSM et SLC-180.....152
- Figure 97: distribution des quantités de séquences alignées sur les chromosomes selon la longueur de ceux-ci en paires de base. La ligne en pointillé indique la corrélation linéaire de ces points (droite de tendance). Le ratio pour le chromosome X a été indiqué en orange et celui du chromosome Y en rouge. Pour tous les échantillons sauf Artz-ST81, le ratio du chromosome X était proche ou très proche de la droite de tendance, indiquant la présence de deux chromosomes X et donc que l'individu était de sexe féminin. Pour Arzt-St81, le ratio est deux fois inférieur à la moyenne, pour le chromosome X et pour le chromosome Y, indiquant la présence d'un seul exemplaire de chaque chromosome, et donc que l'individu était de sexe mâle. On note que pour tous les autres échantillons, le ratio pour le chromosome Y était très proche de 0, indicatif d'un sexe féminin.154
- Figure 98: Distribution des phyla bactériens dans les différentes parties des échantillons soumises à analyse, et visualisée par l'algorithme Krona tool [201]. La légende est identique pour l'ensemble des graphiques, permettant une comparaison entre les différents échantillons. Nous ne disposons pas d'os pétreux pour les échantillons de SLC et de BSM, et seulement du tartre pour ce dernier. La dent de l'échantillon SLC-180 avait été analysée avant notre intérêt pour le tartre. Il n'a pu être analysé rétrospectivement.....155
- Figure 99: *Heatmap* obtenue grâce à *MetaPhlan* (mode Very Sensitive Local pour Bowtie2) [199] représentant les 100 genres bactériens les plus représentés au sein des échantillons analysés. Dans la partie supérieure gauche de la *heatmap*, il est observé un cluster de genres bactériens communément présents dans la communauté bactérienne buccale, en bonne santé ou à l'hygiène déficiente. Ce cluster correspondait aux échantillons de tartre analysés ici. Les kystes et les os présentaient les mêmes genres bactériens, beaucoup plus divers, liés à la grande diversité de bactéries présentes dans les sols.....156

- Figure 100 : distribution des ratios nombre de séquences/longueur de chromosome pour chacun des échantillons analysés. Le ratio pour le chromosome X est représenté en noir et le ratio pour le chromosome Y en rouge. Il est observé une bonne corrélation pour tous les autosomes, quel que soit l'échantillon, même avec un nombre de séquences restreint. Seul MR2037, avec seulement 289 séquences humaines, présente une mauvaise distribution et des autosomes qui ont un ratio comparable aux chromosomes sexuels.....161
- Figure 101 : représentation sous forme de *heatmap* des 50 espèces bactériennes majoritaires identifiées dans les différents échantillons de tartre analysés, comparés entre elles et à un échantillon externe (os costal, à gauche). Le groupe sain est repéré par les orbes vertes, le groupe intermédiaire par l'orbe jaune et le groupe pathologique par l'orbe orange.....163
- Figure 102 : dentiste en uniforme arrachant une dent à un marin à la mâchoire déjà dégarnie. Lithographie colorée, Wellcome Library (18^{ème} siècle (?)).....164
- Figure 103 : dendrogramme de l'espèce *T. denticola*, calculé à partir des données de BLAST sur le site NCBI, consulté le 28/08/2016. Les génomes assemblés sont entourés en violet, tandis que les souches les plus éloignées sont identifiées par un cadre rouge.....167
- Figure 104 : représentation graphique sous forme d'arbre couvrant minimal (MST, Minimum Spanning Tree) des relations phylogénétiques inférées des SNPs identifiés entre les différentes souches et séquences analysées de *Treponema denticola*. Les souches publiées sont représentées en violet. La souche OMZ_758 appartient à l'espèce *T. putidum*, espèce la plus proche de *T. denticola* [223]. La couleur des autres souches est corrélée à leur positionnement dans les clusters hautement pathologique (orange) ou sain (vert). Le nombre de SNPs séparant les souches entre elle est indiqué sur chaque branche. La ligne en pointillés n'est pas proportionnelle au nombre de SNPs.168
- Figure 105 : représentation graphique sous forme d'arbre couvrant minimal (MST, Minimum Spanning Tree) des relations phylogénétiques inférées des SNPs identifiés entre les différentes souches et séquences analysées de *Tannerella forsythia*. Les souches publiées sont représentées en violet, La couleur des autres souches est corrélée à leur positionnement dans les clusters pathologique (jaune) et hautement pathologique (orange). Le nombre de SNPs séparant les souches entre elle est indiqué sur chaque branche.170

INDEX DES TABLEAUX

Tableau 1 : liste des échantillons analysés au cours de l'étude.....	58
Tableau 2 : sélection des couples d'amorces amplifiant des fragments de taille croissante dans les souris et les deux bactéries sélectionnées.	72
Tableau 3 : liste des cibles et amorces utilisées pour la détection par qPCR de <i>M. tuberculosis</i>	73
Tableau 4 : liste des échantillons réalisés à partir des prélèvements fournis par O. Putelat sur les ossements de la porte de Bourgogne.....	74
Tableau 5 : liste des couples d'amorces testés via TestPrime afin de déterminer les meilleures associations. A noter que le couple E343F-W926R a été testé sur SSU r123 RefNR, riche de 500.769 séquences bactériennes.....	78
Tableau 6 : liste des amorces analysées pour déterminer les meilleurs couples. Les bases dégénérées sont repérées par leur couleur rouge. La localisation des séquences est donnée par leur position 5' sur <i>E. coli</i>	79
Tableau 7 : ADN total et équivalents génomes des bactéries injectées. ND : Non déterminé. En vert les échantillons négatifs pour les agents injectés.	86
Tableau 8 : dosage de <i>B. atrophaeus</i> et <i>X. arboricola</i> dans les échantillons de souris immunodéprimées	87
Tableau 9 : valeurs d'index histologique d'Oxford correspondant au degré de modification diagénétique (adapté de [205]).	93
Tableau 10 : données de dénombrement sur les prélèvements issus du second protocole. ND : non déterminé. Ind : indénombrable, le nombre de colonies étant trop élevé pour être compté.....	99
Tableau 11 : pourcentage des séquences assignées, au sein du phylum protéobactéries, à partir des amplicons du fragment HVR5. Les ordres sélectionnés représentent 85% des séquences assignées pour le sol témoin et le sol contact. Les cellules grisées représentent les ordres minoritaires dans le sol témoin, qui ont fortement crû dans le sol au contact des cadavres. Les valeurs en rouge indiquent les populations actives.	118
Tableau 12 : données de dénombrement sur les prélèvements issus du premier protocole. ND : non déterminé. Ind : indénombrable, le nombre de colonies étant trop élevé pour être compté.....	126

Tableau 13 : résultats du séquençage des banques créées à partir des extraits osseux et des fèces. Le nombre d'OTUs est donné après élimination des séquences présentes en une seule copie (singleton), au niveau taxonomique du genre (niveau le plus bas obtenu avec usearch). Les valeurs aberrantes (longueur et taux de GC) ont été repérées en rouge.....	131
Tableau 14 : liste des échantillons issus de Bure, Bourogne, Vellechevreux et Birgelsgaerten, avec le résultat de l'amplification par qPCR d'un fragment de l'ADN mitochondrial et l'assignation d'espèce déduite de cette séquence,.....	147
Tableau 15 : liste des échantillons analysés pour les sites de Horbourg et Bernolsheim. Un squelette peut être analysé à plusieurs reprises avec des os différents, voire des extraits différents.....	150
Tableau 16 : synthèse des données obtenues lors du séquençage des tartres du Mary-Rose, et effet des processus de filtrage des données sur la quantité finale de séquences analysées,.....	158
Tableau 17 : synthèse des valeurs obtenues à partir de l'alignement des séquences sur le génome humain de référence. Trois échantillons présentent un très faible nombre de séquences humaines, et un pourcentage très bas (MR2017, MR2037 et MR2049). Le grand nombre de séquences obtenues compense cette rareté, sauf pour MR2037 pour lequel il n'est obtenu que 289 séquences humaines. Le plus fort taux est de 4,44% et le taux moyen de séquences humaines 0,74%).	160
Tableau 18 : cartographie des séquences des différents échantillons sur les génomes humain et des trois bactéries du complexe rouge. La longueur des barres dans les cellules est proportionnelle à la quantité de séquences rapportées aux échantillons entre eux. Les échantillons modernes présentent la proportion majoritaire de séquences cartographiées sur le génome humain, car prélevés sur des patients vivants. Les couleurs des échantillons sont corrélées à leur regroupement dans la représentation en <i>heatmap</i>	165
Tableau 19 : rapport d'analyse de similarité entre les séquences disponibles sur le site www.ncbi.nlm.nih.gov consulté le 28/08/2016, comparées à la souche de référence ATCC 35405.....	166

Références Bibliographiques

1. Higuchi, R., et al., *DNA sequences from the quagga, an extinct member of the horse family*. Nature, 1984. **312**(5991): p. 282-4.
2. Paabo, S., *Molecular cloning of Ancient Egyptian mummy DNA*. Nature, 1985. **314**(6012): p. 644-5.
3. Woodward, S.R., N.J. Weyand, and M. Bunnell, *DNA sequence from Cretaceous period bone fragments*. Science, 1994. **266**(5188): p. 1229-32.
4. Green, R.E., et al., *A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing*. Cell, 2008. **134**(3): p. 416-26.
5. Meyer, M., et al., *A high-coverage genome sequence from an archaic Denisovan individual*. Science, 2012. **338**(6104): p. 222-6.
6. Aufderheide, A.C. and C. Rodriguez-Martin, *The Cambridge Encyclopedia of Paleopathology*, ed. C.U. Press. 2011. 498.
7. Ruffer, M.A., *Arthritis deformans and spondylitis in ancient Egypt*. Journal of Pathology and Bacteriology, 1918. **22**: p. 212-226.
8. Thucydide, *Peloponnese Wars - 2nd Book*. 411 BC.
9. Papagrigorakis, M.J., et al., *DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens*. Int J Infect Dis, 2006. **10**(3): p. 206-14.
10. Tours, G.d., *Historia Francorum - Liber IV*. 594.
11. Boccace, *Decameron*. 1349-1353.
12. Schmid, B.V., et al., *Climate-driven introduction of the Black Death and successive plague reintroductions into Europe*. Proc Natl Acad Sci U S A, 2015. **112**(10): p. 3020-5.
13. Spigelman, M. and E. Lemma, *The Use of the Polymerase Chain Reaction (PCR) to detect Mycobacterium tuberculosis in Ancient Skeletons*. International Journal of Osteoarchaeology, 1993. **3**: p. 137-143.
14. Raoult, D., et al., *Molecular identification by "suicide PCR" of Yersinia pestis as the agent of medieval black death*. Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12800-3.
15. Theves, C., et al., *Molecular identification of bacteria by total sequence screening: determining the cause of death in ancient human subjects*. PLoS One, 2011. **6**(7): p. e21733.
16. Haas, C.J., et al., *Detection of Leprosy in Ancient Human Skeletal Remains by Molecular Identification of Mycobacterium leprae*. American Journal of Clinical Pathology, 2000. **114**: p. 428-436.
17. Zink, A., et al., *Corynebacterium in ancient Egypt*. Med Hist, 2001. **45**(2): p. 267-72.
18. Miller, R.L., et al., *Diagnosis of Plasmodium falciparum infections in mummies using the rapid manual ParaSight-F test*. Trans R Soc Trop Med Hyg, 1994. **88**(1): p. 31-2.
19. Nerlich, A.G., et al., *Plasmodium falciparum in ancient Egypt*. Emerg Infect Dis, 2008. **14**(8): p. 1317-9.
20. Tsangaras, K. and A.D. Greenwood, *Museums and disease: using tissue archive and museum samples to study pathogens*. Ann Anat, 2012. **194**(1): p. 58-73.
21. Bos, K.I., et al., *A draft genome of Yersinia pestis from victims of the Black Death*. Nature, 2011. **478**(7370): p. 506-10.
22. Schuenemann, V.J., et al., *Genome-wide comparison of medieval and modern Mycobacterium leprae*. Science, 2013. **341**(6142): p. 179-83.

23. Harbeck, M., et al., *Yersinia pestis* DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. PLoS Pathog, 2013. **9**(5): p. e1003349.
24. Tumpey, T.M., et al., *Characterization of the reconstructed 1918 Spanish influenza pandemic virus*. Science, 2005. **310**(5745): p. 77-80.
25. Biagini, P., et al., *Variola virus in a 300-year-old Siberian mummy*. N Engl J Med, 2012. **367**(21): p. 2057-9.
26. Saiki, R.K., et al., *Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia*. Science, 1985. **230**(4732): p. 1350-4.
27. Hofreiter, M., et al., *DNA sequences from multiple amplification reveal artifacts induced by cytosine deamination in ancient DNA*. Nucleic Acids Res, 2001. **29**(23): p. 4793-4799.
28. Paabo, S., et al., *Genetic analyses from ancient DNA*. Annu Rev Genet, 2004. **38**: p. 645-79.
29. Matheus, P., et al., *Pleistocene brown bears in the mid-continent of North America*. Science, 2004. **306**(5699): p. 1150.
30. Bunce, M., et al., *Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle*. PLoS Biol, 2005. **3**(1): p. e9.
31. Shapiro, B., et al., *Rise and fall of the Beringian steppe bison*. Science, 2004. **306**(5701): p. 1561-5.
32. Prost, S., et al., *Influence of climate warming on arctic mammals? New insights from ancient DNA studies of the collared lemming *Dicrostonyx torquatus**. PLoS One, 2010. **5**(5): p. e10447.
33. Willerslev, E., et al., *Ancient biomolecules from deep ice cores reveal a forested southern Greenland*. Science, 2007. **317**(5834): p. 111-4.
34. Poinar, H.N., et al., *Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA*. Science, 2006. **311**(5759): p. 392-4.
35. Rasmussen, M., et al., *Ancient human genome sequence of an extinct Palaeo-Eskimo*. Nature, 2010. **463**(7282): p. 757-62.
36. Green, R.E., et al., *A draft sequence of the Neandertal genome*. Science, 2010. **328**(5979): p. 710-22.
37. Reich, D., et al., *Genetic history of an archaic hominin group from Denisova Cave in Siberia*. Nature, 2010. **468**(7327): p. 1053-60.
38. Hershkovitz, I., et al., *Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean*. PLoS One, 2008. **3**(10): p. e3426.
39. Drancourt, M., et al., *Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: An approach to the diagnosis of ancient septicemia*. PNAS, 1998. **95**: p. 12637-12640.
40. Bos, K.I., et al., *Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis*. Nature, 2014. **514**(7523): p. 494-497.
41. Willerslev, E., et al., *Diverse plant and animal genetic records from Holocene and Pleistocene sediments*. Science, 2003. **300**(5620): p. 791-5.
42. Haile, J., et al., *Ancient DNA typing of archaeological pig remains corroborates historical records*. Journal of Archaeological Science, 2010. **37**(1): p. 174-177.
43. Romanowski, G., M.G. Lorenz, and W. Wackernagel, *Adsorption of plasmid DNA to mineral surfaces and protection against DNase I*. Appl Environ Microbiol, 1991. **57**(4): p. 1057-61.

44. Ogram, A., et al., *DNA adsorption to soils and sediments*. Environ Sci Technol, 1988. **22**(8): p. 982-4.
45. Cai, P., Q.Y. Huang, and X.W. Zhang, *Interactions of DNA with Clay Minerals and Soil Colloidal Particles and Protection against Degradation by DNase*. Environ Sci Technol, 2006. **40**: p. 2971-2976.
46. Smith, C.I., et al., *Neanderthal DNA. Not just old but old and cold?* Nature, 2001. **410**(6830): p. 771-2.
47. Briggs, A.W., et al., *Patterns of damage in genomic DNA sequences from a Neandertal*. Proc Natl Acad Sci U S A, 2007. **104**(37): p. 14616-21.
48. Bennett, E.A., et al., *Library construction for ancient genomics: single strand or double strand?* Biotechniques, 2014. **56**(6): p. 289-90, 292-6, 298, passim.
49. Ginolhac, A., et al., *mapDamage: testing for damage patterns in ancient DNA sequences*. Bioinformatics, 2011. **27**(15): p. 2153-5.
50. Jonsson, H., et al., *mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters*. Bioinformatics, 2013. **29**(13): p. 1682-4.
51. Cooper, A. and H.N. Poinar, *Ancient DNA: do it right or not at all*. Science, 2000. **289**(5482): p. 1139.
52. Charruau, P., et al., *Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates*. Mol Ecol, 2011. **20**(4): p. 706-24.
53. Geigl, E.-M., *Palaeogenetics of cattle domestication: Methodological challenges for the study of fossil bones preserved in the domestication centre in Southwest Asia*. Comptes Rendus Palevol, 2008. **7**(2-3): p. 99-112.
54. Pruvost, M., et al., *Freshly excavated fossil bones are best for amplification of ancient DNA*. Proc Natl Acad Sci U S A, 2007. **104**(3): p. 739-44.
55. Geigl, E.M. and T. Grange, *Eurasian wild asses in time and space: morphological versus genetic diversity*. Ann Anat, 2012. **194**(1): p. 88-102.
56. Pruvost, M., et al., *Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art*. Proc Natl Acad Sci U S A, 2011. **108**(46): p. 18626-30.
57. Champlot, S., et al., *An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications*. PLoS One, 2010. **5**(9).
58. Pruvost, M. and E.-M. Geigl, *Real-time quantitative PCR to assess the authenticity of ancient DNA amplification*. Journal of Archaeological Science, 2004. **31**(9): p. 1191-1197.
59. Pruvost, M., T. Grange, and E.M. Geigl, *Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples: application to ancient DNA studies*. Biotechniques, 2005. **38**(4): p. 569-75.
60. Meyer, M., et al., *Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins*. Nature, 2016. **531**(7595): p. 504-7.
61. Meyer, M., et al., *A mitochondrial genome sequence of a hominin from Sima de los Huesos*. Nature, 2014. **505**(7483): p. 403-6.
62. Guimaraes, S., et al., *A cost-effective high-throughput metabarcoding approach powerful enough to genotype ~44 000 year-old rodent remains from Northern Africa*. Mol Ecol Resour, 2016.
63. Dabney, J., et al., *Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments*. Proc Natl Acad Sci U S A, 2013. **110**(39): p. 15758-63.

64. Schuch, R. and V.A. Fischetti, *The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations*. PLoS One, 2009. **4**(8): p. e6532.
65. Van Ert, M.N., et al., *Global genetic population structure of Bacillus anthracis*. PLoS One, 2007. **2**(5): p. e461.
66. Achtman, M., et al., *Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis*. PNAS, 1999. **96**(24): p. 14043-14048.
67. Ziegler, P., *The Black Death*, ed. A.S.P. Inc. 1999, Wolfeboro Falls, NH. 312.
68. Benedictow, O.J., *The Black Death 1346-1353: the complete history*, ed. B. Press. 2004.
69. Devignat, R., *Variétés de l'espèce Pasteurella pestis*. Bulletin World Health Organisation, 1951. **4**: p. 247-263.
70. Drancourt, M., et al., *Genotyping, Orientalis-like Yersinia pestis, and plague pandemics*. Emerg Infect Dis, 2004. **10**(9): p. 1585-92.
71. Wiechmann, I. and G. Grupe, *Detection of Yersinia pestis DNA in two early medieval skeletal finds from Aschheim (Upper Bavaria, 6th century A.D.)*. Am J Phys Anthropol, 2005. **126**(1): p. 48-55.
72. Drancourt, M., et al., *Yersinia pestis Orientalis in Remains of Ancient Plague Patients* Emerg Infect Dis, 2007. **13**(2): p. 332-333.
73. Achtman, M., et al., *Microevolution and history of the plague bacillus, Yersinia pestis*. Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17837-42.
74. Morelli, G., et al., *Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity*. Nat Genet, 2010. **42**(12): p. 1140-3.
75. Schuenemann, V.J., et al., *Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death*. Proc Natl Acad Sci U S A, 2011. **108**(38): p. E746-52.
76. Haensch, S., et al., *Distinct clones of Yersinia pestis caused the black death*. PLoS Pathog, 2010. **6**(10): p. e1001134.
77. Vergnaud, G., *Yersinia pestis genotyping*. Emerg Infect Dis, 2005. **11**(8): p. 1317-8; author reply 1318-9.
78. Wagner, D.M., et al., *Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis*. The Lancet Infectious Diseases, 2014. **14**(4): p. 319-326.
79. Onion, R. *19th-Century Maps Tracking Major Diseases Across the United States*. 2014 [cited 2016 03-08]; Available from: http://www.slate.com/blogs/the_vault/2014/04/24/medical_maps_sidney_carnegie_maps_of_major_diseases.html.
80. Rodriguez-Campos, S., et al., *Overview and phylogeny of Mycobacterium tuberculosis complex organisms: implications for diagnostics and legislation of bovine tuberculosis*. Res Vet Sci, 2014. **97** Suppl: p. S5-S19.
81. Evinger, S., et al., *New skeletal tuberculosis cases in past populations from Western Hungary (Transdanubia)*. Homo, 2011. **62**(3): p. 165-83.
82. Bhanu, N.V., et al., *Two mycobacterium fortuitum strains isolated from pulmonary tuberculosis patients in Delhi harbour IS6110 homologue*. Diagn Microbiol Infect Dis, 2004. **48**(2): p. 107-10.
83. Coros, A., E. DeConno, and K.M. Derbyshire, *IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species*. J Bacteriol, 2008. **190**(9): p. 3408-10.

84. Müller, R., C.A. Roberts, and T.A. Brown, *Complications in the study of ancient tuberculosis: non-specificity of IS6110 PCRs*. Science and Technology of Archaeological Research, 2015. **1**(1): p. STAR20141120548.
85. Bouwman, A.S., et al., *Genotype of a historic strain of Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A, 2012. **109**(45): p. 18511-6.
86. Chan, J.Z., et al., *Metagenomic analysis of tuberculosis in a mummy*. N Engl J Med, 2013. **369**(3): p. 289-90.
87. Comas, I., et al., *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans*. Nat Genet, 2013. **45**(10): p. 1176-82.
88. Bryant, J.M., et al., *Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data*. BMC Infectious Diseases, 2013. **13**(110).
89. Namouchi, A., et al., *After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection*. Genome Res, 2012. **22**(4): p. 721-34.
90. WHO, *global leprosy update, 2013; reducing disease burden*. Weekly epidemiological record, 2014. **89**(36): p. 389-400.
91. Rafi, A., et al., *Mycobacterium leprae DNA from ancient bone detected by PCR*. The Lancet, 1994. **343**: p. 1360-1361.
92. Haas, C.J., et al., *Molecular evidence for different stages of tuberculosis in ancient bone samples from Hungary*. Am J Phys Anthropol, 2000. **113**(3): p. 293-304.
93. Montiel, R., et al., *DNA sequences of Mycobacterium lepraerecovered from ancient bones*. FEMS Microbiology Letters, 2003. **226**(2): p. 413-414.
94. Taylor, G.M., et al., *Detection and strain typing of ancient Mycobacterium leprae from a medieval leprosy hospital*. PLoS One, 2013. **8**(4): p. e62406.
95. Taylor, G.M., et al., *A Mediaeval Case of Lepromatous Leprosy from 13–14th Century Orkney, Scotland*. Journal of Archaeological Science, 2000. **27**(12): p. 1133-1138.
96. Schwarz, S., L. Skytte, and K.L. Rasmussen, *Pre-Columbian treponemal infection in Denmark?- a paleopathological and archaeometric approach*. Heritage Science, 2013. **1**(19): p. 1-12.
97. von Hunnius, T.E., et al., *Digging deeper into the limits of ancient DNA research on syphilis*. Journal of Archaeological Science, 2007. **34**(12): p. 2091-2100.
98. Brouqui, P., et al., *Chronic Bartonella quintana bacteremia in homeless patients*. N Engl J Med, 1999. **340**(3): p. 184-9.
99. Drancourt, M., et al., *Bartonella quintana in a 4000-Year-Old Human Tooth*. Journal of Infectious Disease, 2005. **191**: p. 607-611.
100. Geyik, M.F., et al., *Musculoskeletal involvement in brucellosis in different age groups: a study of 195 cases*. Swiss Med Wkly, 2002. **132**: p. 98-105.
101. Kay, G.L., et al., *Recovery of a medieval Brucella melitensis genome using shotgun metagenomics*. MBio, 2014. **5**(4): p. e01337-14.
102. D'Anastasio, R., et al., *Possible brucellosis in an early hominin skeleton from sterkfontein, South Africa*. PLoS One, 2009. **4**(7): p. e6439.
103. Warinner, C., et al., *Pathogens and host immunity in the ancient human oral cavity*. Nat Genet, 2014. **46**(4): p. 336-44.
104. Dewhirst, F.E., et al., *The human oral microbiome*. J Bacteriol, 2010. **192**(19): p. 5002-17.

105. McLean, J.S., *Advancements toward a systems level understanding of the human oral microbiome*. Front Cell Infect Microbiol, 2014. **4**: p. 98.
106. S., S.S. and A.D. Haffajee, *Periodontal microbial ecology*. 2005. **Periodontology** **2000**(38): p. 135-187.
107. Palakuru, S.K., V.K. Lakshman, and K.G. Bhat, *Microbiological analysis of oral samples for detection of Mycobacterium tuberculosis by nested polymerase chain reaction in tuberculosis patients with periodontitis*. Dent Res J, 2012. **9**(6): p. 688-693.
108. Yang, C.J., et al., *Unexpectedly high prevalence of Treponema pallidum infection in the oral cavity of human immunodeficiency virus-infected patients with early syphilis who had engaged in unprotected sex practices*. Clin Microbiol Infect, 2015. **21**(8): p. 787 e1-7.
109. Li, Y., et al., *On the origin of smallpox: Correlating variola phylogenics with historical smallpox records*. PNAS, 2007. **104**(40): p. 15787-15792.
110. Tumpey, T.M., et al., *Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus*. Science, 2005. **310**: p. 77-80.
111. Rasmussen, S., et al., *Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago*. Cell, 2015. **163**(3): p. 571-82.
112. Cui, Y., et al., *Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis*. PNAS, 2013. **110**(2): p. 577-582.
113. Moodley, Y., et al., *Age of the association between Helicobacter pylori and man*. PLoS Pathog, 2012. **8**(5): p. e1002693.
114. Blouin, Y., et al., *Significance of the identification in the Horn of Africa of an exceptionally deep branching Mycobacterium tuberculosis clade*. PLoS One, 2012. **7**(12): p. e52841.
115. Kay, G.L., et al., *Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe*. Nat Commun, 2015. **6**: p. 6717.
116. Allentoft, M.E., et al., *The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils*. Proc Biol Sci, 2012. **279**(1748): p. 4724-33.
117. Orlando, L., et al., *Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse*. Nature, 2013. **499**(7456): p. 74-8.
118. Goncharov, A., et al., *Draft Genome Sequence of Enterococcus faecium Strain 58m, Isolated from Intestinal Tract Content of a Woolly Mammoth, Mammuthus primigenius*. Genome Announc, 2016. **4**(1).
119. Eisenhofer, R., A. Cooper, and L.S. Weyrich, *Isolating Viable Ancient Bacteria: What You Put In Is What You Get Out*. Genome Announc, 2016. **4**(4).
120. Whitman, W., D. Coleman, and W. Wiebe, *Prokaryotes: the unseen majority*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(12): p. 6578-83.
121. Eisen, J., *Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes*. PLoS Biology, 2007. **5**(3): p. e82.
122. Woese, C.R., *Bacterial evolution*. Microbiology Review, 1987. **51**: p. 221-271.
123. Wayne, L.G., et al., *Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics*. International Journal of Systematic Bacteriology, 1987. **37**(4): p. 463-464.
124. Torsvik, V., J. Goksoyr, and F.L. Daae, *High diversity in DNA of soil bacteria*. Appl Environ Microbiol, 1990. **56**(3): p. 782-7.
125. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: the unseen majority*. Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6578-83.

126. Hugenholtz, P., *Exploring prokaryotic diversity in the genomic era*. Genome biology, 2002. **3**(2): p. REVIEWS0003.
127. Pace, N.R., *A molecular view of microbial diversity and the biosphere*. Science, 1997. **276**(5313): p. 734-40.
128. Delmont, T.O., et al., *Accessing the soil metagenome for studies of microbial diversity*. Appl Environ Microbiol, 2011. **77**(4): p. 1315-24.
129. Amann, R., W. Ludwig, and K. Schleifer, *Phylogenetic identification and in situ detection of individual microbial cells without cultivation*. Microbiological reviews, 1995. **59**(1): p. 143-69.
130. Petrosino, J.F., et al., *Metagenomic pyrosequencing and microbial identification*. Clin Chem, 2009. **55**(5): p. 856-66.
131. Simon, C. and R. Daniel, *Metagenomic analyses: past and future trends*. Appl Environ Microbiol, 2011. **77**(4): p. 1153-61.
132. Tringe, S.G. and P. Hugenholtz, *A renaissance for the pioneering 16S rRNA gene*. Curr Opin Microbiol, 2008. **11**(5): p. 442-6.
133. Hazen, T.C., A.M. Rocha, and S.M. Techtmann, *Advances in monitoring environmental microbes*. Curr Opin Biotechnol, 2013. **24**(3): p. 526-33.
134. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data*. Nat Methods, 2010. **7**(5): p. 335-6.
135. Caporaso, J.G., et al., *Moving pictures of the human microbiome*. Genome Biol, 2011. **12**(5): p. R50.
136. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms*. ISME J, 2012. **6**(8): p. 1621-4.
137. Gloor, G.B., et al., *Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products*. PLoS One, 2010. **5**(10): p. e15406.
138. Lazarevic, V., et al., *Metagenomic study of the oral microbiota by Illumina high-throughput sequencing*. J Microbiol Methods, 2009. **79**(3): p. 266-71.
139. von Wintzingerode, F., U. Göbel, and E. Stackebrandt, *Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis*. FEMS microbiology reviews, 1997. **21**(3): p. 213-29.
140. Acinas, S.G., et al., *PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample*. Applied and Environmental Microbiology, 2005. **71**(12): p. 8966-8969.
141. Polz, M. and C. Cavanaugh, *Bias in template-to-product ratios in multitemplate PCR*. Applied and environmental microbiology, 1998. **64**(10): p. 3724-30.
142. Konstantinidis, K.T., A. Ramette, and J.M. Tiedje, *The bacterial species definition in the genomic era*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2006. **361**(1475): p. 1929-1940.
143. Handelsman, J., et al., *Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products*. Chemistry and biology, 1998. **5**(10): p. R245-9.
144. Lasken, R.S. and J.S. McLean, *Recent advances in genomic DNA sequencing of microbial species from single cells*. Nat Rev Genet, 2014. **15**(9): p. 577-84.
145. Riesenfeld, C.S., P.D. Schloss, and J. Handelsman, *METAGENOMICS: Genomic Analysis of Microbial Communities*. Annual Review of Genetics, 2004. **38**(1): p. 525-552.
146. Handelsman, J., *Metagenomics: application of genomics to uncultured microorganisms*. Microbiology and molecular biology reviews : MMBR, 2004. **68**(4): p. 669-85.

147. Tyson, G., et al., *Community structure and metabolism through reconstruction of microbial genomes from the environment*. Nature, 2004. **428**(6978): p. 37-43.
148. Delmont, T.O., et al., *Accessing the Soil Metagenome for Studies of Microbial Diversity*. Applied and Environmental Microbiology, 2010. **77**(4): p. 1315-1324.
149. Quince, C., T.P. Curtis, and W.T. Sloan, *The rational exploration of microbial diversity*. The ISME Journal, 2008. **2**(10): p. 997-1006.
150. Vogel, T.M., et al., *TerraGenome: a consortium for the sequencing of a soil metagenome*. Nature Reviews Microbiology, 2009. **7**(252): p. doi:10.1038/nrmicro2119.
151. Delmont, T.O., et al., *Structure, fluctuation and magnitude of a natural grassland soil metagenome*. The ISME Journal, 2012. **6**(9): p. 1677-1687.
152. Roh, S.W., et al., *Comparing microarrays and next-generation sequencing technologies for microbial ecology research*. Trends in biotechnology, 2010. **28**(6): p. 291-299.
153. Bentley, D.R., *Whole-genome re-sequencing*. Current Opinion in Genetics & Development, 2006. **16**(6): p. 545-552.
154. Behrensmeyer, A.K. and J.H. Miller, *Building links between Ecology and Paleontology using taphonomic studies of recent vertebrate communities*. Paleontology in Ecology and Conservation, ed. J. Louys. 2012: Springer Verlag Berlin.
155. Lauber, C.L., et al., *Vertebrate decomposition is accelerated by soil microbes*. Appl Environ Microbiol, 2014. **80**(16): p. 4920-9.
156. Payne, J.A., *A Summer Carrion Study of the Baby Pig *Sus Scrofa* Linnaeus*. Ecology, 1965. **46**(5): p. 592-602.
157. Carter, D. and M. Tibbett, *Microbial decomposition of skeletal muscle tissue (*Ovis aries*) in a sandy loam soil at different temperatures*. Soil Biology and Biochemistry, 2006. **38**(5): p. 1139-1145.
158. Forbes, S.L., *Decomposition Chemistry in a Burial Environment*. Soil analysis in forensic taphonomy - Chemical and Biological Effects of Buried Human Remains, ed. M. Tibbett and D.O. Carter. 2008, Boca Raton: CRC Press. 366.
159. Dadour, I.R. and M.L. Harvey, *The Role of Invertebrates in Terrestrial Decomposition: Forensic Applications*. Soil analysis in forensic taphonomy - Chemical and Biological Effects of Buried Human Remains, ed. M. Tibbett and D.O. Carter. 2008, Boca Raton: CRC Press. 366.
160. Tibbett, M. and D.O. Carter, *Cadaver Decomposition and Soil: Processes*. Soil analysis in forensic taphonomy - Chemical and Biological Effects of Buried Human Remains, ed. M. Tibbett and D.O. Carter. 2008, Boca Raton: CRC Press. 366.
161. Hopkins, D.W., *The Role of Soil Organisms in Terrestrial Decomposition*. Soil analysis in forensic taphonomy - Chemical and Biological Effects of Buried Human Remains, ed. M. Tibbett and D.O. Carter. 2008, Boca Raton: CRC Press. 366.
162. Dix, J. and M. Graham, *Time of death, decomposition and identification: an atlas*. Forensic Pathology Atlases. Causes of Death Series, ed. J. Dix. 2000, Boca Raton: CRC Press LLC. 117.
163. Carter, D.O., D. Yellowlees, and M. Tibbett, *Cadaver decomposition in terrestrial ecosystems*. Naturwissenschaften, 2007. **94**(1): p. 12-24.
164. Forbes, S.L., B.B. Dent, and B.H. Stuart, *The effect of soil type on adipocere formation*. Forensic Sci Int, 2005. **154**(1): p. 35-43.

165. Carter, D.O., D. Yellowlees, and M. Tibbett, *Moisture can be the dominant environmental parameter governing cadaver decomposition in soil*. *Forensic Sci Int*, 2010. **200**(1-3): p. 60-6.
166. Wang, Y. and P.Y. Qian, *Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies*. *PLoS One*, 2009. **4**(10): p. e7401.
167. Putelat, O. *L'homme, l'animal et l'Ajoie. Du premier Moyen Âge à l'aube du second millénaire. Ostéologie des sites de Courtedoux, Creugenat ; de Bure, Montbion et de Bourogne, Cimetière communal*. in *Actes des journées archéologiques frontalières de l'Arc jurassien*. 2005. Presses Universitaires de Franche-Comté et Porrentruy.
168. Barrand-Emam, H., et al., *Artzenheim "Lotissement les Violettes" Alsace, Haut-Rhin (68). Un ensemble funéraire du Premier Moyen-Âge (fin 6ème - fin 9ème) et une occupation du début du Bronze final*, in *Rapport Final d'Opération d'Archéologie préventive*. 2013.
169. Guthmann, E., *L'exploitation des ressources animales dans le néolithique moyen danubien (5000-4500 av. j.-c.). L'exemple de la plaine du Rhin supérieur*, in *École doctorale Sciences humaines et sociales – Perspectives européennes (Strasbourg)*. in progress, Strasbourg: Strasbourg.
170. DENAIRE, A., et al., *Kolbsheim « Vogeseblick », du village Néolithique ancien à la position de la Bruche de 1914*, in *Rapport final d'Opération d'archéologie préventive*, A.-A. SRA Alsace, Editor. 2013.
171. Bell, L.S., J.A. Thorp, and A. Elkerton, *The sinking of the Mary Rose warship: a medieval mystery solved?* *Journal of Archaeological Science*, 2009. **36**: p. 166-173.
172. Gorgé, O., et al., *Analysis of ancient DNA in Microbial Ecology*, in *Microbial Environmental Genomics*, F. Martin and S. Uroz, Editors. 2015, Springer New York. p. 289-315.
173. Gansauge, M.T. and M. Meyer, *Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA*. *Nat Protoc*, 2013. **8**(4): p. 737-48.
174. Deagle, B.E., J.P. Eveson, and S.N. Jarman, *Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces*. *Frontiers in Zoology*, 2006. **3**(11).
175. Rychlik, W., *OLIGO 7 primer analysis software*. *Methods Mol Biol*, 2007. **402**: p. 35-60.
176. Tourasse, N.J., et al., *The Bacillus cereus group: novel aspects of population structure and genome dynamics*. *J Appl Microbiol*, 2006. **101**(3): p. 579-93.
177. Kolsto, A.B., N.J. Tourasse, and O.A. Okstad, *What sets Bacillus anthracis apart from other Bacillus species?* *Annu Rev Microbiol*, 2009. **63**: p. 451-76.
178. Moayeri, M., et al., *Anthrax Pathogenesis*. *Annu Rev Microbiol*, 2015. **69**: p. 185-208.
179. Donoghue, H., *Insights gained from palaeomicrobiology into ancient and modern tuberculosis*. *Clin Microbiol Infect*, 2011. **17**: p. 821-829.
180. Kamerbeek, J., et al., *Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology*. *J Clin Microbiol*, 1997. **35**(4): p. 907-14.
181. Vergnaud, G., et al., *Analysis of the three Yersinia pestis CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA*. *Adv Exp Med Biol*, 2007. **603**: p. 327-38.

182. Eisenach, K.D., et al., *Polymerase chain reaction amplification of a repetitive DNA sequence specific for Mycobacterium tuberculosis*. J Infect Dis, 1990. **161**(5): p. 977-81.
183. *Technical Data Set : Cereus-Ident-Agar*, M. Millipore, Editor. 1.46092.0020.
184. Baker, G.C., J.J. Smith, and D.A. Cowan, *Review and re-analysis of domain-specific 16S primers*. Journal of Microbiological Methods, 2003. **55**(3): p. 541-555.
185. Corless, C.E., et al., *Contamination and sensitivity issues with a real-time universal 16S rRNA PCR*. J Clin Microbiol, 2000. **38**(5): p. 1747-52.
186. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies*. Nucleic Acids Res, 2013. **41**(1): p. e1.
187. Mohammadi, T., et al., *Optimization of Real-Time PCR Assay for Rapid and Sensitive Detection of Eubacterial 16S Ribosomal DNA in Platelet Concentrates*. Journal of Clinical Microbiology, 2003. **41**(10): p. 4796-4798.
188. Hofstadler, S.A., et al., *TIGER: the universal biosensor*. International Journal of Mass Spectrometry, 2005. **242**(1): p. 23-41.
189. Spangler, R., N.L. Goddard, and D.S. Thaler, *Optimizing Taq polymerase concentration for improved signal-to-noise in the broad range detection of low abundance bacteria*. PLoS One, 2009. **4**(9): p. e7010.
190. Watanabe, I., Y. Kodama, and S. Harayama, *Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting*. Journal of Microbiological Methods, 2001. **44**: p. 253-262.
191. Winsley, T., et al., *Capturing greater 16S rRNA gene sequence diversity within the domain Bacteria*. Appl Environ Microbiol, 2012. **78**(16): p. 5938-41.
192. Liu, Z., et al., *Short pyrosequencing reads suffice for accurate microbial community analysis*. Nucleic Acids Res, 2007. **35**(18): p. e120.
193. Lu, J., et al., *Use of PCR with Universal Primers and Restriction Endonuclease Digestions for Detection and Identification of Common Bacterial Pathogens in Cerebrospinal Fluid*. J Clin Microbiol, 2000. **38**(6): p. 2076-2080.
194. Huws, S.A., et al., *Specificity and sensitivity of eubacterial primers utilized for molecular profiling of bacteria within complex microbial ecosystems*. J Microbiol Methods, 2007. **70**(3): p. 565-9.
195. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample*. Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4516-22.
196. Masella, A.P., et al., *PANDAseq: paired-end assembler for illumina sequences*. BMC Bioinformatics, 2012. **13**: p. 31.
197. Martin, M., *Cutadapt removes adapter sequences from igh-throughput sequencing reads*. EMBnet.Journal, 2011. **17**(1): p. 10-12.
198. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.
199. Segata, N., et al., *metagenomic microbial community profiling using unique clade-specific marker genes*. Nature Methods, 2012. **9**(8): p. 811-814.
200. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-2079.
201. Ondov, B.D., N.H. Bergman, and A.M. Phillippy, *Interactive metagenomic visualization in a Web browser*. BMC Bioinformatics, 2011. **12**: p. 385.
202. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads*. Nat Methods, 2013. **10**(10): p. 996-8.

203. Edgar, R.C., et al., *UCHIME improves sensitivity and speed of chimera detection*. *Bioinformatics*, 2011. **27**(16): p. 2194-200.
204. Blankenberg, D., et al., *Manipulation of FASTQ data with Galaxy*. *Bioinformatics*, 2010. **26**(14): p. 1783-5.
205. Hedges, R.E. and A.R. Millar, *Measurements and relationships of diagenetic alteration of bone from three archaeological sites*. *Journal of Archaeological Science*, 1995. **22**: p. 201-209.
206. Maia, V.H., et al., *DNA barcoding Bromeliaceae: achievements and pitfalls*. *PLoS One*, 2012. **7**(1): p. e29877.
207. Caruthers, M.H., et al., *Deoxyoligonucleotide synthesis via the phosphoramidite method*. *Gene Amplif Anal*, 1983. **3**: p. 1-26.
208. DeSantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB*. *Appl Environ Microbiol*, 2006. **72**(7): p. 5069-72.
209. Awad, A.H., et al., *Seasonal evaluation of fine and coarse culturable bacterial aerosols from residences within a rural and an urban city in Egypt*. *Int J Environ Health Res*, 2013. **23**(4): p. 269-80.
210. Lee, J.L. and R.E. Levin, *A comparative study of the ability of EMA and PMA to distinguish viable from heat killed mixed bacterial flora from fish fillets*. *J Microbiol Methods*, 2009. **76**(1): p. 93-6.
211. Vesper, S., et al., *Quantifying fungal viability in air and water samples using quantitative PCR after treatment with propidium monoazide (PMA)*. *J Microbiol Methods*, 2008. **72**(2): p. 180-4.
212. Wang, S. and R.E. Levin, *Discrimination of viable Vibrio vulnificus cells from dead cells in real-time PCR*. *J Microbiol Methods*, 2006. **64**(1): p. 1-8.
213. Patel, P., et al., *Development of an ethidium monoazide-enhanced internally controlled universal 16S rDNA real-time polymerase chain reaction assay for detection of bacterial contamination in platelet concentrates*. *Transfusion*, 2012. **52**(7): p. 1423-32.
214. Pinhasi, R., et al., *Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone*. *PLoS One*, 2015. **10**(6): p. e0129102.
215. Mittnik, A., et al., *A Molecular Approach to the Sexing of the Triple Burial at the Upper Paleolithic Site of Dolni Vestonice*. *PLoS One*, 2016. **11**(10): p. e0163019.
216. Ozga, A.T., et al., *Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus*. *Am J Phys Anthropol*, 2016. **160**(2): p. 220-8.
217. Liu, B., et al., *Deep sequencing of the oral microbiome reveals signatures of periodontal disease*. *PLoS One*, 2012. **7**(6): p. e37919.
218. Camelo-Castillo, A.J., et al., *Subgingival microbiota in health compared to periodontitis and the influence of smoking*. *Front Microbiol*, 2015. **6**: p. 119.
219. Seshadri, R., et al., *Comparison of the genome of the oral pathogen Treponema denticola with other spirochete genomes*. *Proc Natl Acad Sci U S A*, 2004. **101**(15): p. 5646-51.
220. Blouin, Y., *A new scenario for the early evolution of Mycobacterium tuberculosis*. 2014, Université Paris Sud - Paris XI.
221. Maiden, M.C., *Multilocus sequence typing of bacteria*. *Annu Rev Microbiol*, 2006. **60**: p. 561-88.
222. Mo, S., et al., *Multilocus sequence analysis of Treponema denticola strains of diverse origin*. *BMC Microbiology*, 2013. **13**(24): p. 12.

223. Munoz, R., et al., *Release LTPs104 of the All-Species Living Tree*. Syst Appl Microbiol, 2011. **34**(3): p. 169-70.
224. Chen, T., et al., *The bioinformatics resource for oral pathogens*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W734-40.
225. Metcalf, J.L., et al., *A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system*. Elife, 2013. **2**: p. e01104.
226. Metcalf, J.L., et al., *Microbial community assembly and metabolic function during mammalian corpse decomposition*. Science, 2016. **351**(6269): p. 158-62.



MINISTÈRE DE LA DÉFENSE

ANNEXE 1

Protocole n° 12130101

Modèle d'infection avec *Bacillus atrophaeus* et *Xanthomonas arboricola*

Évaluation de la persistance d'ADN bactériens dans l'environnement après septicémie simulée chez la souris

Version n	Date	Nature des modifications	
A	27/09/2012	Création	
		Prénom NOM	Fonction (Entité) Date et signature
Rédaction		Olivier GORGÉ	DRB/DBIO
Chef de département		Guilhem LARIGAUDERIE	DRB/DBIO
Vérification		Mickaël VALIDIRE	Correspondant qualité
Approbation			Président de séance du Comité Consultatif pour l'Utilisation des modèles Animaux

Le texte et les indications de ce document sont la propriété intellectuelle du DGA MNRBC, Toute reproduction, même partielle, et tout droit d'exploitation industrielle ou commerciale sont subordonnés à une autorisation écrite et préalable du DGA MNRBC,

DGA MNRBC FM 202009 version 3 Document à diffusion électronique contrôlée.

L'utilisateur doit valider tout exemplaire papier avant emploi,

1. OBJET DU PROTOCOLE

La lutte contre la prolifération biologique nécessite de disposer de moyens de mise en évidence d'activités passées ou masquées. Le laboratoire référent, en charge de la recherche de preuves de ces activités illicites, peut donc être confronté à des échantillons anciens, ou dégradés, dans lequel l'agent biologique étudié par le proliférant peut n'être plus présent que par son acide nucléique (ADN ou ARN). La décroissance de l'ADN dans l'environnement est mal connue, alors qu'il serait l'élément principal dans la mise en évidence et la charge de la preuve pour la démonstration d'activités interdites passées. En effet, la détection de l'ADN est la technique la plus sensible à l'heure actuelle, et il est possible de discriminer une souche pathogène autochtone d'une souche importée par génotypage, toujours à partir de cet ADN.

L'objet de ce protocole est de simuler une septicémie à *B. anthracis* et *Y. pestis* chez la souris en utilisant les simulants *B. atrophaeus* et *Xanthomonas arboricola* respectivement. Ce protocole permettra d'évaluer la possibilité de détecter dans l'environnement les ADN de ces bactéries après sacrifice et enfouissement des dépouilles. Deux bactéries ont été choisies pour représenter les bactéries Gram+ et les bactéries Gram-, avec la particularité de sporuler pour la bactérie Gram+ retenue. La détection sera réalisée par PCR quantitative à intervalles de temps définis, afin de déterminer la décroissance de la quantité de ces ADN, et estimer ainsi la pertinence à mener des tentatives de détection d'ADN pathogènes dans l'environnement a posteriori, lorsque des activités passées ont été décrites et que le site devient accessible (îles de Gruinart ou Vozrozhdenye par exemple). Le projet s'inscrit dans la problématique de la filière de la preuve, spécifiquement dans le domaine biologique.

1.1. Description de l'étude :

Un groupe de souris sera artificiellement contaminé par injection intra-veineuse d'agents bactériens simulants. Après sacrifice, ces souris seront enterrées dans de petites cages dans l'emprise de DGA MNRBC, et des prélèvements seront effectués à intervalles de temps définis. L'ADN bactérien sera recherché dans les prélèvements, permettant de caractériser l'évolution de la charge bactérienne, sa dispersion dans l'environnement et sa rémanence dans les restes (principalement os et dents) des souris.

L'étude est menée sur 2 agents bactériens non pathogènes : *Bacillus atrophaeus* et *Xanthomonas arboricola*.

Le premier est l'agent de classe I couramment utilisé dans les campagnes d'essais. Le second vient d'être autorisé par le CGA pour les essais d'épandage à Cazeaux, suite à plusieurs études montrant sa complète innocuité pour l'environnement.

Une pré-étude sera menée sur quatre souris, afin de valider le modèle de simulation de septicémie. Ces quatre souris subiront pour moitié une injection d'une suspension de *B. atrophaeus* et pour moitié de *X. arboricola*. Elles seront sacrifiées après 24 heures et l'ADN des bactéries sera recherché dans les os, les dents et les muscles de ces souris. Si la recherche est positive, l'étude sera étendue à l'ensemble de souris. Dans le cas contraire, un second essai sera réalisé avec des souris immuno-déprimées, dans les mêmes conditions et avec les mêmes attendus. Ensuite, la moitié des souris (neutres ou immunodéprimées) subira l'injection de 0,5 mL³ d'une suspension de formes végétatives de *B. atrophaeus* et la seconde de formes végétatives de *X. arboricola*. Ces suspensions auront un titre de 10¹⁰ cfu/mL, soit un titre sanguin de 3,10⁹ bactéries/mL (volume sanguin approximatif de la souris 1,6 à 1,8 mL). Pour l'anthrax, la concentration sanguine en formes végétatives, au moment du décès, a été mesurée à 10⁷ cfu/mL, Pour *Y. pestis*, la bactériémie peut monter à 10⁸ cfu/mL. Considérant le caractère non pathogène des bactéries simulantes utilisées, il est estimé qu'une dose 10 à 100 fois plus élevée que la concentration maximale observée lors d'une septicémie réelle est à même soit de provoquer une septicémie chez l'animal, soit d'être encore présent lors de son sacrifice.

Le calendrier de prélèvements sera le suivant : T0, T+1 mois, T+3 mois, T+6 mois, T+12 mois, T+18 mois, T+24 mois et T+30 mois, soit 8 lots de souris. Chaque lot comprend deux souris contaminées par agent bactérien non pathogène et une souris témoin, soit 5 souris par lot. Afin de prévenir la disparition d'un lot (animal fouisseur, travaux, etc, ...), 4 lots supplémentaires seront prévus, soit un total de 12 lots de 5 souris. Au total, 68 souris seront nécessaires, deux lots de 4 souris pour les pré-études, 2 lots de 20 souris et un lot de 10 pour l'étude proprement dite (lots ensuite répartis selon les contingences du prélèvement en 12 lots de 5 souris).

3 Volume maximum prévu et compatible avec une souris,

1.2. Justification du modèle utilisé par rapport à des méthodes alternatives physiques :

Dans cette étude, il est modélisé un processus biologique de septicémie, suivi de l'enfouissement dans le sol des restes animaux et leur analyse après un temps déterminé de dégradation. La multiplicité des paramètres biologiques entrant en jeu (distribution dans l'organisme, interaction bactéries et différents tissus, multiplication, ...) nécessite d'utiliser un modèle animal.

L'utilisation d'animaux pour cette étude s'avère donc indispensable.

1.3. Choix de l'hôte

Le choix s'est porté en première intention sur des souris neutres, afin de représenter au mieux ce qui peut se passer dans la nature avec la contamination naturelle d'animaux par un agent pathogène. Si l'organisme de la souris est en mesure d'éliminer l'agent avant le sacrifice de celle-ci, l'étude sera réalisée sur des souris immuno-déprimées.

2. REFERENCES

2.1. Textes régissant les conditions légales d'expérimentation sur animaux vivants

2.1.1. Ministère de l'agriculture, de l'alimentation, de la pêche et des affaires rurales

Décret n° 2003-768 du 1er août 2003 relatif à la partie Réglementaire du livre II du code rural, J.O n° 181 du 7 août 2003 page 13688, NOR: AGRD0300394D

Décret n° 2003-768 du 1er août 2003 relatif à la partie Réglementaire du livre II du code rural, J.O n° 181 du 7 août 2003 page 37235, ANNEXES, NOR: AGRD0300394D

2.1.2. Ministère de la défense

Arrêté du 29/10/90 relatif aux conditions de l'expérimentation sur les animaux vivants pratiquée dans les établissements relevant du ministre de la défense.

2.1.3. Communauté Européenne

Journal Officiel des Communautés Européennes L358, Conseil 86/609/CEE :
Directive du Conseil du 24 novembre 1986, concernant le rapprochement des dispositions législatives, réglementaires et administratives des États membres relatives à la protection des animaux utilisés à des fins expérimentales ou à d'autres fins scientifiques.

2.1.4. Ministère des affaires étrangères

Décret n° 2001-486 du 6 juin 2001 portant publication de la Convention européenne sur la protection des animaux vertébrés utilisés à des fins expérimentales ou à d'autres fins scientifiques, adoptée à Strasbourg le 18 mars 1986 et signée par la France le 2 septembre 1987, NOR : MAEJ0130033D, JO du 8 juin 2001, p. 9094

2.1.5. Agrément de l'établissement pratiquant l'expérimentation animale

DGA MNRBC : Arrêté ministériel d'agrément en date du 23 juillet 2012 (ministère de la défense)

Autre établissement éventuel : Néant

2.2. Procédures, modes opératoires

DGA MNRBC MO 307002 : élimination des déchets solides à incinérer au DGA MNRBC

CEB FM 202008 : fiche de réception des animaux

CEB MO 200170 animalerie souris : hébergement et entretien

CEB MO 202200 : commande et réception des animaux

CEB NO 202038 : Modification de l'organisation du comité consultatif pour l'utilisation des modèles animaux (CCUA)

3. TERMINOLOGIE ET SIGLES

Néant

4. MATERIELS ET METHODES

4.1. Expérimentateurs

Autorisations nominatives :

Niveau et formation

Niveau I:

Philippe MERY, ICT, vétérinaire diplômé de l'école vétérinaire de Toulouse (certificat n° 2012/22).

Rôle dans l'étude : Responsable de l'expérimentation, commandes, suivi sanitaire des animaux, participe à la contamination.

Jérémy RAITEUX, ICT, formation à l'expérimentation animale niveau I, en mars 2009 au centre de formation à l'expérimentation animale d'Alfort (ENV), (certificat n° 2012/23).

Rôle dans l'étude : participe à la contamination, assure la dissection des animaux du lot T0, gestion des résultats.

Françoise RAYNAUD

Rôle dans l'étude : Françoise RAYNAUD réalise l'injection intra-veineuse des suspensions bactériennes dans les souris (certificat n° 2012/24).

Niveau III:

Personnels affectés à l'entretien des animaux, des locaux et du matériel

David LORY

Olivier GORGE,

Rôle dans l'étude : demandeur de l'étude, conception, analyse et interprétation des résultats

4.2. Animaux

Espèce, race, souche : Souris BALB/c ou BALB/c nude mouse

Sexe : femelle

Statut sanitaire : IOPS

Poids à la commande : 18 - 20 g

Provenance : Charles River France

Début d'étude prévu le : Novembre 2012

Durée de l'étude : 30 mois

Durée de la période d'acclimatation : 1 semaine

Nombre d'animaux par lot : **20 souris par lot à contaminer, 10 souris dans le lot témoin et deux lots de 4 souris pour la pré-étude.**

Nombre de lots : 5

4.3. Environnement

4.3.1. Localisation de l'expérimentation (Bt ; pièce)

Local de réception : Bâtiment 311

Local d'hébergement acclimatation : Bâtiment 371, pièce 11

Local de traitement : Bâtiment 371

Local d'hébergement après administration : Bâtiment 371, pièce 11

4.3.2. Unités animales

Les conditions d'hébergement des animaux sont précisées dans le tableau suivant :

		Conditions d'hébergement pendant la période d'acclimatation	Les conditions d'hébergement pendant la période d'essai
Mode de traitement de l'air :	<i>Température</i>	20-22 °C	20-22 °C
	<i>Humidité Relative</i>	55 %	55 %
	<i>Pression relative</i>	surpression	surpression
	<i>Filtration air</i>	oui	oui
	<i>Système de sécurité</i>	oui	oui
Éclairage :	<i>Naturel</i>	non	non
	<i>Artificiel</i>	oui	oui
	<i>Cycle de lumière</i>	oui - 12h/12h	oui - 12h/12h

4.3.3. Cages et nombre d'animaux par cage

Cages : Innocage souris avec couvercle statique cage type II, composition PET, dimension (hors tout) 37x23x14, surface 522 cm², hauteur utile 13 cm, fournisseur Innovive. Elles sont jetables et fabriquées à partir de plastique recyclé.

Nombre d'animaux par cage : 5 animaux par cage.

Identification des individus ou des lots : les lots sont identifiés individuellement

Nature litière : sciure de bois stérile

Recueil des urines et des fèces : non

4.3.4. Entretien

Les cages et litières sont changées 2 fois par semaine.

4.3.5. Mode d'alimentation

Nature de l'aliment : aliment SDS rongeurs

Ration : ad libitum

Mode de distribution : mangeoire

4.3.6. Mode d'abreuvement

Qualité : eau potable

Mode de distribution : biberon

Contrôle des quantités absorbées : non

4.4. Protocole expérimental précis

4.4.1. Description précise de l'ensemble du processus que va subir l'animal

4.4.1.1. Préparation de l'animal

Néant

4.4.1.2. L'expérimentation proprement dite

Dissection des animaux de la pré-étude pour prélever les fémurs, les dents et la queue après euthanasie avec une dose létale de pentobarbital afin de rechercher l'ADN des bactéries injectées.

Sacrifice de tous les animaux à J+1 par une dose létale de pentobarbital.

Dissection des animaux du lot T0 pour prélever les os, les dents et la queue.

Enfouissement des autres lots.

Quantification des agents biologiques dans les os et les dents pour chaque lot (et queue pour lot T0 et lots réservés à la pré-étude).

4.5. En fonction de ce qui précède évaluation de la douleur et de la contrainte provoquée

Durée de la douleur : néant

Durée du processus de la pathologie induite : défini lors de la pré-étude.

Point limite : 24H post-injection.

Moyens mis en jeu pour lutter contre la douleur : aucun

4.6. Classification normalisée de Russel et Burch

Classement proposé : C

4.7. Devenir des animaux

Euthanasiés :

Méthode : injection létale de pentobarbital, dose injectée 100 mg/kg

Pratiquée par : expérimentateurs déclarés

Conservés au Laboratoire : non

Cédés : non

Les animaux seront éliminés dans les conteneurs adaptés, Les conteneurs prétraités par autoclavage par le laboratoire sont étiquetés, transportés et traités selon la procédure DGA MNRBC PR 307002.

ANNEXE 2 - PROTOCOLE DE PRELEVEMENT

Ecouvillon sec :

- J0 : Tremper l'écouvillon dans un tube de BHI, puis faire un isolement sur gélose sélective. A l'issue, plonger l'écouvillon dans le BHI et mettre à incuber 24h à 30°C
- J1 :
 - inspecter la gélose sélective et repiquer toute colonie blanche (*B. anthracis* et *B. cereus / thuringiensis* mutés) sur gélose TSA;
 - Prélever 100µL de BHI et étaler sur gélose sélective
 - Plonger une oese dans le BHI et faire un isolement sur gélose TSA
- J2 :
 - inspecter les géloses sélectives J1 et repiquer toute colonie blanche
 - analyse génétique sur les colonies issues du repiquage de la gélose sélective J0
- J3 :
 - inspecter les géloses J2
 - analyse génétique sur les colonies issues du repiquage de la gélose J1

Ecouvillon humide :

- J0 : faire un isolement sur gélose sélective. Plonger l'écouvillon et le reste de liquide dans du BHI et mettre à incuber 24h à 30°C
- J1 :
 - inspecter la gélose sélective et repiquer toute colonie blanche
 - Prélever 100µL de BHI et étaler sur gélose sélective
 - Plonger une oese dans le BHI et faire un isolement sur gélose TSA
- J2 :
 - inspecter les géloses sélectives J1 et repiquer toute colonie blanche sur TSA
 - analyse génétique sur les colonies issues du repiquage de la gélose sélective J0
- J3 :
 - inspecter les géloses J2

- analyse génétique sur les colonies blanches issues du repiquage de la gélose J1

Sédiment :

- J0 : Mettre en suspension 0,5g de sédiments dans de l'eau distillée. Prélever 100µL, les ajouter à du BHI et mettre à incuber 24h à 30°C. Prélever 100µL, faire un étalement sur gélose sélective et mettre à incuber 24h à 30°C
- J1 :
 - inspecter la gélose sélective et repiquer toute colonie blanche ;
 - Prélever 100µL de BHI et étaler sur gélose sélective
 - Plonger une oese dans le BHI et faire un isolement sur gélose sélective
- J2 :
 - inspecter les repiquages J1 et repiquer toute colonie blanche
 - analyse génétique sur les colonies blanches issues de la gélose J0
- J3 :
 - inspecter les géloses J2
 - analyse génétique sur les colonies blanches issues du repiquage de la gélose J1

ANNEXE 3 - SCRIPTS

Usearch2 taxo script

```
#!/bin/bash
mkdir merged_reads
mkdir fastq
mkdir no_singleton
mkdir centroids
mkdir otus
mkdir no-chimera
mkdir labelled
mkdir taxo
mkdir krona
mkdir csv-tsv_files

cp /Users/olivier/Desktop/scripts/Taxon_parsing_1.1.py .
cp /Users/olivier/Desktop/scripts/Label_change_1.1.py .

#pandaseq
for f1 in *R1_001.fastq.gz
do f2=${f1%%R1_001.fastq.gz}"R2_001.fastq.gz" && /Applications/pandaseq -f $f1 -r $f2 -B -l
40 >${f1%%R1_001.fastq.gz}merged.fa -p TACGGRAGGCAGCAG -q CGTATTACCGCGGCTGCTGG -
g ${f1%%R1_001.fastq.gz}log.txt
done

#move initial files to dedicated folder
mv *.gz merged_reads/

#clustering
for file in *.fa
do /Applications/usearch8 -threads 4 -cluster_fast $file -id 0.99 -centroids $file.centroids.fa -
sizeout
done
for file in *.centroids.fa
do /Applications/usearch8 -threads 4 -cluster_otus $file -id 0.99 -otus $file.otus -uparseout
```

```

$file.up -relabel OTU_ -sizein -sizeout
done
gzip *.centroids.fa
mv *.centroids.fa.gz centroids/
gzip *.fastq
mv *.fastq.gz fastq/
mv *.otus otus/
mv *.up otus/
cd otus/

#singleton removal
for file in *.otus
do /Applications/usearch8 -threads 4 -sortby size $file -fastaout $file.nosingle -minsize 2
done

gzip *.otus
gzip *.up
mv *.nosingle ../no-singleton
cd ../no-singleton

#chimera filtering
for file in *.nosingle
do /Applications/usearch8 -thread 4 -uchime_ref $file -db
/Users/olivier/Desktop/Usearch_db/gold.fa -strand plus -nonchimeras $file.nochimera
done
gzip *.nosingle
mv *.nochimera ../no-chimera
cd ..

#label centroids (sur .otus précédemment créés)
python Label_change_1.1.py no-chimera/ "nochimera"
cd no-chimera/
gzip *.nochimera
mv *.fa ../labelled
cd ../labelled

```

```

#taxo assign
for file in *.fa
do /Applications/usearch8-old -utax $file -db /Users/olivier/Desktop/Usearch_db/rdp_16s.fa -
taxconfs /Users/olivier/Desktop/Usearch_db/rdp_16s_short.tc -tt
/Users/olivier/Desktop/Usearch_db/rdp_16s.tt -utaxout $file.taxo
done
gzip *.fa
mv *.taxo ../taxo

#parsing results
cd ..
python Taxon_parsing_1.1.py taxo/ all_samples.csv
awk -F";" 'NR>1 {print $1. $4. $5. $6. $7. $8 .$9 >> ($3 ".csv"); close($3 ".csv")}' all_samples.csv
mv all_samples.csv csv-tsv_files/
for file in *.csv
do tr " " "\t" <$file >$file.tsv
done
#create krona file
for file in *.tsv
do ktImportText -o $file.html $file
done
mv *.html krona/
mv *.csv csv-tsv_files/

```

metaphlan2taxo

```
#!/bin/bash
# This metaphlan_script allows batch analyses of multiple fastq files in the same folder.
# To run this script, you need to have metaphlan2.1, graphlan and KronaTools installed, as well
as their various dependencies as described on the corresponding download sites. Here, they
have been installed in a directory called Software in Home.
# To allow direct access to the software from the working directory, install the software in the
Path by writing the following 4 lines in the ~/.bashrc file (without # ):
# export PATH=~/.Software/metaphlan2:$PATH
# export PATH=~/.Software/graphlan/:$PATH
# export PATH=~/.Software/KronaTools-2.5/:$PATH
# mpa_dir=~/.Software/metaphlan2

#Run the script as follows saving a logfile to test it when you modify something: sh
metaphlan2_script.sh >>report_script_metaphlan.txt 2>&1

#Even though this command is in the Path, it needs to be repeated specifically within the script
mpa_dir=~/.metaphlan2

#Run two parallel metaphlan analyses using two bowtie2 parameters: very sensitive (VS files)
and very sensitive local (VSL files) removing the reads that are mapped on less than 28 bases in
this later case.
for file in *.fastq
do
metaphlan2.py $file --mpa_pkl ${mpa_dir}/db_v20/mpa_v20_m200.pkl --bowtie2db
${mpa_dir}/db_v20/mpa_v20_m200 --bt2_ps very-sensitive-local --bowtie2out
${file}.VSL.bowtie2.bz2 --min_alignment_len 28 --nproc 4 --input_type multifastq >
${file}.VSL.meta.txt
done
for file in *.fastq
do
metaphlan2.py $file --mpa_pkl ${mpa_dir}/db_v20/mpa_v20_m200.pkl --bowtie2db
${mpa_dir}/db_v20/mpa_v20_m200 --bt2_ps very-sensitive --bowtie2out ${file}.VS.bowtie2.bz2
--nproc 4 --input_type multifastq > ${file}.VS.meta.txt
done
```

```

# Clean extension tags
for file in *.fastq.VSL.bowtie2.bz2
do
mv "$file" "${file%.fastq.VSL.bowtie2.bz2}.VSL.bowtie2.bz2"
done
for file in *.fastq.VS.bowtie2.bz2
do
mv "$file" "${file%.fastq.VS.bowtie2.bz2}.VS.bowtie2.bz2"
done
for file in *.fastq.VSL.meta.txt
do
mv "$file" "${file%.fastq.VSL.meta.txt}.VSL.meta.txt"
done
for file in *.fastq.VS.meta.txt
do
mv "$file" "${file%.fastq.VS.meta.txt}.VS.meta.txt"
done

#Generate a table of the hits
for file in *.VSL.bowtie2.bz2
do
metaphlan2.py $file --mpa_pkl ${mpa_dir}/db_v20/mpa_v20_m200.pkl -t clade_profiles --
bowtie2db ${mpa_dir}/db_v20/mpa_v20_m200 --nproc 4 --min_alignment_len 28 --input_type
bowtie2out > ${file}.VSL.metatable.txt
done

# Clean extension tags
for file in *.VSL.bowtie2.bz2.VSL.metatable.txt
do
mv "$file" "${file%.VSL.bowtie2.bz2.VSL.metatable.txt}.VSL.metatable.txt"
done

#Rerun a metaphlan analysis by filtering away the data generated with the very-sensitive-local
parameters to keep only bacteria and archae
for file in *.VSL.bowtie2.bz2

```

```

do
metaphlan2.py $file --mpa_pkl ${mpa_dir}/db_v20/mpa_v20_m200.pkl --bowtie2db
${mpa_dir}/db_v20/mpa_v20_m200 --ignore_viruses --ignore_eukaryotes --nproc 4 --
min_alignment_len 28 --input_type bowtie2out > ${file}.VSL_BA.meta.txt
done

# Clean extension tags
for file in *.VSL.bowtie2.bz2.VSL_BA.meta.txt
do
mv "$file" "${file%.VSL.bowtie2.bz2.VSL_BA.meta.txt}.VSL_BA.meta.txt"
done

#Format the data to allow a krona graph display
for file in *.meta.txt
do
python ${mpa_dir}/utils/metaphlan2krona.py -p $file -k ${file}.krona.txt
done

# Clean extension tags
for file in *.meta.txt.krona.txt
do
mv "$file" "${file%.meta.txt.krona.txt}.krona.txt"
done

#Generate a krona graph that will be legible on a computer with Krona graph installed but
without requiring an internet connection
for file in *.krona.txt
do
ktImportText $file -l -o ${file}.chart.html
done

# Clean extension tags
for file in *.krona.txt.chart.html
do
mv "$file" "${file%.krona.txt.chart.html}.chart.html"
done

```

```

# Merge Data to do a heat map analysis. here specific to the VSL and VSL_BA data
python      ${mpa_dir}/utils/merge_metaphlan_tables.py      *.VSL.meta.txt      >
merged_abundance_table_VSL.txt

python      ${mpa_dir}/utils/merge_metaphlan_tables.py      *.VSL_BA.meta.txt   >
merged_abundance_table_VSL_BA.txt

# Do the heatmap analysis and draw the heatmap grap. I propose various analyses that might be
useful to have a global view of your data. which can be amended.

python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.01 --tax_lev s
-s log --in merged_abundance_table_VSL.txt --out abundance_heatmap_VSL_Species_25.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev s
-s log --in merged_abundance_table_VSL.txt --out abundance_heatmap_VSL_Species_40.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev g
-s log --in merged_abundance_table_VSL.txt --out abundance_heatmap_VSL_Genera_40.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev f -
s log --in merged_abundance_table_VSL.txt --out abundance_heatmap_VSL_Families_40.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.01 --tax_lev s
-s      log      --in      merged_abundance_table_VSL_BA.txt      --out
abundance_heatmap_VSL_BA_Species_25.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev s
-s      log      --in      merged_abundance_table_VSL_BA.txt      --out
abundance_heatmap_VSL_BA_Species_40.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev g
-s      log      --in      merged_abundance_table_VSL_BA.txt      --out
abundance_heatmap_VSL_BA_Genera_40.png
python ${mpa_dir}/utils/metaphlan_hclust_heatmap.py -c bbcry --top 40 --minv 0.01 --tax_lev f -
s      log      --in      merged_abundance_table_VSL_BA.txt      --out
abundance_heatmap_VSL_BA_Families_40.png

#unzip and rename the product for the .VSL.bowtie2.bz2 files
bzip2 -dk *.VSL.bowtie2.bz2
for file in *.VSL.bowtie2
do
mv "$file" "${file%.VSL.bowtie2}.fastq.bowtie2"

```



```

done

#execute the Get infos script to recover the name of the sequences and the sequences that
mapped to the metaphlan database
for file in *.fastq
do
python ${mpa_dir}/utils/Get_infos_1.1.py $file ${file}.bowtie2 ${mpa_dir}/utils/m2c.txt
done

#rename the output to recover the .VSL tag
for file in *.fastq.meta_info
do
mv "$file" "${file%.fastq.meta_info}.VSL.meta_info"
done

#cut the 4th column with sequences
for file in *.VSL.meta_info
do
cut -f 4 $file > ${file}.sequences
done

#count the lines
wc -l *.sequences | sed 's/.meta_info.sequences//' >> readscount_VSL.txt

#Sort sequences and count unique reads
for file in *.VSL.meta_info.sequences
do
sort $file | uniq -c > ${file}.sorted
done

#count the unique lines
wc -l *.sorted | sed 's/.meta_info.sequences.sorted//' >> uniquereadscount_VSL.txt

#clean-up to do the second round with the .VS files
rm *.sequences
rm *.sequences.sorted

```

```

rm *bowtie2

#unzip and rename the product for the .VS.bowtie2.bz2 files
bzip2 -dk *.VS.bowtie2.bz2
for file in *.VS.bowtie2
do
mv "$file" "${file%.VS.bowtie2}.fastq.bowtie2"
done

#execute the Get infos script to recover the name of the sequences and the sequences that
mapped to the metaphlan database
for file in *.fastq
do
python ${mpa_dir}/utils/Get_infos_1.1.py $file ${file}.bowtie2 ${mpa_dir}/utils/m2c.txt
done

#rename the output to recover the .VS tag
for file in *.fastq.meta_info
do
mv "$file" "${file%.fastq.meta_info}.VS.meta_info"
done

#cut the 4th column with sequences
for file in *.VS.meta_info
do
cut -f 4 $file > ${file}.sequences
done

#count the lines
wc -l *.sequences | sed 's/\.meta_info.sequences//' >> readscout_VS.txt

#Sort sequences and count unique reads
for file in *.VS.meta_info.sequences
do
sort $file | uniq -c > ${file}.sorted
done

```

```
#count the unique lines
```

```
wc -l *sorted | sed 's/.meta_info.sequences.sorted//' >> uniquereadscount_VS.txt
```

```
#Merge all counts in one file. remove unnecessary spaces. recover only the initial file name in the first column and organize in the order VSI total. VSL unique. VS total. VS unique
```

```
paste readscount_VSL.txt uniquereadscount_VSL.txt readscount_VS.txt uniquereadscount_VS.txt
```

```
| sed 's/\+ //g' | gawk '{print $2. $1. $3. $5. $7}' | sed 's/.VSL\ //g' > readscountmerged.txt
```

```
#Make some order and put files into folders.
```

```
mkdir Input_fastq
```

```
mkdir Bowtie
```

```
mkdir Meta
```

```
mkdir Meta_tables
```

```
mkdir Krona
```

```
mkdir Heatmaps
```

```
mkdir Meta_info
```

```
mv *.fastq Input_fastq
```

```
mv *.bowtie2.bz2 Bowtie
```

```
mv *.meta.txt Meta
```

```
mv *.metatable.txt Meta_tables
```

```
mv *.krona.txt Krona
```

```
mv *.chart.html Krona
```

```
mv merged_abundance_table_VSL.txt Heatmaps
```

```
mv merged_abundance_table_VSL_BA.txt Heatmaps
```

```
mv *.png Heatmaps
```

```
mv *.meta_info Meta_info
```

```
rm *.sequences
```

```
rm *.sequences.sorted
```

```
rm *bowtie
```

shotgun2taxo (à partir des fichiers sam)

```
#!/bin/bash
```

```
# crée les répertoires
```

```
mkdir NC_counted
```

```
mkdir taxo_assigned
```

```

mkdir stats

# copie les scripts
cp /Volumes/SAMSUNG/scripts/compteurID_v2.pl .
cp /Volumes/SAMSUNG/scripts/matchTaxID_v4Internet.pl .
cp /Volumes/SAMSUNG/bacteria_db/bact_GenBank_summary .
cp -R /Volumes/SAMSUNG/bacteria_db/taxdump .

#extrait la longueur de chaque read
for file in *.sam;
do awk '{print length($10)}' $file > $file.sum;
done

#fait la moyenne de la taille des reads pour chaque banque
for file in *.sum ;
do awk '{total += $1} END {print total/NR}' $file > $file.mean ;
done

#compte le nombre de reads de chaque taille et les trie en fonction de cette taille
for file in *.sum; do sort $file | uniq -c | sort -nr -k2 > $file.sum2; done

#change les suffixes des fichiers
rename 's/sam.sum.sum2/txt/' *

#swape les champs hits et lenght
for file in *.txt; do awk '{ t = $1; $1 = $2; $2 = t; print; }' $file > $file.txt2; done

#insère le nom du fichier en première ligne
for file in *.txt2;
do echo $file > tempfile; cat $file >> tempfile; mv tempfile $file;
done

#insère les champs hits et taille en seconde ligne
for file in *.txt2; do awk 'NR==2 {print "lenght hits"} 1' $file > $file.txt3; done

#collationne l'ensemble des données de longueur de reads des banques en un fichier. une

```

banque par champ

```
paste *.txt3 > stats/libraries.sums
```

#corrige le nom des colonnes et insère une tab

```
sed 's/\.txt.txt2/\t/g' <stats/libraries.sums | sed -e 's/\s/\t/g' > stats/libraries.csv
```

#insère le nom du fichier en première ligne

```
for file in *.mean;
```

```
do echo $file >tempfile; cat $file >>tempfile; mv tempfile $file;
```

```
done
```

#collationne l'ensemble des données de longueur de reads des banques en un fichier. une banque par champ

```
paste *.mean > stats/libraries.means
```

```
rm *.mean
```

```
rm *.sum
```

```
rm *.txt
```

```
rm *.txt2
```

```
rm *.txt3
```

#recherche tous les NC_XXXXXX.x. pondère et collationne les occurrences et les exporte dans un nouveau fichier texte à deux champs. nombre d'occurrences et NC

```
for file in *.sam;
```

```
do perl compteurID_v2.pl $file $file.NC_counted;
```

```
done
```

#compte le nombre d'espèces bactériennes différentes trouvées

```
for file in *.NC_counted;
```

```
do wc -l $file >> stats/diversity.txt;
```

```
#done
```

#lance le script perl (qui doit être présent dans le répertoire avec taxdump et bact_genBank_summary)

```
for file in *.NC_counted ;
```

```
do perl matchTaxID_v4Internet.pl $file bact_GenBank_summary;
```

done

```
# mv *.final taxo_assigned/  
mv *.taxo.txt taxo_assigned/  
mv *.NC_counted NC_counted
```

Titre : Diagenèse de l'ADN bactérien et analyses métagénomiques de pathologies bactériennes du passé

Mots clés : ADN ancien, pathogène, bactérie, NGS, taphonomie

Cette étude a eu pour objet la mise en évidence de traces d'ADN bactérien pathogène dans des échantillons animaux et humains anciens, et ainsi améliorer les connaissances sur l'évolution des maladies au cours du temps. En parallèle, nous avons étudié les phénomènes de dégradation de l'ADN dans le sol sur des cadavres de souris enterrés après avoir été contaminées par des bactéries non pathogènes. Cette étude des processus taphonomiques s'est étalée sur trois ans et a permis de montrer une disparition rapide des bactéries simulantes, remplacé par l'ADN des bactéries du sol, qui colonisent rapidement la dépouille et dégradent tant l'ADN endogène (murin) qu'exogène (bactérien). Cette disparition rapide explique la grande difficulté à mettre

en évidence des pathogènes dans des échantillons anciens, à de rares exceptions près. Notre étude n'a pas permis de détecter d'agents pathogènes particuliers dans les échantillons que nous avons étudiés, mais nous avons mis en évidence l'intérêt d'analyser certains types de restes pour accéder à une information génétique préservée. Le tartre dentaire est un bon indicateur de la flore buccale de l'hôte et les kystes calcifiés assurent une bonne préservation de l'ADN endogène, moins soumis à contamination et digestion par les bactéries de l'environnement. Les kystes présentent en règle générale une teneur en ADN endogène supérieure à tous les autres tissus étudiés jusqu'à présent dans la littérature.

Title : Bacterial DNA diagenesis and metagenomic analyses of past bacterial pathologies

Keywords : ancient DNA, pathogen, bacteria, NGS, taphonomy

Abstract : The aim of this study was the identification of pathogenic bacterial DNA traces in ancient animal and human samples, and thus improve knowledge of past diseases that affect humankind over time. In parallel, we studied the DNA degradation phenomena in the soil on the buried corpses of mice after being contaminated by non-pathogenic bacteria. This study of taphonomic processes was spread over three years and has shown a rapid disappearance of simulant bacteria, replaced with the DNA of soil bacteria that colonize the body quickly after burial and degrade both the endogenous DNA (murine) that exogenous (bacteria). This quick degradation can explain the high

difficulty to detect and identify bacterial pathogens in old samples, with very few exceptions. Despite the fact in our study we were not able to detect specific pathogens in the samples we have studied, we have shown the interest to analyse certain types of remnants to access preserved and informative genetic data. Dental calculus is a good indicator of the oral flora of the host and calcified cysts ensure good preservation of the endogenous DNA, less subject to contamination and digestion by bacteria from the environment. Cysts generally have an endogenous DNA content higher than all other tissues examined up to now in the literature.

