

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE STRASBOURG

Discipline : Sciences du vivant

Spécialité : Aspects moléculaires et cellulaires de la biologie

présentée par

Caroline BOUAKAZE

Analyse de polymorphismes ponctuels d'ADN extraits de tissus humains anciens : application à l'étude de populations sibériennes anciennes

Soutenue publiquement le 15 février 2010

Membres du jury

Directeur de thèse

Monsieur Bertrand LUDES, Professeur à l'Université de Strasbourg

Rapporteurs externes

Monsieur Angel CARRACEDO, Professeur à l'Université de Saint-Jacques de Compostelle

Monsieur José BRAGA, Professeur à l'Université Paul Sabatier de Toulouse

Examineur

Monsieur Benoît JAULHAC, Professeur à l'Université de Strasbourg

Membre invité

Monsieur Jacky DE MONTIGNY, Professeur à l'Université de Strasbourg

REMERCIEMENTS

Je tiens tout d'abord à remercier mon Directeur de thèse, le Professeur Bertrand LUDES, de m'avoir accueillie au sein de son équipe de recherche depuis janvier 2006, pour un stage de quelques mois, puis pour m'avoir donné la possibilité de réaliser ce travail de thèse. Ce travail n'aurait pu être réalisé sans sa confiance, son soutien permanent malgré son emploi du temps chargé, ses conseils avisés et sa clairvoyance, ce dont je le remercie infiniment.

Mes plus sincères remerciements vont au Docteur Christine KEYSER qui a largement contribué à faire de ces années de thèse des années inoubliables. Je tiens à la remercier pour sa présence quotidienne à mes côtés durant ces quatre années de recherche. Sa patience, ses précieux conseils, ses encouragements et son soutien dans les moments les plus difficiles m'ont énormément aidé à persévérer et je lui en suis extrêmement reconnaissante.

Je tiens à exprimer toute ma reconnaissance aux membres du jury, les Professeurs Angel CARRACEDO, Benoît JAULHAC, José BRAGA et Jacky DE MONTIGNY, qui ont immédiatement accepté de consacrer une partie de leur temps précieux à l'évaluation de ce travail. C'est un grand honneur pour moi d'avoir pu soumettre ce travail à leur regard scientifique.

Je remercie chaleureusement les Professeurs Eric CRUBEZY et Henri DABERNAT de l'équipe d'Anthropologie Moléculaire et Imagerie de Synthèse de Toulouse pour m'avoir permis de réaliser ces études et pour l'intérêt qu'ils ont toujours porté à mon travail. Merci aussi pour la confiance qu'ils m'ont accordée et leur accueil lors des séjours effectués dans leur équipe.

Je tiens à remercier le Docteur Sylvie DE MARTINO de l'Institut de Bactériologie de Strasbourg et les Docteurs Nicolas VEZIRIS et Wladimir SOUGAKOFF du Centre National de Référence des Mycobactéries et de la Résistance des Mycobactéries aux Antituberculeux pour leur participation dans le projet sur la tuberculose.

Angéla, je te remercie pour ton aide précieuse dans la mise au point de l'analyse par spectrométrie de masse et dans la réalisation de certaines des figures de ce manuscrit.

Daniel, mille mercis pour votre aide enrichissante dans l'utilisation de plusieurs logiciels de bioinformatique et pour votre disponibilité quand j'avais des soucis techniques.

Prisca et Aurélie, je vous remercie pour les innombrables services que vous m'avez rendus. Prisca, cela a été un réel plaisir de travailler à tes côtés ; bravo pour ton incroyable sens pratique des choses et de l'organisation qui m'a énormément facilité le quotidien.

Marie et Fanny, je vous remercie pour les fous rires, les petites pauses café et surtout pour votre soutien durant ces derniers mois. Merci aussi pour toutes les discussions scientifiques

(et autres...) et pour vos conseils dans la rédaction de ce mémoire. Bon courage à toutes les deux pour la fin de votre thèse !

Annie, je vous remercie d'avoir toujours pris le temps de répondre à mes questions et d'avoir été disponible pendant mes moments difficiles. Merci beaucoup pour votre attention et votre bonne humeur.

Je remercie très sincèrement toutes les personnes de l'Institut de Médecine Légale de Strasbourg pour l'ambiance chaleureuse qu'elles créent dans le service et pour la gentillesse qu'elles m'ont témoignée tout au long de cette thèse.

Une petite pensée pour toutes les personnes avec qui j'ai eu la chance de partager quelques semaines, mois ou années au sein de l'institut : Sarah, merci pour ton aide et ton soutien au début de cette thèse, les anciens post-docs et thésards, Beti, Sylvain et Didier, merci pour tous vos conseils, et un grand merci aux stagiaires pour leur aide technique, Sophie R., Sophie L. et Arnaud.

Je souhaite remercier les responsables de l'équipe OpenLAB, le Professeur Catherine FLORENTZ et les Docteurs Laurence DROUARD et Michel LABOUESSE pour m'avoir permis de découvrir le métier d'enseignant. Merci également aux moniteurs de l'année 2009/2010, Judith, Isabelle, Thoueiba, Elise et Sara, pour la bonne ambiance des T.P. Je souhaite à vous toutes ainsi qu'aux nouveaux moniteurs une bonne continuation.

Un grand merci à tous ceux qui ont contribué, de près ou de loin, à ce travail et que je n'ai pas cité explicitement.

Enfin un immense merci à toute ma famille, et en particulier à mes parents et ma sœur, ainsi qu'à François et à mes amis que j'ai eu la chance d'avoir à mes côtés et qui m'ont toujours soutenue et encouragée depuis de nombreuses années. Tout ce travail n'aurait pas été possible sans vous...

Table des matières

Table des matières	1
Liste des figures	5
Liste des tableaux	7
Liste des abréviations et acronymes	9
Introduction générale.....	11
I. Rappels fondamentaux.....	13
I.1 Les études d'ADN ancien	15
I.1.1 Historique.....	15
I.1.2 Les particularités des études d'ADN ancien.....	17
I.1.2.1 L'ADN ancien : une molécule rare et dégradée.....	17
I.1.2.1.1 Les mécanismes de dégradation de l'ADN.....	17
I.1.2.1.2 Les conséquences sur la réaction d'amplification par PCR.....	19
I.1.2.1.3 Les facteurs influençant l'état de préservation de l'ADN.....	20
I.1.2.1.3.1 Les conditions de conservation des échantillons anciens	20
I.1.2.1.3.2 Les sources d'ADN ancien	21
I.1.2.2 La co-extraction d'inhibiteurs de la réaction d'amplification par PCR.....	21
I.1.2.3 Les contaminations	22
I.1.2.4 Les critères d'authenticité	23
I.1.3 Les champs d'application des études d'ADN ancien	25
I.1.3.1 L'étude d'ADN humain ancien.....	25
I.1.3.1.1 Les grandes énigmes historiques.....	25
I.1.3.1.2 La relation entre l'Homme de Néandertal et les hommes modernes	26
I.1.3.1.3 Le recrutement des ensembles funéraires	26
I.1.3.1.4 L'histoire des migrations de populations humaines.....	27
I.1.3.2 L'étude d'ADN de pathogènes anciens	28
I.1.3.2.1 Contexte général	29
I.1.3.2.2 L'exemple de la tuberculose	29
I.2 Les marqueurs moléculaires.....	33
I.2.1 Etudes de génomes humains anciens	33
I.2.1.1 L'ADN humain : rappels fondamentaux.....	33
I.2.1.2 Les marqueurs moléculaires.....	35

I.2.1.2.1 Les polymorphismes de longueur	35
I.2.1.2.1.1 Le locus de l'amélogénine	35
I.2.1.2.1.2 Les STR (Short Tandem Repeats)	35
I.2.1.2.2 Les SNP (Single Nucleotide Polymorphisms)	38
I.2.1.2.2.1 Les polymorphismes ponctuels de l'ADNmt	38
I.2.1.2.2.2 Les polymorphismes ponctuels de l'ADNnu	40
I.2.2 Etudes de génomes de pathogènes anciens : l'exemple de la tuberculose	44
I.2.2.1 Diagnostic de la tuberculose	44
I.2.2.2 Identification des espèces du CMTB	44
I.2.2.3 Typage moléculaire	45
I.2.2.4 Analyses phylogénétiques	47
II. Problématique et approche adoptée	49
II.1 Pourquoi les SNP ?	51
II.2 Les méthodes de typage des SNP	52
II.2.1 Le typage des SNP par spectrométrie de masse MALDI-TOF	53
II.2.1.1 Les intérêts	53
II.2.1.2 Le principe	54
II.2.2 Le typage des SNP par miniséquençage à l'aide du kit SNaPshot™	56
II.2.2.1 Les intérêts	56
II.2.2.2 Le principe	57
II.3 Les populations sibériennes étudiées	59
II.4 La stratégie méthodologique adoptée	61
III. Matériel et Méthodes	63
III.1 Echantillons biologiques	65
III.1.1 Echantillons contemporains	65
III.1.2 Echantillons anciens	66
III.2 Marqueurs moléculaires	69
III.2.1 Génome humain	69
III.2.1.1 SNP du chromosome Y	69
III.2.1.2 SNP de la région codante de l'ADNmt	69
III.2.1.3 SNP autosomaux	70
III.2.2 Génome des bactéries du CMTB	71
III.2.2.1 SNP	71
III.2.2.2 Séquence d'insertion IS6110	73

III.3	Extraction et quantification de l'ADN	73
III.3.1	Extraction, purification et concentration de l'ADN	73
III.3.1.1	Echantillons humains contemporains.....	73
III.3.1.2	Echantillons humains anciens	73
III.3.2	Quantification de l'ADN par PCR en temps réel.....	74
III.4	Analyse des SNP	76
III.4.1	Amplification par PCR des régions d'ADN contenant les SNP d'intérêt.....	76
III.4.1.1	Conception des amorces PCR	76
III.4.1.2	Validation des amorces PCR.....	77
III.4.1.3	PCR multiplexes.....	78
III.4.2	Typage des SNP par spectrométrie de masse MALDI-TOF.....	79
III.4.2.1	Réaction d'extension d'amorce ou PEX	79
III.4.2.1.1	Conception des amorces PEX	79
III.4.2.1.2	Validation des amorces PEX.....	79
III.4.2.1.3	Réactions PEX multiplexes.....	81
III.4.2.2	Spectrométrie de masse MALDI-TOF.....	81
III.4.3	Typage des SNP par miniséquençage à l'aide du kit SNaPshot™	83
III.4.3.1	Développement des réactions à partir des échantillons contemporains	83
III.4.3.1.1	Réactions de miniséquençage ou SBE	83
III.4.3.1.1.1	Conception des amorces SBE.....	83
III.4.3.1.1.2	Validation des amorces SBE	84
III.4.3.1.1.3	Réactions de miniséquençage multiplexes.....	84
III.4.3.1.2	Electrophorèse capillaire	85
III.4.3.1.3	Interprétation des données.....	86
III.4.3.2	Optimisation des réactions pour les échantillons anciens	86
III.5	Analyse de la séquence IS6110	87
III.6	Précautions prises lors de l'analyse des échantillons anciens	88
IV.	Résultats	91
IV.1	Analyse des SNP par spectrométrie de masse MALDI-TOF.....	93
IV.1.1	Résultats obtenus avec la méthode de purification « <i>genopure oligo</i> ».....	93
IV.1.2	Résultats obtenus avec la méthode de purification « <i>genostrep</i> »	96
IV.2	Analyse des SNP par miniséquençage à l'aide du kit SNaPshot™	97
IV.2.1	Analyse des SNP du génome humain	97
IV.2.1.1	SNP du chromosome Y	97

IV.2.1.2 SNP de la région codante de l'ADNmt.....	99
IV.2.1.3 SNP autosomaux	101
IV.2.2 Analyse des SNP du génome des agents pathogènes de la tuberculose.....	104
IV.2.2.1 SNP permettant d'identifier les bactéries du CMTB au niveau de l'espèce et de caractériser les groupes génétiques PGG.....	104
IV.2.2.2 SNP permettant la caractérisation des groupes phylogénétiques SCG	108
IV.3 Analyse de la séquence d'insertion IS6110	110
IV.4 Validation des résultats obtenus à partir des échantillons anciens.....	112
V. Discussion	115
V.1 Méthodes utilisées pour l'analyse des SNP.....	118
V.1.1 GenoSNIP/spectrométrie de masse MALDI-TOF	118
V.1.2 Miniséquençage à l'aide du kit SNaPshot™/électrophorèse capillaire.....	120
V.2 Apport de l'analyse des SNP à l'histoire du peuplement de la Sibérie.....	121
V.2.1 Analyse des SNP du chromosome Y.....	122
V.2.2 Analyse des SNP de la région codante de l'ADNmt.....	124
V.2.3 Analyse des SNP autosomaux	124
V.2.4 Hypothèses émises sur l'origine des sujets sibériens anciens étudiés.....	126
V.2.5 Intérêts de l'analyse des SNP de génomes humains anciens.....	127
V.3 Apport de l'analyse des SNP à l'histoire sanitaire de la Sibérie	128
V.3.1 Analyse de SNP du génome des agents pathogènes de la tuberculose.....	129
V.3.2 Hypothèses émises sur la présence de la tuberculose en Yakoutie	131
V.3.3 Intérêts de l'analyse des SNP du génome d'agents pathogènes anciens.....	132
VI. Conclusion et perspectives.....	133
VII. Références bibliographiques.....	139
VIII. Annexes	159
IX. Publications.....	169

Liste des figures

Figure 1 : Principe de la PCR.....	16
Figure 2 : Processus de dégradation chimique de l'ADN.	18
Figure 3 : Mal de Pott, enfant de la nécropole d'Adaïma, Haute-Egypte, 3600 avant J.-C....	31
Figure 4 : Carte schématique de l'ADNmt humain.....	34
Figure 5 : Interactions entre les principales protéines impliquées dans la mélanogénèse chez l'Homme (d'après Bouakaze et al., 2009).....	43
Figure 6 : Détermination des groupes phylogénétiques PGG grâce aux SNP <i>gyrA</i> ⁹⁵ et <i>katG</i> ⁴⁶³	48
Figure 7 : Principe de l'analyse d'un SNP par spectrométrie de masse MALDI-TOF.....	54
Figure 8 : Principe d'un spectromètre de masse MALDI-TOF.....	55
Figure 9 : Principe de l'analyse d'un SNP par miniséquençage à l'aide du kit SNaPshot™.	57
Figure 10 : Les différentes étapes d'une série de réactions multiplexes pour l'analyse simultanée de plusieurs SNP par miniséquençage à l'aide du kit SNaPshot™ (d'après Sanchez et al., 2006).....	58
Figure 11 : Carte géographique de la Sibérie.	59
Figure 12 : Photos des échantillons yakoutes anciens.....	68
Figure 13 : Arbre phylogénétique présentant les SNP du chromosome Y sélectionnés.	69
Figure 14 : Arbre phylogénétique présentant les SNP de la région codante de l'ADNmt sélectionnés.	70
Figure 15 : Arbre phylogénétique présentant les 8 SNP sélectionnés pour la différenciation les espèces du CMTB.....	72
Figure 16 : Spectres obtenus pour l'analyse du locus rs16891982 à partir de 3 extraits d'ADN humain contemporains qui présentent des génotypes homozygote C/C (A.), homozygote G/G (B.) et hétérozygote C/G (C.).	93
Figure 17 : Spectres obtenus pour l'analyse des loci <i>gyrA</i> ⁹⁵ et <i>katG</i> ⁴⁶³ à partir de l'extrait d'ADN de <i>M. tuberculosis</i> H37Rv.....	95
Figure 18 : Comparaison des méthodes de purification des produits d'extension « genopure oligo » (A.) et « genostrep » (Bruker Daltonics) (B.).	96
Figure 19 : Electrophorégramme obtenu pour l'analyse de 13 SNP du chromosome Y à partir d'un échantillon ancien affilié à l'haplogroupe Y R1a1.	99

Figure 20 : Electrophorégramme obtenu pour l'analyse de 9 SNP de la région codante de l'ADNmt à partir d'un échantillon ancien affilié à l'haplogroupe mitochondrial U5a1..... 100

Figure 21 : Electrophorégrammes obtenus pour l'analyse de 10 SNP autosomaux à partir d'1 ng d'ADN extrait à partir d'échantillons contemporains issus d'origines géographiques européenne (A.), asiatique (B.) et africaine (C.) et d'un échantillon ancien (D.). 102

Figure 22 : Diagramme triangulaire généré à l'aide du logiciel STRUCTURE représentant les affinités génétiques entre différentes populations humaines actuelles (européenne, asiatique et africaine) et les échantillons anciens provenant de la région de Krasnoyarsk. 104

Figure 23 : Electrophorégrammes obtenus pour l'analyse des 8 SNP permettant de différencier les espèces du CMTB et de caractériser les 3 groupes génétiques PGG à partir d'échantillons contemporains correspondant aux 7 espèces du CMTB..... 107

Figure 24 : Electrophorégrammes obtenus pour l'analyse des 4 SNP permettant de différencier les espèces *M. tuberculosis* et *M. bovis* et de caractériser les 3 groupes génétiques PGG à partir des échantillons yakoutes anciens provenant des tombes Batta Tcharana (A.), Atakh (B.) et Odjuluun 2 (C.) et des contrôles positifs *M. tuberculosis* H37Rv (D.) et *M. bovis* CIP 102426 (E.)..... 108

Figure 25 : Electrophorégrammes obtenus pour l'analyse des 8 SNP permettant de différencier les 7 groupes génétiques SCG à partir des échantillons *M. tuberculosis* H37Rv affilié au groupe SCG 6b (A.) et *M. bovis* CIP 102426 affilié au groupe SCG 7 (B.)..... 110

Figure 26 : Photo d'un gel d'agarose obtenu après la migration par électrophorèse de différents produits d'amplification des réactions de PCR IS-1 et IS-2 pour les échantillons Bouogaryma 2 et Batta Tcharana. 111

Figure 27 : Répartition géographique actuelle de l'haplogroupe Y R1a1 (M17)..... 123

Figure 28 : Répartition géographique des 6 lignées phylogénétiques de *M. tuberculosis* et *M. africanum* décrites par Gagneux et al (2006). 131

Liste des tableaux

Tableau 1 : Critères d'authenticité des résultats des études d'ADNa (d'après Cooper et Poinar, 2000).	24
Tableau 2 : Liste des échantillons anciens provenant de la région de Krasnoyarsk.	66
Tableau 3 : Liste des échantillons anciens provenant de Yakoutie.....	67
Tableau 4 : Informations concernant les SNP autosomaux sélectionnés.....	71
Tableau 5 : Détermination des SCG grâce aux SNP sélectionnés.	72
Tableau 6 : Paramètres appliqués pour l'analyse par spectrométrie de masse MALDI-TOF. 82	
Tableau 7 : Allèles obtenus pour les 13 SNP du chromosome Y d'intérêt analysés à partir des 10 échantillons masculins anciens provenant de la région de Krasnoyarsk.....	98
Tableau 8 : Allèles obtenus pour les 8 SNP permettant d'identifier les espèces du CMTB et de caractériser les 3 groupes génétiques PGG à partir des 56 échantillons bactériens contemporains testés.	105
Tableau 9 : Allèles obtenus pour les 8 SNP permettant de caractériser les 7 groupes génétiques SCG à partir de 37 échantillons bactériens contemporains.....	109

Liste des abréviations et acronymes

A	Adénine
AB	« Applied Biosystems »
ADN	Acide DésoxyriboNucléique
ADNa	Acide DésoxyriboNucléique ancien
ADNmt	Acide DésoxyriboNucléique mitochondrial
ADNnu	Acide DésoxyriboNucléique nucléaire
AIM	« Ancestry Informative Marker »
AMPc	Adénosine MonoPhosphate cyclique
BSA	« Bovine Serum Albumin »
C	Cytosine
CMTB	Complexe <i>Mycobacterium tuberculosis</i>
Da	Dalton
DCT	« DaupaChrome Tautomerase »
ddNTP	didésoxyNucléotide TriPhosphate
dNTP	désoxyNucléotide TriPhosphate
DR	« Direct Repeat»
G	Guanine
HPLC	« High Performance Liquid Chromatography »
HVR	« HyperVariable Region »
IPC	« Internal PCR Control »
IS6110	« Insertion Sequence 6110 »
LSP	« Large Sequence Polymorphism »
α -MSH	« α -Melanocyte Stimulating Hormone »
MAFSO	Mission Archéologique Française en Sibérie Orientale
MALDI-TOF	« Matrix Assisted Laser Desorption/Ionization-Time Of Flight »
MC1R	« MelanoCortin 1 Receptor »
MIRU-VNTR	« Mycobacterial Interspersed Repetitive Unit-Variable Number of Tandem Repeat »
MITF	« MIcroptalmia-associated Transcription Factor »
MOTT	« Mycobacteria other than <i>tuberculosis</i> »

NCBI	« National Center for Biotechnology Information »
nt	nucléotides
pb	paires de bases
PCR	« Polymerase Chain Reaction »
PEX	« Primer EXtension »
PGG	« Principal Genetic Group »
POP-4 TM	« Performance Optimum Polymer 4 »
rCRS	« revised Cambridge Reference Sequence »
RD	« Region of Difference »
rpm	rotation par minute
RFLP	« Restriction Fragment Length Polymorphism »
RFU	« Relative Fluorescence Unit »
SAP	« Shrimp Alcaline Phosphatase »
SBE	« Single Base Extension »
SCG	« SNP Cluster Group »
SNP	« Single Nucleotide Polymorphism »
spoligotyping	« spacer oligonucleotide genotyping »
STR	« Short Tandem Repeat »
T	Thymine
TYR	Tyrosinase
TYRP1	« Tyrosinase Related Peptide 1 »
U	Uracile

Introduction générale

Depuis une vingtaine d'années, il est possible d'étudier les acides désoxyribonucléiques (ADN) préservés dans des échantillons anciens grâce au développement des techniques de biologie moléculaire et, en particulier, à l'invention de la réaction de polymérisation en chaîne (PCR pour Polymerase Chain Reaction). Ces études dites d'ADN ancien (ADNa) visent à extraire et analyser l'ADN d'organismes humains, animaux, végétaux ou de micro-organismes à partir de substrats environnementaux ou biologiques datant d'une centaine d'années à environ 100 000 ans. Ces études permettent de mieux cerner l'évolution des espèces, des populations et des génomes et présentent donc des champs d'application nombreux et variés.

Si la plupart des études d'ADNa publiées à ce jour rapportent l'analyse d'ADN mitochondrial (ADNmt) en raison des caractéristiques particulières de ce type d'ADN, l'intérêt des paléogénéticiens pour l'ADN nucléaire (ADNnu) est grandissant. La majorité des études d'ADNa humain visant à analyser l'ADNnu ont porté sur l'amplification d'un fragment du gène de l'amélogénine (pour la détermination du sexe des individus) et sur l'analyse de microsatellites ou STR (pour Short Tandem Repeats) au niveau des autosomes (pour l'identification individuelle et la détermination de liens de proche parenté) et du chromosome Y (pour l'identification de lignées paternelles). Toutefois, l'ADNa ne peut souvent être exploité par le biais des STR étant donné que leur analyse nécessite l'amplification de fragments d'ADN dont la taille n'est souvent pas compatible avec l'état de dégradation de ces molécules anciennes. Afin d'augmenter les chances de succès d'analyse de l'ADNa, il s'avère donc nécessaire d'étudier des marqueurs nucléaires présentant une région informative de plus petite taille que les STR. C'est le cas des polymorphismes ponctuels de séquence ou SNP (pour Single Nucleotide Polymorphisms) qui constituent la forme la plus abondante de variations génétiques dans le génome humain. Ils font déjà l'objet d'un réel intérêt dans de nombreux domaines et présentent de nombreuses caractéristiques qui pourraient être intéressantes pour les études d'ADNa.

Dans ce contexte, l'objectif du travail de thèse présenté dans ce mémoire a été de déterminer si l'analyse des SNP permettait d'étendre les possibilités d'étude des génomes anciens. Ce travail a été réalisé au sein du laboratoire d'anthropologie moléculaire rattaché à l'équipe

d'accueil 4438 intitulée « Physiopathologie et Médecine Translationnelle ». L'un des principaux thèmes de recherche développés par ce laboratoire depuis plusieurs années porte sur l'histoire des populations sibériennes anciennes. Afin de déterminer ce que pouvait apporter l'analyse des SNP à cette thématique, nous avons sélectionné plusieurs types de SNP du génome humain et du génome de certains pathogènes. Des développements permettant le typage de ces SNP ont été réalisés au moyen de différentes méthodes sur des échantillons contemporains d'ADN humain et bactérien. Par la suite, certains de ces protocoles ont pu être appliqués avec succès à des échantillons humains anciens provenant de deux régions sibériennes distinctes. Etant donné que les résultats obtenus ont permis d'apporter de nouveaux éléments quant à l'origine et à l'état sanitaire des populations anciennes considérées, ce travail de thèse atteste bien de l'utilité des SNP pour les études d'ADNa.

Après un bref rappel de l'historique, des particularités et des champs d'application des études d'ADNa ainsi que des marqueurs moléculaires couramment utilisés, les raisons pour lesquelles les SNP semblent être des candidats de choix pour ces études seront énoncées dans la première partie de ce mémoire. La stratégie méthodologique qui a été adoptée dans ce travail sera ensuite détaillée puis les résultats obtenus à partir des différents échantillons analysés pour chacun des ensembles de SNP choisis seront exposés. Finalement, ces résultats seront discutés au regard des données actuelles afin de dégager les apports de l'analyse des SNP sélectionnés et de définir les perspectives de ces développements.

Ce travail de thèse a été validé par la publication de plusieurs articles qui sont joints à la fin de ce mémoire.

I. Rappels fondamentaux

I.1 Les études d'ADN ancien

I.1.1 Historique

La preuve que des molécules d'ADNa pouvaient être extraites et analysées à partir d'échantillons anciens a été apportée pour la première fois au milieu des années 1980 à partir de tissus mous d'un couagga taxidermisé, espèce du genre *Equus* proche du zèbre qui s'est éteinte à la fin du XIX^{ème} siècle (Higuchi et al., 1984). La première séquence nucléotidique humaine, isolée à partir de tissus mous prélevés sur une momie égyptienne datée de 2400 ans, n'a pas tardé à être publiée (Paabo, 1985; Paabo et al., 1985). Jusqu'à la fin des années 1980, le nombre de ces études est resté relativement restreint puisqu'elles se sont très vite heurtées à des limitations techniques liées à la nature dégradée de l'ADNa (Paabo et al., 1989).

Les études d'ADNa ont réellement pris leur essor grâce à l'avènement de la PCR (Saiki et al., 1985 ; Mullis et Faloona, 1987). Cette technique permet d'amplifier de l'ADN en plusieurs centaines de milliers de copies à partir d'infimes quantités de départ voire une seule copie en théorie (Figure 1). Elle a été utilisée avec succès pour la première fois sur de l'ADN extrait à partir d'os et de dents anciens au début des années 1990 (Hagelberg et al., 1989; Hanni et al., 1990). Ces travaux ont démontré que les tissus durs étaient de très bonnes, voire meilleures, sources d'ADNa que les tissus mous. Aujourd'hui encore, toutes les études d'ADNa comprennent une étape d'amplification d'ADN par PCR traditionnelle ou une variante.

Au début des années 1990, l'enthousiasme extrême des scientifiques à l'égard des études d'ADNa alimenta une course à la recherche de la séquence d'ADN la plus ancienne. Plusieurs travaux spectaculaires publiés dans des revues scientifiques internationales très réputées ont rapporté l'extraction et l'amplification d'ADN extrait à partir de végétaux datant d'environ 20 millions d'années (Golenberg et al., 1990; Soltis et al., 1992), d'os de dinosaures datés d'environ 80 millions d'années (Woodward et al., 1994) et d'insectes vieux de plus de 100 millions d'années piégés dans de l'ambre (Cano et al., 1993). Très rapidement discrédités, ces travaux ont néanmoins mis en évidence l'une des difficultés majeures de l'étude d'ADNa : la contamination (Gibbons, 1994; Austin et al., 1997; Gutierrez et Marin, 1998). Ces travaux ont abouti à la publication de critères d'authentification des séquences d'ADNa (Cooper et Poinar, 2000) et de nombreux progrès méthodologiques ont été réalisés par la suite grâce à une meilleure connaissance des caractéristiques de l'ADNa.

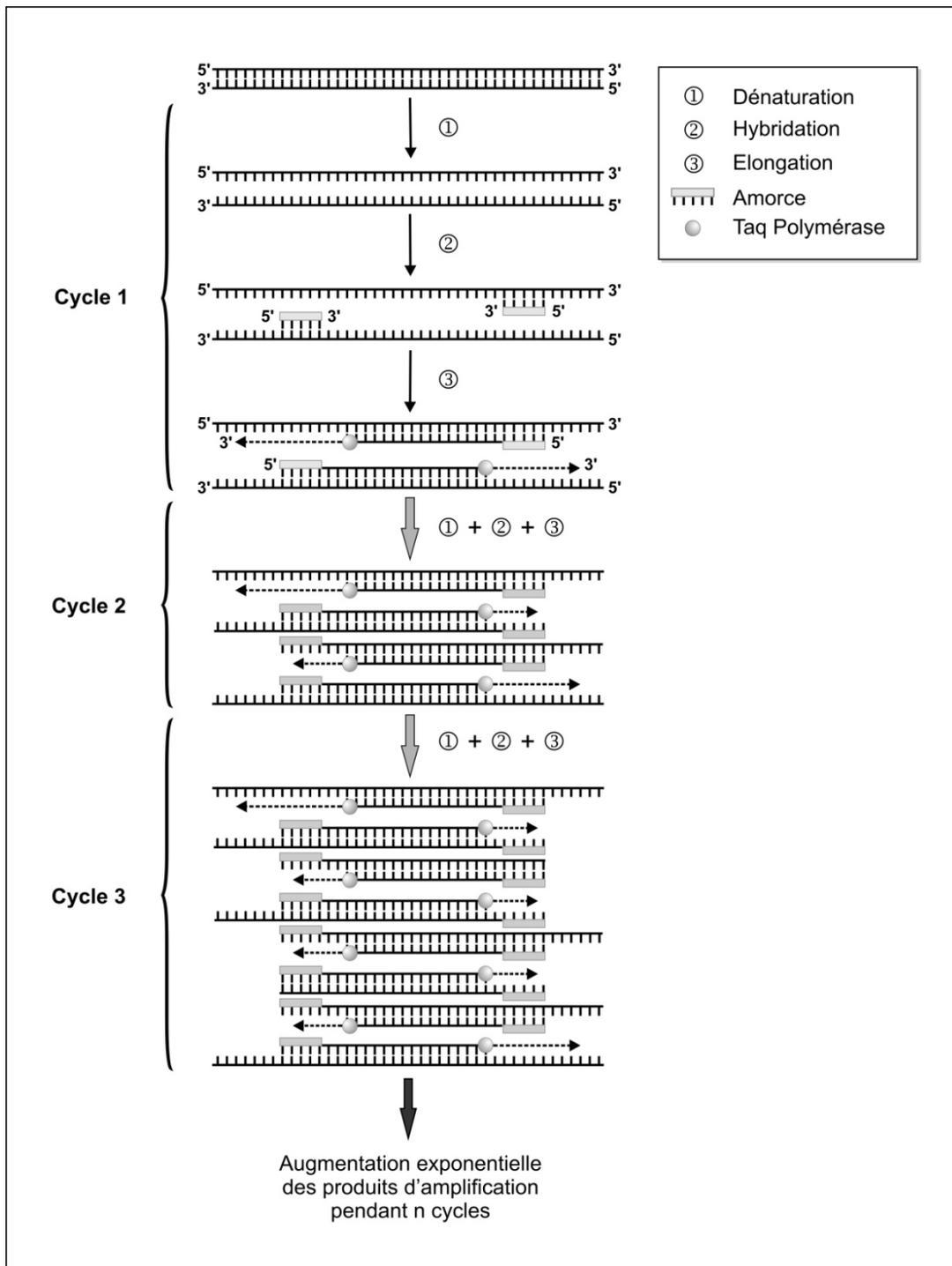


Figure 1 : Principe de la PCR.

I.1.2 Les particularités des études d'ADN ancien

I.1.2.1 L'ADN ancien : une molécule rare et dégradée

L'état de préservation de l'ADN est variable d'un échantillon à l'autre et la persistance de l'ADN dans les restes anciens n'est pas éternelle. Cela s'explique par le fait que l'ADN est une molécule instable au cours du temps puisqu'elle est sans cesse soumise à des processus de dégradation (Lindahl, 1993).

I.1.2.1.1 Les mécanismes de dégradation de l'ADN

Dans les cellules vivantes, les processus de dégradation spontanés de l'ADN sont généralement contrebalancés par des mécanismes spécifiques de réparation qui assurent le maintien de l'intégrité du génome. Cependant après la mort de l'organisme, les dommages subis par l'ADN ne sont plus réparés et s'accumulent au cours du temps. Deux types de mécanismes de dégradation *post-mortem* peuvent être distingués : les processus enzymatiques et les processus chimiques.

Dès la mort de l'organisme, la rupture des compartiments cellulaires entraîne la libération d'enzymes lysosomales comprenant des nucléases. Les endo-nucléases vont cliver les liaisons phosphodiester de l'ADN, et par conséquent, le fragmenter et les exo-nucléases vont s'attaquer à l'extrémité des fragments formés et réduire encore leur taille. Ce phénomène d'autolyse est ensuite complété par l'action des enzymes libérées par les organismes saprophytes (insectes, champignons et bactéries) qui se succèdent sur l'organisme mort et permettent la décomposition des tissus.

Les processus de dégradation chimique sont des mécanismes beaucoup plus lents. Comme dans les cellules vivantes, il s'agit majoritairement de phénomènes d'hydrolyse et d'oxydation (Lindahl, 1993). Les réactions d'hydrolyse peuvent accentuer le phénomène de fragmentation de l'ADN en s'attaquant aux liaisons phosphodiester. De plus, elles peuvent rompre les liaisons N-glycosidiques liant les bases azotées aux désoxyriboses, ce qui, en entraînant la formation de sites abasiques, fragiliserait la structure de l'ADN et favoriserait sa fragmentation. Ce type de réactions peut également entraîner des modifications au niveau des bases azotées telles que la perte de groupements amines induisant fréquemment la transformation d'une cytosine (C) en uracile (U). Les réactions d'oxydation peuvent également résulter en la formation de bases modifiées ou de dommages au niveau du squelette ribose-phosphate. D'autres types de réactions chimiques peuvent survenir au cours du temps

et conduire à la formation de liaisons intra-moléculaires, *e.g.* l'alkylation qui peut entraîner un mésappariement entre deux brins d'ADN, ou inter-moléculaires, *e.g.* la réaction de Maillard qui produit des complexes stables ADN-protéine (Willerslev et Cooper, 2005). Ces différents types de dégradation chimique sont représentés sur la figure 2.

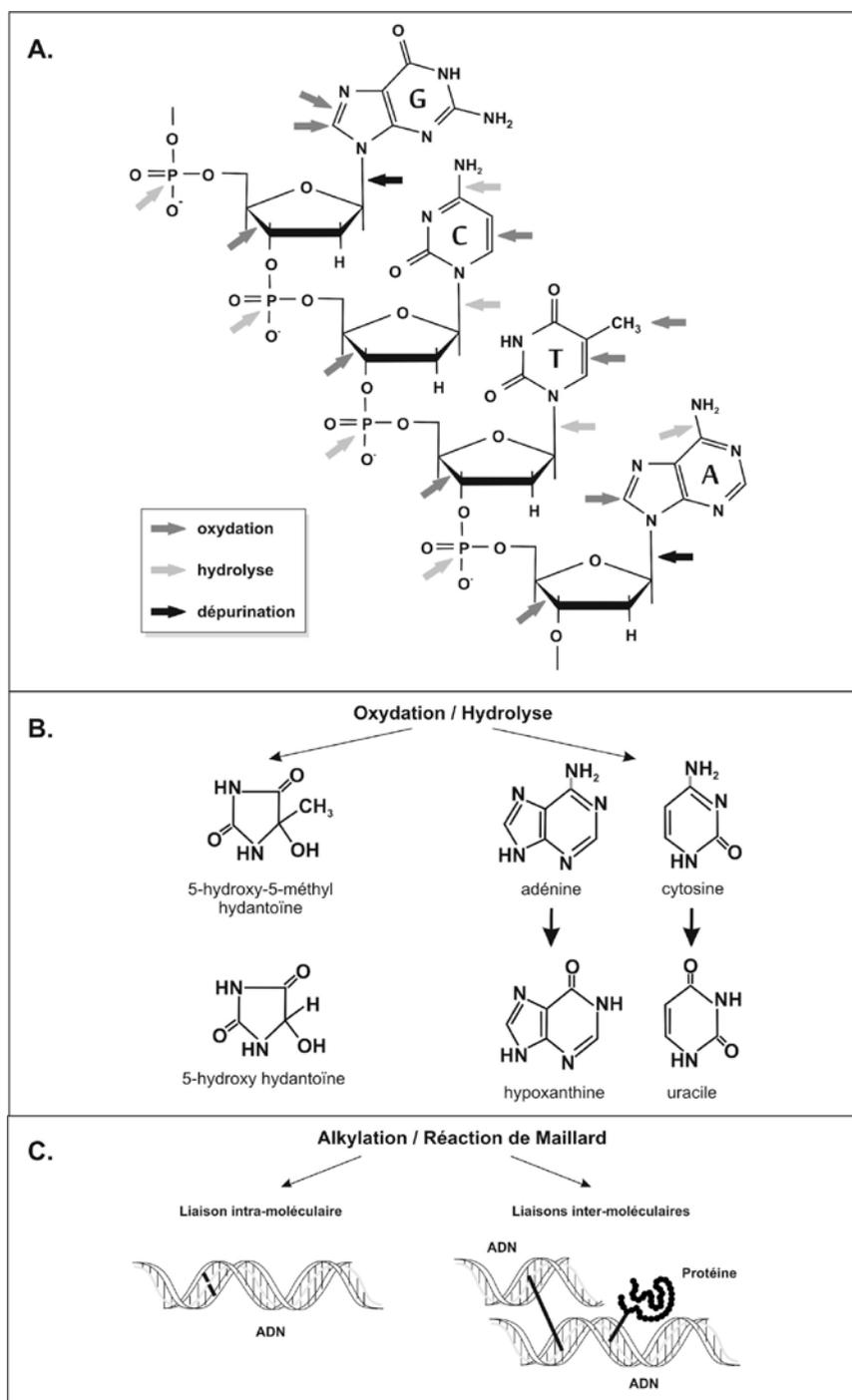


Figure 2 : Processus de dégradation chimique de l'ADN.

Les sites de lésions sur la molécule d'ADN (A.) et des exemples de bases modifiées engendrées par ces processus (B.) et de réactions chimiques conduisant à la formation de liaisons moléculaires (C.) sont représentés sur cette figure d'après Hofreiter et al., 2001 et Willerslev et Cooper, 2005.

1.1.2.1.2 Les conséquences sur la réaction d'amplification par PCR

Les dommages *post-mortem* de l'ADN, qui comprennent des cassures de brins, des sites abasiques ou encore des nucléotides modifiés, se répercutent à la fois sur la qualité et la quantité de l'ADN, et, de ce fait, limitent l'amplification par PCR.

La fragmentation, qui est un processus encore aujourd'hui non réparable et donc irréversible, est l'une des principales difficultés des études d'ADNa. Il a été démontré que la majorité des spécimens anciens ne contiennent plus d'ADN amplifiable et que les autres ne contiennent généralement plus que des fragments d'ADN endogène dont la taille est comprise entre 100 et 500 pb (Willerslev et Cooper, 2005). Par conséquent, seuls des fragments de petite taille sont amplifiables par PCR à partir des extraits d'ADNa. L'utilisation de différentes paires d'amorces permettant d'amplifier des fragments d'ADN chevauchants est une méthode souvent utilisée pour analyser des séquences anciennes plus longues.

Les modifications chimiques de l'ADN, telles que la création de sites abasiques ou de bases modifiées, peuvent conduire à l'incorporation de bases incorrectes lors de l'amplification. Par exemple, la désamination de C en U (un analogue de la thymine (T)) par réaction d'hydrolyse peut provoquer une erreur de lecture (T à la place de C) et donc d'élongation de la Taq polymérase. Ainsi, des transitions C → T (ou guanine (G) → adénine (A) selon le brin séquencé) sont fréquemment observées dans les séquences anciennes (Gilbert et al., 2003; Brotherton et al., 2007). Ces erreurs de la Taq polymérase sont facilement détectables grâce à la comparaison de plusieurs séquences obtenues à partir de différentes réactions d'amplification indépendantes les unes des autres ou de différents clones issus du clonage des produits PCR d'une seule réaction d'amplification.

Alors que dans le cas précédent l'élongation est possible mais des bases erronées sont générées, d'autres modifications chimiques de l'ADN peuvent complètement bloquer l'élongation des produits PCR. C'est le cas des hydantoïnes qui sont des produits d'oxydation des pyrimidines. Récemment, les ADN polymérases chimériques hautement fidèles et dont l'activité polymérase ne serait pas bloquée par des composés tels que les hydantoïnes font l'objet de nouveaux développements (Gilbert et Willerslev, 2007). Par ailleurs, les liaisons

moléculaires, *e.g.* les produits de Maillard, mobilisent l'ADN endogène et, de ce fait, bloquent également l'amplification par PCR par encombrement stérique.

L'état fragmenté et modifié de l'ADN peut également mener à des événements de recombinaisons aléatoires lors de l'amplification par PCR. Ce phénomène, appelé « jumping PCR », produit généralement des amplicons de taille plus grande que celle attendue. Il s'agit de contaminations aspécifiques ou d'amplicons chimères générés lors des premiers cycles d'amplification au cours desquels l'information génétique présente en différents loci est recombinée aléatoirement. Les mécanismes à l'origine de ces artefacts d'amplification sont encore mal connus.

1.1.2.1.3 Les facteurs influençant l'état de préservation de l'ADN

1.1.2.1.3.1 Les conditions de conservation des échantillons anciens

Les processus de dégradation enzymatique et chimique de l'ADN sont influencés par des facteurs physico-chimiques tels que le pH, la température, la pression, les taux d'oxygène et d'humidité. Par exemple, une dessiccation rapide, des températures faibles et des concentrations élevées en sels sembleraient favoriser la dégradation ou l'inactivation des nucléases et, par conséquent, ralentiraient les processus enzymatiques de fragmentation de l'ADN (Hofreiter et al., 2001). A l'inverse, certaines conditions telles que des températures élevées, la présence de rayonnements ionisants, un pH acide, un taux élevé en oxygène et un degré d'humidité important favoriseraient l'activité enzymatique et/ou les réactions chimiques et donc la dégradation de l'ADN. Ainsi, le principal facteur qui conditionne la préservation de l'ADN au cours du temps, n'est pas le temps lui-même, mais bien le milieu environnemental dans lequel les spécimens anciens ont été conservés depuis la mort de l'organisme.

La durée de préservation de l'ADN peut varier de quelques années à des centaines de milliers d'années suivant les conditions taphonomiques. Il a été estimé que l'ADN était préservé dans un milieu tempéré jusqu'à environ 10 000 ans et que seul un environnement froid permettait de conserver l'ADN jusqu'à 100 000 ans (Lindahl, 1993; Poinar et al., 1996; Willerslev et al., 2004). En raison de températures constamment faibles et de l'absence d'oxygène, les glaciers polaires et le pergélisol (sols gelés en permanence) semblent être les milieux de conservation qui permettraient de préserver l'ADN le plus longtemps possible (Willerslev et al., 2004).

I.1.2.1.3.2 Les sources d'ADN ancien

En théorie, tout tissu ancien composé à l'origine de cellules vivantes est susceptible de contenir de l'ADN. Néanmoins, dans les mêmes conditions de conservation, l'ADN est mieux préservé dans les os et les dents que dans les tissus mous (Cooper et al., 1992). C'est pourquoi les tissus durs sont aujourd'hui les substrats les plus utilisés dans les études d'ADNa. La meilleure préservation de l'ADN dans ces tissus semblerait être due à l'adsorption de l'ADN sur la matrice minérale de ces tissus constituée essentiellement d'hydroxyapatite de calcium. Par ailleurs, les os et les dents sont mieux conservés au cours du temps et sont donc les substrats les plus abondants. Néanmoins, d'autres restes peuvent également contenir de l'ADNa tels que les tissus mous, les cheveux, les poils ou encore les coprolithes.

Il ne s'agit pas ici de faire une liste exhaustive et détaillée des différents substrats qui renferment de l'ADNa mais de retenir leur diversité et que l'état de préservation de l'ADN n'est pas le même quelque soit le type de reste biologique analysé.

I.1.2.2 La co-extraction d'inhibiteurs de la réaction d'amplification par PCR

Il semble logique de penser qu'un ADN bien préservé sera facilement amplifiable ; or ce n'est pas forcément le cas. Des molécules, co-purifiées avec l'ADNa lors de l'étape d'extraction, peuvent souvent diminuer l'efficacité de la réaction d'amplification voire même la rendre impossible. Il peut s'agir de substances contenues dans le sol comme les acides humiques et fulviques (Tuross, 1994) ou de constituants même de l'organisme mort comme l'hème de l'hémoglobine (Akane et al., 1994), le collagène de type I (Scholz et al., 1998), et la mélanine contenue dans les cheveux (Uchihi et al., 1992). Le mode d'action de ces inhibiteurs est encore mal connu.

Néanmoins, la présence d'inhibiteurs de PCR dans un extrait d'ADN peut être facilement détectée grâce à l'utilisation de techniques d'amplification par PCR en temps réel, *e.g.* le kit Quantifiler™ Human DNA Quantification Kit (Applied Biosystems, AB) (cf. Matériel et Méthodes). De plus, des méthodes peuvent être utilisées pour limiter l'action de ces inhibiteurs. Une simple dilution en cascade est souvent réalisée à partir des extraits d'ADN où la présence d'inhibiteurs a été détectée. Elle permet de diluer les inhibiteurs et donc de réduire leurs effets mais elle implique également la dilution de l'ADNa qui est généralement déjà présent en faible quantité. L'ajout de certains additifs à différentes étapes permettrait également de s'affranchir de l'action de certains inhibiteurs. Par exemple, l'ajout de PTB (N-phenacylthiazolium bromide) dans le tampon de lyse lors de l'étape d'extraction d'ADN

permettrait de dissocier les produits de Maillard et donc de libérer l'ADN (Poinar et al., 1998). De même, l'ajout d'albumine de sérum bovin (BSA pour Bovine Serum Albumin) dans le mélange réactionnel de PCR diminuerait l'action de certains inhibiteurs grâce à son affinité pour les porphyrines, composés phénoliques présents dans les acides humiques ou fulviques et les tannins du sol (Kreider, 1996; O'Rourke et al., 2000).

I.1.2.3 Les contaminations

S'il est possible de détecter et de s'affranchir des effets des inhibiteurs de la PCR, en revanche ce n'est pas toujours le cas pour les contaminations. Le risque de contamination est un problème majeur des études d'ADNa. Il est principalement lié à l'état dégradé de l'ADN extrait de tissus anciens. Tout apport, même en quantité infime, d'ADN moderne exogène est susceptible d'entrer en compétition avec les quelques rares molécules préservées d'ADNa endogène lors de la PCR. Dans ce cas, l'avantage principal de la PCR qui est son extrême sensibilité est aussi sa principale limitation puisque l'ADN contaminant, qui est une bien meilleure matrice que l'ADN endogène, est préférentiellement amplifié. Les études d'ADNa se focalisant sur l'analyse d'ADN humain ancien sont particulièrement sensibles aux contaminations en raison de la difficulté de différencier de l'ADN humain ancien endogène et de l'ADN moderne contaminant provenant de manipulateurs.

A chaque moment, depuis l'inhumation des restes anciens jusqu'à l'amplification par PCR dans le laboratoire, toute personne en contact direct ou indirect avec les restes anciens ou les extraits d'ADNa est une source potentielle de contamination. Au sein du laboratoire, en plus du personnel de laboratoire, le risque peut provenir de réactifs ou de consommables et même des locaux. En effet, les microgouttelettes d'ADN, générées puis dispersées par les mouvements d'air créés lors de l'ouverture des tubes ou des transferts de liquide, peuvent être à l'origine de contaminations croisées entre différents échantillons (Willerslev et Cooper, 2005). Cela est particulièrement problématique lorsqu'il s'agit de microgouttelettes d'ADN amplifié qui contiennent des millions de copies d'ADN. Par ailleurs, la poussière présente dans le laboratoire peut, elle aussi, renfermer une quantité insoupçonnée et non négligeable d'ADN (Toothman et al., 2008). Les ADN humain et bactérien sont omniprésents dans tous les laboratoires (Willerslev et Cooper, 2005).

Néanmoins, de nombreuses précautions pour éviter de contaminer les échantillons peuvent être prises. Sur le site de fouilles, toutes les personnes en contact avec des échantillons anciens devraient porter un masque facial et des gants lors des manipulations. Aussi, l'ADN

de toutes ces personnes devrait-il être prélevé et analysé pour les mêmes marqueurs que ceux analysés sur les échantillons. La comparaison des résultats obtenus à partir de l'échantillon ancien et ceux obtenus à partir des personnes génotypés sur le site de fouille permet en effet d'éliminer un risque de contamination.

Au sein du laboratoire, le risque de contamination par le personnel est contrôlé par le port obligatoire d'une blouse, d'une double paire de gants et d'un masque facial ainsi que par le génotypage de toute personne ayant ou ayant eu accès au laboratoire. Concernant l'organisation du laboratoire, deux parties sont généralement physiquement séparées : le laboratoire d'ADN ancien où sont réalisées toutes les étapes en amont de l'amplification par PCR et le laboratoire d'ADN moderne où a réellement lieu la PCR ainsi que toutes les étapes post-PCR. Le stockage et le pré-traitement (décontamination, cryobroyage...) des tissus biologiques anciens sont généralement réalisés dans une pièce séparée du laboratoire de biologie moléculaire où sont réalisées l'extraction d'ADN et la préparation du mélange réactionnel de PCR. Ces étapes sont souvent réalisées sous des hottes à flux laminaires équipées de lampes à UV. Les paillasse, les réactifs et le matériel sont toujours décontaminés entre l'analyse de chaque échantillon pour éviter les contaminations croisées. Cette décontamination peut être réalisée au moyen d'une irradiation aux UV (254 nm), d'un autoclavage ou d'un traitement avec de l'hypochlorite de sodium ou des solutions contenant des DNases. Le cheminement des personnes, des réactifs et du matériel entre les laboratoires d'ADN ancien et moderne ne doit se faire que dans le sens pré-PCR vers post-PCR.

Une procédure de décontamination des échantillons anciens pour éliminer les contaminations de surface est le plus souvent réalisée. Selon le type de substrat à analyser, une décontamination physique, qui consiste à abraser quelques millimètres de la surface du spécimen à l'aide d'un scalpel, de papier de verre ou d'une mini-perceuse Dremel[®], ou une décontamination chimique, telle que le traitement à l'hypochlorite de sodium ou une brève irradiation aux UV, sera choisie.

I.1.2.4 Les critères d'authenticité

Déjà au début des études d'ADNa, les scientifiques se sont rendu compte qu'il fallait établir des critères pour s'assurer que les résultats ne provenaient pas d'ADN contaminants mais bien d'ADNa autrement dit de l'authenticité des résultats. Au départ, de simples précautions pour détecter les contaminations telles que l'utilisation de témoins négatifs d'extraction et l'analyse de multiples extraits d'ADN par échantillon biologique ainsi que l'observation d'une

corrélation inverse entre l'efficacité de l'amplification par PCR et la taille des produits PCR ont été recommandées (Paabo, 1989). A mesure, ces 3 critères ont évolué pour conduire en 2000 à la publication d'une liste de 9 critères nécessaires à l'authentification des résultats des études d'ADNa (Cooper et Poinar, 2000). Ces critères sont bien connus de la communauté des paléogénéticiens (Tableau 1).

Séparation physique des zones de travail. Il est essentiel que toutes les étapes réalisées sur les échantillons anciens en amont de l'étape d'amplification par PCR (y compris la préparation du mélange réactionnel de PCR) soient effectuées dans un laboratoire d'ADNa dédié à cet effet.

Utilisation de contrôles de réactions. Il est indispensable de réaliser des contrôles négatifs en parallèle de chaque extraction et amplification. De plus, les contrôles positifs devraient être évités puisqu'ils constituent un risque de contamination.

Observation d'une corrélation inverse entre l'efficacité de PCR et la taille des amplicons. L'ADNa étant fragmenté, l'efficacité de l'amplification devrait être inversement proportionnelle à la taille des produits amplifiés.

Reproductibilité des résultats. Les amplifications doivent être multipliées à partir d'un même extrait et de différents extraits d'un même sujet. Les résultats obtenus à partir de ces amplifications doivent être concordants.

Clonage. Les séquences nucléiques doivent être vérifiées grâce au clonage des produits amplifiés et au séquençage de nombreux clones. Cette précaution permettrait de détecter les hétérogénéités dans les amplicons pouvant provenir de contaminations, de la dégradation des molécules d'ADN ou de phénomènes de « jumping PCR ».

Réplication des analyses dans un autre laboratoire. Seul le fait que les mêmes résultats aient été obtenus par deux laboratoires indépendants permet d'éliminer le risque de contamination intra-laboratoire.

Estimation de l'état de préservation biochimique. Il est possible d'estimer l'état de préservation de l'ADN dans les échantillons anciens au moyen de différentes méthodes indirectes tels que l'étude des acides aminés.

Quantification de l'ADN. La quantification du nombre de molécules présentes dans l'extrait d'ADNa permet de déterminer dans quelle mesure les résultats sont fiables. Si le nombre de molécules est inférieur à 1000, la contamination sporadique ne peut pas être exclue.

Etude de restes animaux associés. La vérification de la présence d'ADN amplifiable dans des restes associés permet de s'assurer de ne pas détruire inutilement les prélèvements humains anciens d'intérêt. Ces restes animaux associés peuvent également servir de contrôles négatifs si les fragments à amplifier sont humains.

Tableau 1 : Critères d'authenticité des résultats des études d'ADNa (d'après Cooper et Poinar, 2000).

La plupart de ces critères sont acceptés mais tous ne font pas l'unanimité. D'ailleurs, rares sont les études d'ADNa qui les respectent scrupuleusement tous. Il est important de souligner que la validation de tous les critères n'est pas une preuve de l'authenticité des résultats obtenus, en particulier, lorsque l'échantillon ancien a été contaminé en amont de l'analyse au laboratoire (Paabo et al., 2004). C'est pourquoi ces dernières années, certains chercheurs se

sont demandés si la publication de ces critères n'avait pas entraîné moins de prudence et de réflexion dans les études d'ADNa (Gilbert et al., 2005). Ainsi, il a été proposé de valider les résultats des études d'ADNa non pas en respectant tous ces critères à la lettre mais en adoptant une approche plus réfléchie et un raisonnement d'auto-critique. Il s'agit là d'expliquer comment les données anciennes ont été obtenues et pourquoi elles semblent être authentiques avec suffisamment de détails pour qu'il n'y ait aucune ambiguïté sur l'authenticité des résultats.

I.1.3 Les champs d'application des études d'ADN ancien

Les études d'ADNa suscitent aujourd'hui un intérêt certain dans de nombreux domaines tels que (i) la systématique et la paléontologie puisqu'il est désormais possible de définir les positions phylogénétiques d'espèces éteintes, *e.g.* le couagga (Higuchi et al., 1984) ou le mammouth (Noro et al., 1998), en comparant leur génome à ceux de leurs homologues actuels, (ii) l'histoire et l'évolution des populations animales et végétales et en particulier l'impact de la domestication au néolithique, (iii) l'histoire et l'évolution des populations humaines (iv) et l'étude des maladies infectieuses du passé ou la paléomicrobiologie.

Dans ce travail de thèse, nous nous limiterons à détailler les applications des études d'ADNa réalisées à partir de restes humains et, plus précisément, aux études d'ADNa humain et celles de pathogènes infectieux.

I.1.3.1 L'étude d'ADN humain ancien

I.1.3.1.1 Les grandes énigmes historiques

L'analyse d'ADN humain à partir de restes humains anciens a déjà permis d'élucider certaines grandes énigmes historiques grâce à l'identification de restes appartenant à des personnes célèbres. Par exemple, l'identification des restes squelettiques des 7 membres de la famille Romanov comprenant le Tsar Nicholas II de Russie, sa femme et ses 5 enfants (Gill et al., 1994; Coble et al., 2009; Rogaev et al., 2009) a permis de démasquer l'imposture d'Anna Anderson qui prétendait être l'une des filles. Par ailleurs, les analyses génétiques réalisées sur le cœur présumé de Louis XVII, fils du roi de France Louis XVI et de Marie-Antoinette ont pu confirmer la mort de Louis XVII à Paris le 8 juin 1795 (Jehaes et al., 2001).

1.1.3.1.2 La relation entre l'Homme de Néandertal et les hommes modernes

La relation phylogénétique entre les Néandertaliens et les Hommes anatomiquement modernes est un sujet de discussion depuis la découverte en 1856 du premier Homme de Néandertal dans la grotte de Feldhofer située dans la vallée de Néander en Allemagne. Des analyses génétiques ont été réalisées à partir de ce spécimen en 1997 et par la suite sur une quinzaine d'autres spécimens néandertaliens retrouvés en Europe et en Asie Centrale (Krings et al., 1997; Krings et al., 2000; Ovchinnikov et al., 2000; Schmitz et al., 2002; Serre et al., 2004; Beauval et al., 2005; Lalueza-Fox et al., 2005; Lalueza-Fox et al., 2006; Orlando et al., 2006; Krause et al., 2007b). Les résultats de ces travaux sont concordants et suggèrent que les Néandertaliens n'auraient pas contribué au pool génétique mitochondrial des hommes modernes ou très faiblement. Autrefois considérés comme une sous-espèce au sein de l'espèce *Homo sapiens* sur la base de caractères discrets, ils sont désormais considérés comme formant un groupe monophylétique, la lignée néandertalienne ou *Homo neanderthalensis*, séparé de celui des Hommes anatomiquement modernes ou *Homo sapiens* (Hodgson et Disotell, 2008). Ainsi, les Néandertaliens seraient apparus il y a plusieurs centaines de milliers d'années, auraient cohabité avec les hommes modernes pendant plus de 10 000 ans sans trace apparente de métissage et auraient disparu il y a environ 30 000 ans. Aujourd'hui, les causes de cette extinction ne sont toujours pas clairement identifiées.

1.1.3.1.3 Le recrutement des ensembles funéraires

Outre l'analyse génétique d'échantillons anciens isolés, les études d'ADNa peuvent également être réalisées à partir de plusieurs sujets inhumés dans un même ensemble sépulcral ancien qu'il s'agisse d'une nécropole ou d'une sépulture collective isolée. Le but de telles études est d'obtenir des éléments quant aux pratiques funéraires des populations humaines du passé. Il est possible de définir l'organisation et le mode de recrutement d'un ensemble sépulcral en déterminant le sexe des sujets analysés et/ou les possibles relations de parenté. Différents degrés de parenté peuvent être étudiés en fonction des marqueurs moléculaires analysés : proche parenté ou parenté directe (parents-enfants), parenté du côté maternel (même lignée maternelle) ou du côté paternel (même lignée paternelle).

Les études d'ADNa qui se sont cantonnées au sexage génétique sont peu nombreuses puisque leur intérêt reste relativement limité. Néanmoins, la détermination du sexe des juvéniles ou immatures étant peu fiable avec les méthodes de l'anthropologie classique, la détermination

génétique du sexe de juvéniles a déjà permis d'apporter des informations relatives à des pratiques d'infanticides (Faerman et Bar-Gal, 1998).

Récemment, une étude d'ADNa réalisée sur une sépulture collective datant de 4 600 ans découverte en Allemagne a révélé la présence de 4 membres d'une même famille : le père, la mère et deux enfants (Haak et al., 2008). Cette étude a apporté la preuve que les liens biologiques étaient importants pour cette société et que la structure familiale classique, appelée famille nucléaire (père, mère et enfants), existe au moins depuis la fin de l'Age de Pierre en Europe.

Par ailleurs, l'étude réalisée au sein du laboratoire dans lequel ce travail de thèse a été effectué portant sur la nécropole d'Egyin Gol en Mongolie, qui fut utilisée du III^{ème} siècle avant J.-C. au II^{ème} siècle après J.-C., reste à ce jour l'une des plus importantes études d'ADNa en termes d'échantillons analysés (Keyser-Tracqui et al., 2003). L'analyse génétique de 62 sujets a révélé des liens de filiations directes entre des individus inhumés à proximité les uns des autres. Cela laisse penser que les Xiongnu, guerriers nomades contre lesquels les chinois ont bâti la Grande Muraille de Chine, pratiquaient un recrutement funéraire à caractère familial sectoriel.

1.1.3.1.4 L'histoire des migrations de populations humaines

Le peuplement actuel des différentes régions du monde est le reflet des différentes vagues de migration des populations humaines au cours du temps. Depuis de nombreuses années, l'origine et l'histoire de ces mouvements de populations sont déduites sur la base de données historiques, linguistiques, archéologiques et, plus récemment, génétiques. En effet, ces 10 dernières années, de nombreuses hypothèses de peuplement ont pu être proposées grâce à l'analyse de la diversité génétique des populations actuelles (Underhill et Kivisild, 2007).

En avril 2005, a été initiée une vaste étude d'anthropologie moléculaire, fruit d'une collaboration entre des universités internationales et des industriels, intitulée « Le Projet Génographique » (ou The Genographic Project). Programmé sur 5 ans, ce projet a pour but de cartographier les migrations humaines sur les différents continents grâce au génotypage de centaines de milliers de personnes du monde entier (<https://genographic.nationalgeographic.com/genographic/lan/fr/index.html>).

L'accès direct au génome d'individus du passé par le développement des études d'ADNa permet aujourd'hui d'appréhender directement la diversité génétique de populations anciennes

et d'ancrer les évènements dans le temps. Le peuplement de différentes régions du monde peut désormais être abordé grâce à une approche de génomique comparative : la confrontation de données génétiques anciennes et modernes.

Par exemple, en 2005, une équipe allemande a tranché sur une question débattue depuis longtemps : l'origine des Européens actuels (Haak et al., 2005). Les ancêtres des Européens actuels sont-ils des agriculteurs du Proche-Orient qui auraient migrés en Europe au Néolithique il y a entre 10 000 et 5 000 ans (théorie du remplacement) ou bien des chasseurs-cueilleurs du Paléolithique installés en Europe depuis plus de 40 000 ans qui auraient appris et adopté la culture des nouveaux arrivants (théorie de la diffusion culturelle) ? Les résultats de l'analyse génétique de 24 fermiers néolithiques provenant de différents sites d'Allemagne, de Hongrie et d'Autriche ont révélé que ces fermiers néolithiques présentaient des mutations particulières retrouvées à des fréquences élevées et peu fréquentes chez les Européens actuels. Sur la base de cette particularité génétique, les auteurs en ont conclu que les agriculteurs néolithiques ne sont probablement pas les ancêtres des Européens et que la théorie de la diffusion culturelle est donc la plus vraisemblable. Les conclusions de cet article ont été très rapidement remises en cause (Ammerman et al., 2006; Barbujani et Chikhi, 2006). L'authenticité des données génétiques n'est pas contestée mais ce sont plutôt l'échantillonnage et les hypothèses testées par des simulations qui font l'objet de polémiques. Pour l'instant, les études populationnelles utilisant des données anciennes sont encore probablement trop récentes pour avoir livré de grandes avancées.

I.1.3.2 L'étude d'ADN de pathogènes anciens

L'ADN humain n'est pas le seul type d'ADN qu'il est possible d'extraire à partir de tissus humains anciens. En réalité, l'étape d'extraction d'ADN à partir d'un échantillon biologique ancien permet d'en isoler l'ADN génomique total, c'est-à-dire aussi bien l'ADN endogène de l'organisme mort que l'ADN de tous les micro-organismes ayant pu infecter l'échantillon en question y compris celui de pathogènes. Par la suite, c'est le choix des amorces utilisées lors de la PCR qui déterminera le type d'ADN amplifié. Les études d'ADN de pathogènes anciens font partie d'un domaine de recherche émergent, appelé la paléomicrobiologie, dont la finalité est de détecter, d'identifier et de caractériser l'ADN de micro-organismes, incluant des bactéries, des virus et des parasites, à partir d'échantillons anciens (Drancourt et Raoult, 2005). Là encore il n'y a pas de limite de temps précise, Drancourt et Raoult ont proposé que l'adjectif « ancien » correspondait à des échantillons datant d'avant la mise en évidence des

micro-organismes recherchés. Pour ce faire, les techniques de microscopie et d'immuno-détection sont utilisées mais, à ce jour, la plupart des résultats ont été obtenus grâce aux études d'ADNa.

1.1.3.2.1 Contexte général

Avant le développement des études d'ADNa, le diagnostic de maladies infectieuses ayant affecté les populations anciennes ne pouvait s'effectuer qu'à partir de l'observation visuelle de lésions sur les restes humains. La possibilité d'extraire et d'amplifier de l'ADN de pathogènes anciens a tout d'abord permis de confirmer le diagnostic d'une maladie infectieuse. Les espèces pathogènes étudiées ne sont normalement pas présentes chez des individus sains et la plupart de ces pathogènes n'ont pas de réservoir environnemental connu (Donoghue et Spigelman, 2006). Cette approche moléculaire a déjà été appliquée avec succès aussi bien pour le diagnostic de maladies infectieuses provoquant des lésions observables sur les tissus humains anciens, comme la tuberculose (Salo et al., 1994) ou la lèpre (Haas et al., 2000), et de maladies visuellement indécélables, comme la peste (Drancourt et al., 1998). Les agents pathogènes responsables de grandes épidémies, telles que l'épidémie de peste qui a sévi en France au XVI^{ème} siècle (Drancourt et al., 1998) et la pandémie de grippe « espagnole » de 1918 (Taubenberger et al., 2005), ont ainsi pu être identifiés.

Depuis peu, les études de paléomicrobiologie ne s'attachent plus seulement à diagnostiquer des pathologies et à caractériser les pathogènes à partir de sujets isolés ou de quelques individus. Elles visent désormais à connaître plus précisément l'état sanitaire des populations du passé dans la perspective d'élucider (i) l'histoire évolutive des maladies infectieuses, mais aussi (ii) l'évolution moléculaire des micro-organismes pathogènes et (iii) la relation hôte-pathogène. Ces points sont particulièrement importants puisqu'ils devraient permettre de mieux comprendre les contextes d'émergence de ces maladies infectieuses et de ré-émergence actuelle de certaines de ces maladies. A ce jour, la maladie infectieuse la plus étudiée par cette approche est la tuberculose.

1.1.3.2.2 L'exemple de la tuberculose

La tuberculose est une maladie infectieuse évolutive et contagieuse par voie aérienne. Les bactéries responsables de cette maladie chez les mammifères sont regroupées au sein du complexe *Mycobacterium tuberculosis* (CMTB). La tuberculose humaine est principalement due à *M. tuberculosis*, qui est un pathogène strict de l'homme. *M. bovis*, l'agent principal de la

tuberculose animale en particulier chez les bovins, est également capable de provoquer la maladie chez l'homme *via* le contact avec des animaux malades ou l'ingestion de produits laitiers contaminés. Les autres membres du CMTB comprennent des pathogènes stricts de l'homme (*M. africanum* et *M. canettii*) et des espèces généralement retrouvées chez les animaux (*M. microti*, *M. caprae* et *M. pinnipedii*). Bien qu'ayant un spectre d'hôtes différent, ces espèces présentent 99,9% de similarité de séquence nucléotidique (Sreevatsan et al., 1997; Brosch et al., 2002; Smith et al., 2006). Avant le séquençage complet de plusieurs espèces du CMTB, il a été pensé à tort que le bacille de la tuberculose humaine, *M. tuberculosis*, avait dérivé de *M. bovis* et que ce dernier aurait franchit la barrière d'espèces vers l'homme à l'époque de la domestication du bétail, il y a 10 000 à 15 000 ans (Gibbons, 2008). Ces dernières années, l'analyse d'un grand nombre de souches de *M. tuberculosis* a permis d'individualiser des lignées qui diffèrent en terme de virulence, d'immunogénicité et de distribution géographique (Gagneux et Small, 2007). En réalité, il semble que les souches adaptées à l'homme sont beaucoup plus diverses que celles adaptées aux animaux. Le CMTB serait composé de deux grandes lignées évolutives bien distinctes, l'une avec l'homme comme hôte unique comprenant toutes les lignées *M. tuberculosis*, l'autre touchant diverses espèces de mammifères et plus occasionnellement l'homme et comprenant *M. bovis*. Ces deux lignées présenteraient un ancêtre commun probablement originaire d'Afrique qui aurait été un pathogène de l'homme il y a plus de 40 000 ans voire plusieurs millions d'années. Proche de *M. canettii* ou *M. tuberculosis*, c'est à partir de cet ancêtre que les espèces du CMTB auraient divergé (Brosch et al., 2002; Gutierrez et al., 2005; Hershberg et al., 2008; Wirth et al., 2008).

L'ancienneté de la tuberculose est attestée par l'existence de lésions suggestives d'une telle infection présentes sur de nombreux squelettes et momies datant de plusieurs milliers d'années. Bien que la forme pulmonaire de la tuberculose soit prédominante, la tuberculose peut atteindre n'importe quelle partie du corps, y compris le squelette, suite à une dissémination hématogène des bacilles. Les lésions les plus fréquemment décrites chez les sujets anciens sont localisées au niveau de la colonne vertébrale. Il s'agit de lésions destructives formant des cavités au niveau du corps vertébral d'une à plusieurs vertèbres thoraciques et/ou lombaires, qui peuvent être associées à un effondrement de la ou des vertèbres atteintes voire même de la fusion de ces vertèbres aboutissant parfois à la formation d'une gibbosité. Ces lésions sont évocatrices d'une tuberculose vertébrale ou Mal de Pott, qui est l'atteinte squelettique la plus fréquente et la plus caractéristique (Ortner, 2003). Un exemple de Mal de Pott est représenté sur la figure 3. D'autres sites de lésions moins

fréquents tels que les côtes, les articulations de la hanche et du genou, le crâne ou encore les poumons chez les momies sont aussi décrits dans la littérature. D'ailleurs, le plus ancien cas de tuberculose a été récemment décrit à partir de lésions observées sur des fragments d'un crâne appartenant à un spécimen *d'Homo erectus* datant d'environ 500 000 ans (Kappelman et al., 2008). Cette étude fait l'objet de polémique car les lésions décrites ne sont pas spécifiques de la tuberculose et aucun diagnostic différentiel n'a été émis (Roberts et al., 2009). C'est pourquoi, il est indispensable de compléter le diagnostic paléo-pathologique de la tuberculose au niveau moléculaire pour attester de la présence d'une bactérie du CMTB.



Figure 3 : Mal de Pott, enfant de la nécropole d'Adaïma, Haute-Egypte, 3600 avant J.-C.
(Photo modifiée d'après Crubézy et al., 2006).

Le fait qu'il existe des lésions morphologiques évoquant une tuberculose est l'une des raisons pour lesquelles cette maladie est la plus étudiée dans les analyses d'ADNa. A cela s'ajoute le fait que la composition nucléotidique du génome des mycobactéries, qui présente un taux élevé en bases G et C, ainsi que la composition de leur paroi, riche en acides mycoliques, sembleraient protéger l'ADN des agents pathogènes de la tuberculose contre la dégradation au cours du temps.

A ce jour, les études d'ADNa ont permis de confirmer la présence de la tuberculose humaine en Egypte ancienne (3500-500 avant J.-C.) (Crubezy et al., 1998; Zink et al., 2001; Zink et al., 2003a; Zink et al., 2003b; Donoghue et al., 2009b), en Angleterre à l'époque préhistorique (400-230 avant J.-C.) (Mays et Taylor, 2003; Taylor et al., 2005), en Amérique à des époques pré-colombiennes (Salo et al., 1994; Braun et al., 1998; Konomi et al., 2002) et en Europe à l'époque médiévale (Taylor et al., 1996; Taylor et al., 1999; Mays et al., 2001; Fletcher et al., 2003a; Fletcher et al., 2003b). Les plus anciens cas de tuberculose humaine et animale, confirmés au niveau moléculaire, ont été identifiés, respectivement, à partir d'ossements de deux sujets provenant d'un village néolithique de 9000 ans des côtes d'Israël (Hershkovitz et al., 2008) et d'un bison de 17 000 ans retrouvé en Amérique du Nord (Rothschild et al., 2001). Jusqu'à présent, une seule étude a révélé la présence de tuberculose humaine provoquée par une infection par *M. bovis* chez des sujets sibériens datant de l'Age du Fer (Taylor et al., 2007). Les auteurs expliquent ce résultat par le mode de vie de ces sujets qui était compatible avec une transmission de *M. bovis* à partir d'animaux infectés. La présence d'ADN de *M. tuberculosis* a été détectée dans la plupart des autres études, ce qui est concordant avec le schéma évolutif proposé par les études phylogénétiques précédemment évoquées. La plupart des études portant sur la tuberculose ancienne se sont focalisées sur des restes squelettiques présentant des altérations évocatrices d'une telle infection. Néanmoins, certaines études ont analysé, en complément, des échantillons anciens ne présentant pas de lésions évocatrices de tuberculose voire aucune altération afin d'estimer la fréquence de cette maladie dans les populations anciennes et l'impact qu'elle a pu avoir sur l'histoire de ces populations.

Une revue de la littérature des études d'ADNa concernant la tuberculose (Donoghue, 2009) ainsi qu'une analyse critique de la qualité de ces travaux (Roberts et Ingham, 2008) ont récemment été publiées. Le bilan de ces travaux est assez surprenant puisque sur les 65 études d'ADNa considérées, 90% ne mentionneraient pas l'utilisation de témoins négatifs pour évaluer les risques de contamination et 85% n'ont vraisemblablement pas validé leurs résultats en répliquant l'expérience. Il semblerait donc que les critères d'authentification des études d'ADNa (Cooper et Poinar, 2000) ne soient pas aussi importants pour les études d'ADN de pathogènes que pour celles d'ADN humain (Taylor et al., 2009). Outre le problème de validation des résultats, la spécificité de certaines méthodologies d'analyse (en particulier le spoligotypage) font également l'objet de critiques et de débats au sein des paléomicrobiologistes (Donoghue et al., 2009a; Wilbur et al., 2009). Néanmoins, tous s'accordent à

dire que les études d'ADNa sont nécessaires et qu'elles représentent une source d'information inestimable pour mieux comprendre l'évolution des agents pathogènes et de la maladie (Donoghue et al., 2009a; Smith et al., 2009; Wilbur et al., 2009).

I.2 Les marqueurs moléculaires

I.2.1 Etudes de génomes humains anciens

I.2.1.1 L'ADN humain : rappels fondamentaux

Deux types d'ADN peuvent être individualisés dans l'espèce humaine : l'ADN nucléaire (ADNnu) et l'ADN mitochondrial (ADNmt).

L'ADNnu, localisé dans le noyau des cellules sous forme de chromosomes (22 paires d'autosomes et 1 paire de chromosomes sexuels, XX chez la femme et XY chez l'homme), présente une transmission biparentale sauf pour le chromosome Y. En effet, ce dernier est transmis uniquement du père au fils et il échappe sur pratiquement toute sa longueur à la recombinaison (transmission uniparentale paternelle). Ainsi, tous les individus masculins d'une même descendance ont le même chromosome Y (en l'absence de nouvelle mutation).

L'ADNmt est une molécule d'ADN circulaire de 16 569 paires de bases (pb) localisée dans les mitochondries, organites du cytoplasme cellulaire qui produisent l'énergie nécessaire au fonctionnement des cellules (Pakendorf et Stoneking, 2005). Cet ADN double brin code pour 13 sous-unités de protéines impliquées dans le processus de la phosphorylation oxydative, 22 ARN de transfert et 2 ARN ribosomiaux. Il existe seulement deux régions non codantes dans l'ADNmt : la région de contrôle connue sous le nom de « D-loop » (displacement loop) ou boucle D et la région V. Une carte schématique de l'ADNmt est représentée sur la figure 4. Contrairement à l'ADNnu qui n'est présent qu'en deux exemplaires par cellule (un seul exemplaire par cellule pour le chromosome Y), il existe plusieurs centaines à plusieurs milliers de copies d'ADNmt dans une cellule humaine (Pakendorf et Stoneking, 2005). En effet, l'ADNmt est présent en de multiples exemplaires par mitochondrie et il y a plusieurs mitochondries par cellule. Par ailleurs, l'absence de recombinaison et l'hérédité maternelle de l'ADNmt sont aujourd'hui largement acceptées (Bandelt et al., 2005; Pakendorf et Stoneking, 2005). Cette transmission uniparentale maternelle s'expliquerait par le fait que les mitochondries du spermatozoïde sont regroupées au niveau du flagelle et que ce dernier ne pénètre généralement pas l'ovocyte lors de la fécondation. De plus, il semblerait que les rares

mitochondries paternelles pénétrant dans l'ovocyte soient détruites. Ainsi, tous les enfants ont le même ADNmt que leur mère et l'ADNmt des individus féminins d'une même descendance maternelle est identique.

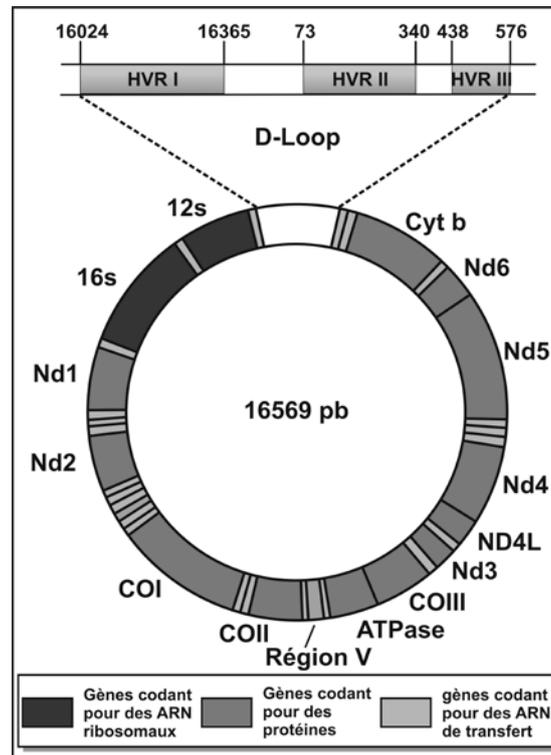


Figure 4 : Carte schématique de l'ADNmt humain.

Quelque soit leur localisation dans la cellule, les molécules d'ADN sont instables et accumulent des variations au cours du temps. Les modifications de l'ADN sont considérées comme des polymorphismes génétiques à partir du moment où elles sont représentées à une fréquence de plus de 1 % dans la population. Lorsqu'une variabilité est observée au niveau de la longueur de la séquence d'ADN étudiée, on parle de polymorphismes de longueur (*e.g.* délétion, insertion ou répétition), et s'il s'agit d'une modification de la séquence de l'ADN elle-même, on parle de polymorphismes de séquence (*e.g.* substitution). A un locus donné, les différentes versions possibles d'un polymorphisme sont les allèles. Le génome nucléaire étant diploïde, l'analyse d'un polymorphisme localisé au niveau de l'ADNnu, et plus précisément au niveau des autosomes, générera un couple d'allèles (l'un hérité du père, l'autre de la mère), qui déterminera le caractère homozygote (allèles identiques) ou hétérozygote (allèles différents) de l'individu pour ce locus. En revanche, l'analyse d'un polymorphisme localisé au niveau de l'ADNmt et du chromosome Y, tous deux haploïdes, ne générera qu'un seul allèle.

I.2.1.2 Les marqueurs moléculaires

I.2.1.2.1 Les polymorphismes de longueur

Les polymorphismes de longueur sont localisés uniquement au niveau de l'ADNnu et généralement dans les régions non codantes. Deux types de marqueurs moléculaires sont généralement utilisés dans les études d'ADNa : le locus de l'amélogénine et les microsatellites ou STR.

I.2.1.2.1.1 Le locus de l'amélogénine

Le gène de l'amélogénine est présent sur les deux chromosomes sexuels, X et Y. Il code pour une protéine de l'émail de la dent. Les deux séquences homologues de ce gène diffèrent par la présence de plusieurs délétions. En 1993, une délétion de 6 pb localisée dans le 1^{er} intron de ce gène, découverte sur le chromosome X, a été utilisée pour la première fois pour la détermination génétique du sexe des individus (Sullivan et al., 1993). Actuellement, cette délétion est toujours utilisée à cet effet dans la majorité des études d'ADNa ainsi que pour les investigations médico-légales. Le principe de cette méthode de sexage repose sur l'amplification par PCR d'une partie du 1^{er} intron du gène de l'amélogénine en utilisant un couple d'amorces s'hybridant en amont et en aval de cette délétion. Un fragment de 106 pb est amplifié à partir du chromosome X et un fragment de 112 pb à partir du chromosome Y. Ce polymorphisme de longueur est parfaitement adapté à l'état de fragmentation de l'ADNa puisque son analyse requiert l'amplification de fragments d'ADN de petite taille. Ce marqueur est généralement inclus dans les kits commerciaux d'identification génétique des individus.

I.2.1.2.1.2 Les STR (Short Tandem Repeats)

Les STR ou microsatellites sont des polymorphismes de répétition. Ils sont constitués de motifs nucléotidiques de 2 à 6 pb (*e.g.* GATA) répétés de 5 à 100 fois les uns à la suite des autres. Ils sont localisés dans les régions non codantes de l'ADNnu au niveau des autosomes et des chromosomes sexuels X et Y.

Le nombre de répétitions des STR autosomaux est très variable d'un individu à l'autre en raison du brassage génétique qu'ils subissent. Leur pouvoir discriminant élevé fait de ces polymorphismes des marqueurs de choix pour l'identification individuelle des individus en médecine légale. Un profil génétique correspond à l'ensemble des génotypes aux loci STR autosomaux d'intérêt et un génotype correspond à la combinaison des deux allèles présents à

un locus donné. Actuellement, l'établissement de profils génétiques repose sur l'étude de plusieurs STR autosomaux grâce à l'utilisation de kits commerciaux tels que les kits d'amplification PCR AmpFISTR® Identifiler® (AB) ou PowerPlex® 16 (Promega). Ces kits permettent l'amplification simultanée de 15 loci STR autosomaux ainsi que du locus de l'amélogénine comme mentionné précédemment. L'avantage de ces kits est qu'ils reposent sur une amplification par PCR multiplexe consistant à utiliser plusieurs couples d'amorces au sein d'une même réaction pour amplifier simultanément plusieurs loci, ce qui permet de travailler à partir de quantités minimales d'ADN. Ces kits présentent un degré de multiplexage élevé puisque les couples d'amorces générant des produits PCR de même gamme de taille sont marqués par des fluorophores différents. Une étape de migration des produits d'amplification par électrophorèse capillaire est ensuite nécessaire pour séparer les différents produits PCR et détecter le nombre de répétition des allèles amplifiés. Dans les études d'ADNa, ces marqueurs sont surtout utilisés pour détecter la présence d'ADN contaminant dans différents extraits d'ADN provenant d'un même sujet. Par ailleurs, de transmission biparentale, ils sont aussi très utiles pour rechercher les liens de proches parentés (parents/enfants).

Les STR localisés dans la région non recombinante du chromosome Y, parfois appelée région mâle-spécifique, sont transmis de manière strictement paternelle. La conséquence directe de l'absence de recombinaison de cette partie du chromosome Y avec le chromosome X est que toute variation génétique dans cette partie va se trouver fixée et donc être transmise sans modification de génération en génération. Un individu masculin possède donc généralement la même combinaison allélique que son frère, son père, son grand-père paternel et tous les autres individus de la même lignée paternelle. En conséquence, ces marqueurs sont utiles pour suivre la transmission des lignées paternelles au cours du temps et avoir accès aux schémas de migrations survenues dans les populations masculines. Ils ne définissent pas directement la lignée paternelle mais l'haplotype Y du sujet. Le terme « haplotype » est équivalent au terme « génotype ». Alors que l'on parle de génotype lorsque deux allèles sont présents à un locus donné, on parle d'haplotype lorsqu'un seul allèle est présent à un locus donné. Ce terme s'applique donc au chromosome Y mais aussi à l'ADNmt, qui sont tous les deux haploïdes. L'haplotype correspond à un ensemble d'allèles transmis en bloc et sans modification à la descendance et présentant donc une histoire évolutive commune. A une échelle plus large, il est possible de regrouper certains haplotypes en haplogroupes, qui sont définis par des polymorphismes spécifiques et correspondent aux lignées. Ainsi les sujets qui partagent les

mêmes allèles pour plusieurs loci STR du chromosome Y font partie du même haplotype Y et donc du même haplogroupe Y ou de la même lignée paternelle. Cependant, l'inverse n'est pas forcément vrai puisqu'un même haplogroupe Y, ou une même lignée paternelle, peut regrouper des haplotypes Y différents. Cela s'explique par le fait que les STR ont un taux de mutation plus élevé que les SNP qui définissent les haplogroupes. Pour caractériser la lignée paternelle d'un sujet, il est nécessaire de compléter l'analyse des STR du chromosome Y par une analyse des SNP du chromosome Y. Des banques de données permettant de calculer les fréquences des haplotypes Y dans diverses populations humaines actuelles sont en accès libre sur internet (*e.g.* <http://www.yhrd.org/>, Y-STR haplotype reference database). Ainsi, il est possible d'étudier la distribution des allèles des STR du chromosome Y dans une population du passé et de la comparer avec celle d'une population de référence. Cela permet d'appréhender directement les changements de fréquences alléliques au cours du temps. L'analyse des STR du chromosome Y est également facilitée par la commercialisation de kits d'amplification tels que les kits d'amplification PCR AmpFISTR[®] Yfiler[®] (AB) et PowerPlex[®] Y (Promega).

Bien que l'analyse des STR soit facile à mettre en œuvre, elle peut souvent se révéler infructueuse à partir d'extraits d'ADN dégradés compte tenu de la fragmentation de l'ADN au cours du temps qui limite fortement la taille des fragments d'ADN amplifiable. Sauf cas exceptionnel, il n'est généralement pas possible d'amplifier des séquences anciennes dont la longueur est supérieure à 300 pb. Or la taille des fragments d'ADN amplifiés avec les kits commerciaux précédemment cités varie généralement de 100 à 400 pb. Une perte de signal est donc souvent observée pour les marqueurs présentant des allèles de grande taille. Néanmoins, ce phénomène peut être utile dans les études d'ADNa puisqu'une corrélation inverse entre l'efficacité d'amplification et la taille des produits PCR est un des critères d'authentification des séquences nucléotidiques anciennes. Cependant, il est beaucoup plus problématique dans le domaine médico-légal où des échantillons dégradés sont aussi souvent rencontrés et qu'une identification ne peut se faire sur la base de profils génétiques partiels. Ces dernières années, les « mini-STR » ont fait l'objet de nouveaux développements. Ce ne sont pas de nouveaux marqueurs en soi mais des marqueurs STR conventionnels pour lesquels la taille des produits d'amplification générés est considérablement réduite. Cela résulte de la conception d'un nouveau couple d'amorces qui s'hybrident le plus près possible du loci STR considéré. Un kit d'amplification simultanée de 8 loci mini-STR a récemment été commercialisé (AmpFISTR[®] MiniFiler[™], AB). Ces récents développements profitent également aux études d'ADNa.

1.2.1.2.2 Les SNP (Single Nucleotide Polymorphisms)

Les SNP ou polymorphismes ponctuels de séquence sont des substitutions dans la majorité des cas. Les transitions, qui correspondent au remplacement d'une purine par une autre purine ($A \leftrightarrow G$) ou d'une pyrimidine par une autre pyrimidine ($C \leftrightarrow T$), sont plus fréquentes que les transversions, qui correspondent au remplacement d'une purine par une pyrimidine ou inversement ($A \leftrightarrow C$, $A \leftrightarrow T$, $G \leftrightarrow C$ ou $G \leftrightarrow T$). Les petites insertions ou délétions (indels) sont souvent comprises dans la dénomination SNP. Par convention, la fréquence de l'allèle le moins représenté doit être au moins égale à 1% ; dans le cas contraire, le terme « mutation » est utilisé. Les SNP constituent les polymorphismes les plus répandus du génome humain et présentent une répartition uniforme (en moyenne toutes les 1000 pb) (Venter et al., 2001; Levy et al., 2007). Ils sont présents sur les régions codantes et non codantes de l'ADN et ce aussi bien sur l'ADNnu que sur l'ADNmt.

1.2.1.2.2.1 Les polymorphismes ponctuels de l'ADNmt

Depuis le début des études d'ADNa, l'ADNmt est le support privilégié de ces analyses génétiques pour deux raisons principales. La première est que l'ADNmt est présent en un plus grand nombre de copies par cellule que l'ADNnu. De ce fait, il semble plus probable de mettre en évidence l'information contenue dans l'ADNmt que celle contenue dans l'ADNnu à partir de tissus biologiques anciens. La seconde raison repose sur les caractéristiques du génome mitochondrial (*i.e.* hérédité strictement maternelle, absence de recombinaison et taux de mutation élevé). Ces particularités font que l'ADNmt évolue exclusivement par l'accumulation au cours du temps de SNP générés par des événements aléatoires et, par conséquent, que les polymorphismes de l'ADNmt sont de bons marqueurs de l'évolution humaine. En outre, ces polymorphismes permettent de différencier des lignées maternelles et de reconstruire les schémas de migration des populations féminines.

Le taux de mutation de l'ADNmt est environ 10 fois plus élevé que celui de l'ADNnu. Cela s'expliquerait par un renouvellement plus rapide des mitochondries et par le fait que les matrices mitochondriales sont des milieux très oxygénés qui favoriseraient la formation de dommages oxydatifs au niveau de l'ADNmt. Ce taux de mutation n'est pas constant quelque soit la région étudiée de l'ADNmt. Il existe en effet 3 régions hypervariables de l'ADNmt (HVR I, HVR II et HVR III) qui contiennent la plupart des polymorphismes de l'ADNmt et sont localisées au niveau de la principale région non codante de l'ADNmt, la région de contrôle ou D-loop (Figure 4).

Dans les études d'ADNa, la région HVR I est la plus étudiée puisqu'elle est la plus grande et présente donc le plus de polymorphismes. Son analyse s'effectue par amplification par PCR puis séquençage classique. La séquence obtenue est alors comparée à une séquence de référence, appelée rCRS (pour revised Cambridge Reference Sequence) (Anderson et al., 1981 ; Andrews et al., 1999), afin de mettre en évidence les polymorphismes qui vont définir un haplotype mitochondrial (parfois appelé mitotype). Les haplotypes mitochondriaux peuvent être uniques ou partagés par différents individus et leur différenciation provient de l'accumulation aléatoire de polymorphismes au niveau de cette région. La comparaison de la fréquence des haplotypes, pour une même population et à différentes époques, permettra de mettre en évidence des modifications qui ont été subies par cette population au cours du temps ou au contraire sa pérennité, alors que la comparaison de la distribution et de la fréquence des haplotypes entre différentes populations permettra de proposer des schémas de dispersion correspondant à différents flux de gènes et par déduction à des mouvements de populations.

Dans certains cas, il est difficile de retrouver des haplotypes partagés entre les populations. Les haplogroupes mitochondriaux auxquels ils sont affiliés sont alors étudiés afin d'avoir une vision plus large des événements. Les individus partageant le même haplotype mitochondrial appartiennent généralement au même haplogroupe mitochondrial ou à la même lignée maternelle. Il existe plus de 25 macro-haplogroupes mitochondriaux et de nombreux sous-haplogroupes dont la répartition géographique actuelle est le reflet des mouvements de populations féminines au cours du temps. Par définition, les haplogroupes mitochondriaux sont des ensembles d'haplotypes qui dérivent d'un haplotype ancestral commun. Par conséquent, les haplotypes affiliés au même haplogroupe présentent des mutations ou SNP en commun à des positions spécifiques de l'ADNmt. Ces loci déterminants peuvent être localisés au niveau de la D-loop et, de ce fait, seul le séquençage de la région HVR I suffit parfois à définir l'haplogroupe du sujet. Cependant, dans la majorité des cas, il s'agit de SNP des régions codantes de l'ADNmt et il est donc nécessaire de les analyser par un autre moyen. Aujourd'hui, la détermination fine de l'haplogroupe d'un individu dans les études de populations actuelles s'effectue de plus en plus par le séquençage complet de l'ADNmt. Pour ce faire, les séquences de fragments d'ADN chevauchants sont assemblées une fois déterminées. Ces séquences sont obtenues par la méthode classique, dite de Sanger, ou par l'utilisation de technologies de séquençage à haut débit, *e.g.* le pyroséquençage à l'aide de la plateforme « Genome Sequencer 20 » ou GS20 (454 Life Sciences/Roche Diagnostics).

Depuis 2008, l'utilisation de ces techniques a permis de reconstituer les génomes mitochondriaux complets de 6 spécimens néandertaliens ainsi que ceux de plusieurs spécimens anciens d'*Homo sapiens* dont celui d'Ötzi (Ermini et al., 2008; Green et al., 2008; Briggs et al., 2009; Ho et Gilbert, (2009)). Cependant, le séquençage complet de l'ADNmt de sujets anciens reste à ce jour exceptionnel. Cela s'explique par le coût élevé des plateformes de séquençage à haut débit mais également par les difficultés techniques du séquençage classique qui sont liées à la faible quantité d'ADN endogène et à sa nature dégradée. La détermination de l'haplogroupe d'un sujet ancien est donc plutôt basée sur le séquençage classique de régions codantes d'intérêt (Endicott et al., 2003; Sampietro et al., 2007) ou sur le typage direct des SNP d'intérêts (Haak et al., 2005; Endicott et al., 2006; Endicott et al., 2009). Néanmoins, toutes ces méthodes nécessitent la réalisation de plusieurs amplifications par PCR et donc l'utilisation d'une certaine quantité d'extrait d'ADNa qui est souvent limitée. C'est pourquoi l'affiliation de l'haplogroupe d'un sujet ancien à un haplogroupe est souvent réalisée par inférence en comparant, par la méthode de « near-matching », l'haplogroupe du sujet à des haplogroupes déjà publiés et pour lesquels les haplogroupes ont déjà été déterminés (Yao et al., 2002).

Malheureusement, l'analyse des polymorphismes de l'ADNmt ne présente pas que des avantages pour les études d'ADNa. Outre la quantité et la qualité de l'ADNa qui sont des facteurs limitants, il existe des phénomènes qui compliquent voire même peuvent fausser les résultats des analyses. Il s'agit par exemple de la présence de mutations réverses ou encore d'hétéroplasmies. Les hétéroplasmies sont définies par la présence de deux populations d'ADNmt, qui sont souvent détectées au niveau de loci appelés « hot spots » ou points chauds de mutation, et correspondent donc à des polymorphismes intra-individuels. Néanmoins, la limitation majeure résulte du fait que l'ADNmt ne représente qu'une infime partie du génome humain (environ 0,0006%) et uniquement l'histoire maternelle des individus (Pakendorf et Stoneking, 2005). C'est pour ces raisons que les « paléogénéticiens » se tournent désormais vers l'analyse des polymorphismes de l'ADNnu.

1.2.1.2.2 Les polymorphismes ponctuels de l'ADNnu

Les SNP localisés dans la région non recombinante du chromosome Y sont transmis de père en fils, de génération en génération, sans modification. Ils présentent un taux de mutation plus faible que celui des STR du chromosome Y. Par conséquent, ils sont plus stables et permettent de suivre des événements plus anciens. Comme mentionné précédemment, ils

définissent les haplogroupes Y ou lignées paternelles. A ce jour, 20 haplogroupes Y majeurs, nommés par des lettres de A à T, et plusieurs sous-haplogroupes sont caractérisés par des loci SNP spécifiques (The Y Chromosome Consortium, 2002; Karafet et al., 2008). Leur distribution géographique actuelle est le reflet des différents mouvements de migration des populations masculines au cours du temps. Certains haplogroupes Y sont continent-, région- voire même population-spécifiques et, de ce fait, sembleraient être des marqueurs utiles pour déterminer l'origine géographique d'un individu (Jobling, 2001; Jobling et Tyler-Smith, 2003). Leur analyse permet également d'étudier la transmission des lignées paternelles au cours du temps et d'avoir accès aux schémas de migration des populations masculines. Le typage direct des SNP d'intérêt permet de déterminer l'haplogroupe Y d'un individu mais le nombre d'études d'ADNa où les haplogroupes Y des sujets étudiés ont été déterminés reste à ce jour restreint. La quantité d'ADN extrait étant un facteur limitant dans la majorité des études d'ADNa, l'affiliation de l'haplotype Y d'un sujet ancien à un haplogroupe Y est souvent réalisée par inférence en le comparant à des haplotypes déjà publiés et pour lesquels les haplogroupes ont déjà été considérés et rares sont les cas où cette affiliation est déterminée directement par typage de SNP (Li et al., 2007; Csanyi et al., 2008).

Les SNP autosomaux sont de loin les polymorphismes ponctuels les moins étudiés dans les études d'ADNa. Cela peut s'expliquer par le fait que la découverte de ces polymorphismes est très récente puisqu'elle coïncide au séquençage complet du génome humain. De ce fait, tous les SNP autosomaux ne sont pas encore caractérisés. Un nombre impressionnant d'études vise aujourd'hui à identifier les SNP impliqués (i) dans la prédisposition à certaines maladies (The Wellcome Trust Case Control Consortium, 2007; McCarthy et al., 2008), (ii) dans les différences inter-individuelles de réponse aux médicaments (Robert et al., 2005; Dai et al., 2008), (iii) dans la structure et la diversité des populations humaines (Auton et al., 2009; Lopez Herraez et al., 2009), (iv) dans les phénotypes pigmentaires des individus (Branicki et al., 2009; Sturm, 2009) ou encore à définir des ensembles de SNP permettant d'identifier les individus (Sanchez et al., 2006; Pakstis et al., 2009) et d'estimer leur origine biogéographique (Halder et al., 2008; Nassir et al., 2009). Un projet international, nommé HapMap, a été lancé fin 2002 (<http://hapmap.ncbi.nlm.nih.gov/>). L'objectif principal de ce projet est d'identifier les SNP et donc les gènes qui jouent un rôle dans la santé et la réponse des individus aux médicaments et aux facteurs environnementaux. Pour ce faire, les séquences génétiques de différents individus originaires d'Afrique, d'Asie et d'Europe, regroupés au sein de groupes de populations, sont comparées et les variations génétiques entre

les différents groupes sont relevées. Les SNP identifiés et les fréquences alléliques de ces SNP au sein des différents groupes de population ont été publiés (The International HapMap Consortium, 2005; The International HapMap Consortium, 2007) et font régulièrement l'objet de mises à jour qui sont en libre accès sur le site internet précédemment cité. Les SNP présentant des différences significatives de fréquences alléliques entre les populations sembleraient être informatifs pour estimer l'origine biogéographique d'un individu. Ils sont appelés AIM (pour Ancestry Informative Markers). Ce sont des marqueurs intéressants pour la génétique des populations. En effet, à l'inverse des SNP du chromosome Y et de l'ADNmt qui reflètent l'histoire évolutive des lignées paternelles et maternelles, les AIM, localisés sur les autosomes, sont le reflet de l'histoire biologique propre à chaque individu puisqu'ils sont hérités de chacun des deux parents (Wolinsky, 2008). De plus, les AIM ainsi que les SNP impliqués dans le phénotype pigmentaire des individus intéressent fortement la criminalistique puisque leur analyse permet de prédire l'origine géographique d'un individu et certains traits physiques (*e.g.* couleur des cheveux, des yeux et de la peau) (Budowle et van Daal, 2008; Kayser et Schneider, 2009). Ces deux types de SNP sont souvent localisés dans des gènes impliqués dans le mécanisme de la mélanogénèse. Bien que plus de 100 gènes aient déjà été découverts chez la souris (Bennett et Lamoreux, 2003), à peine plus de 10 gènes ont été identifiés comme impliqués dans la pigmentation humaine (Sturm, 2009). Le mécanisme de la mélanogénèse chez l'Homme est schématisé sur la figure 5.

Les SNP autosomaux sont des marqueurs récemment utilisés dans les études d'ADNa. Certains phénotypes particuliers incluant la susceptibilité à une maladie auto-immune (Larcombe et al., 2005), la tolérance au lactose (Burger et al., 2007), la couleur des cheveux et de la peau (Lalueza-Fox et al., 2007), l'aptitude au langage (Krause et al., 2007a) ou encore la perception de l'amertume (Lalueza-Fox et al., 2009) ont déjà pu être étudiés chez des spécimens anciens grâce à l'analyse de SNP autosomaux. Les investigations réalisées sur des spécimens néandertaliens ont révélé que certains Néandertaliens avaient la peau claire et les cheveux roux (*via* l'analyse du gène *MCR1* impliqué dans la pigmentation), que certains ne percevaient pas l'amertume comme environ 30% de la population humaine actuelle (*via* l'étude du gène *TAS2R38* impliqué dans la perception gustative) et qu'ils étaient probablement dotés de la faculté d'apprentissage du langage (*via* l'étude du gène *FOXP2* impliqué dans l'acquisition du langage) (Krause et al., 2007a; Lalueza-Fox et al., 2007; Lalueza-Fox et al., 2009). La plupart de ces SNP ont été étudiés par l'amplification par PCR des loci d'intérêts puis le séquençage des amplicons.

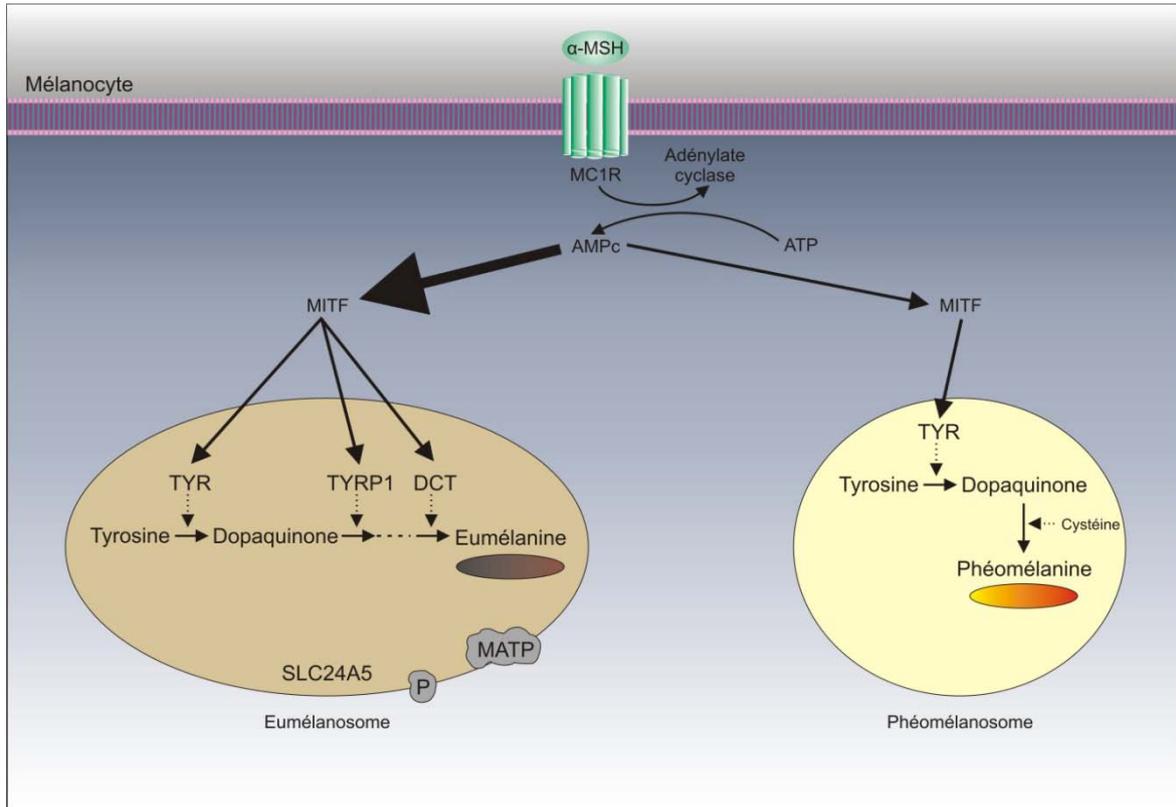


Figure 5 : Interactions entre les principales protéines impliquées dans la mélanogénèse chez l’Homme (d’après Bouakaze et al., 2009).

La pigmentation de la peau, des yeux et des cheveux d’un individu est déterminée par les quantités relatives de deux pigments : l’eumélanine (brune/noire) et la phéomélanine (rouge/jaune). Ils sont synthétisés à partir de la tyrosine dans des organites spécialisés des mélanocytes, les mélanosomes (eumélanosomes et phéomélanosomes). Le gène MC1R, qui code pour le récepteur des mélanocortines de type 1 (MC1R pour MelanoCortin 1 Receptor) ancré dans la membrane du mélanocyte, semble jouer un rôle clé dans la régulation de la mélanogénèse. La synthèse d’eumélanines est stimulée par la fixation de la mélanocortine ou α-MSH (pour α-Melanocyte Stimulating Hormone) sur le récepteur MC1R. Cette liaison va entraîner l’activation de l’adénylate cyclase, avec pour effet la production d’adénosine monophosphate cyclique (AMPc), la synthèse du facteur de transcription MITF (pour Microphthalmia-associated Transcription Factor) et finalement l’activation de l’expression de différentes enzymes de la famille des tyrosinases (*e.g.* TYR pour Tyrosinase, TYRP1 pour Tyrosinase Related Peptide 1, DCT pour DaupaChrome Tautomerase) qui vont conduire à la synthèse de ces pigments bruns/noirs. D’autres protéines sont nécessaires pour la synthèse de ces pigments et la maturation des eumélanosomes telles que la Protéine P (codée par le gène OCA2), SLC24A5 (codée par le gène du même nom) et MATP (codée par le gène SLC45A2). A l’inverse, l’interaction du récepteur MC1R avec la protéine Agouti, antagoniste de l’α-MSH, conduit à une synthèse préférentielle de phéomélanines, de même qu’une concentration élevée de cystéine dans

le mélanosome. De nombreux détails sur la pigmentation humaine sont disponibles dans les références suivantes (Parra, 2007; Sturm, 2009).

I.2.2 Etudes de génomes de pathogènes anciens : l'exemple de la tuberculose

Si les marqueurs moléculaires utilisés dans les études de tuberculoses anciennes sont souvent les mêmes que ceux utilisés en routine dans les laboratoires de bactériologie clinique, les méthodes d'analyse de ces marqueurs diffèrent souvent.

I.2.2.1 Diagnostic de la tuberculose

Le diagnostic de la tuberculose repose sur la détection de micro-organismes du CMTB dans les échantillons d'intérêt. Ainsi, les méthodes moléculaires utilisées pour le diagnostic de la tuberculose sont toutes basées sur l'amplification par PCR d'un fragment d'ADN dont la séquence est conservée parmi les différentes espèces du CMTB et spécifique à ces espèces. Si des tests diagnostiques cliniques sont disponibles dans le commerce (*e.g.* Cobas Amplicor *Mycobacterium tuberculosis*, Roche), ils ne peuvent pas être appliqués à des échantillons anciens en raison de la taille des fragments amplifiés qui n'est pas compatible avec l'état de fragmentation de l'ADNa.

Dans la majorité des études d'ADNa, le diagnostic de la tuberculose repose sur l'amplification par PCR puis le séquençage d'une région de 123 pb ou de 92 pb de la séquence d'insertion *IS6110* (Eisenach et al., 1990; Taylor et al., 1996). Il s'agit d'une séquence d'insertion spécifique des espèces du CMTB et dont la distribution (*i.e.* le nombre de répétitions et les localisations sur le génome) varie d'une souche à l'autre. La séquence d'insertion *IS1081*, qui est aussi présente en de multiples copies dans le génome des espèces du CMTB, est un locus parfois ciblé en complément de *IS6110* (Taylor et al., 2005; Taylor et al., 2007).

I.2.2.2 Identification des espèces du CMTB

Les études d'ADNa récentes ne se contentent plus du diagnostic de la maladie mais visent également à identifier l'espèce du CMTB impliquée. Cette étape d'identification permet de distinguer une tuberculose strictement humaine d'une zoonose (celle provoquée par une infection par *M. bovis*), ce qui est important d'un point de vue épidémiologique (Djelouadji et al., 2008). D'un point de vue clinique, cette identification est également importante pour le

traitement du patient puisque *M. bovis* est naturellement résistant à un des antituberculeux utilisé, la pyrazinamide (Scorpio et Zhang, 1996).

La différenciation des espèces du CMTB repose sur l'analyse de loci polymorphes entre les espèces du CMTB. Il peut s'agir de délétions (Brosch et al., 2002; Parsons et al., 2002) ou de SNP (Niemann et al., 2000; Chimara et al., 2004; Arnold et al., 2005; Huard et al., 2006). Des kits commerciaux d'identification des espèces du CMTB sont disponibles (*e.g.* GenoType[®] MTBC, Hain Lifescience) mais ils n'ont pas été appliqués à des échantillons anciens probablement pour la même raison que celle citée pour les kits diagnostiques de la tuberculose.

Dans les études d'ADNa, seules les espèces *M. tuberculosis* et *M. bovis* sont généralement distinguées puisqu'il s'agit des agents principaux de la tuberculose chez l'Homme. Cette étape est réalisée grâce à l'analyse de SNP spécifiques de l'une des deux espèces (*e.g.* la présence de l'allèle G au locus *oxyR*(285) est caractéristique de *M. tuberculosis* tandis que l'allèle A est retrouvé chez *M. bovis* (Sreevatsan et al., 1996)) ou de délétions (*e.g.* la délétion RD7 n'est pas présente chez *M. tuberculosis* (Brosch et al., 2002)). Ces polymorphismes sont analysés grâce à l'amplification par PCR des régions d'ADN d'intérêt puis, dans le cas des SNP, les amplicons sont séquencés alors que, dans le cas des délétions, une électrophorèse sur gel d'agarose est réalisée afin de vérifier l'absence ou la présence de ces délétions.

I.2.2.3 Typage moléculaire

En bactériologie clinique, le typage moléculaire de *M. tuberculosis* est couramment réalisé en complément du diagnostic moléculaire à des fins épidémiologiques. Il permet de définir des familles génotypiques qui présentent des différences de prévalence, virulence, contagiosité et tendance à la pharmaco-résistance. Il est particulièrement utile pour (i) l'identification de chaînes de transmission dans la communauté ou à l'hôpital, (ii) l'identification de facteurs de risque associés à la transmission récente de tuberculose, (iii) la différenciation des cas de réinfections et de rechutes et (iv) l'identification de familles émergentes de bacilles (Groupe de travail du conseil supérieur d'hygiène publique de France, 2004). De nombreuses méthodes de typage différant par le type de polymorphisme ciblé ont été développées (Barnes et Cave, 2003; Mathema et al., 2006). Trois de ces méthodes, qui sont la technique IS6110-RFLP, le typage MIRU-VNTR et le spoligotypage ont été standardisées et sont aujourd'hui les plus couramment utilisées.

La méthode IS6110-RFLP utilise la séquence IS6110 comme cible. Cette méthode est basée sur la technique RFLP (pour Restriction Fragment Length Polymorphism) qui consiste en la digestion de l'ADN par une enzyme de restriction, la séparation des fragments d'ADN sur gel d'agarose, le transfert sur membrane et enfin la révélation par hybridation à une sonde constituée d'un fragment de la séquence IS6110. Un profil de bandes de différentes tailles est alors observé en fonction du nombre de copies et des positions de la séquence IS6110 dans le génome de la souche étudiée (van Embden et al., 1993).

Le typage MIRU-VNTR est le plus récemment décrit. Il utilise des séquences répétitives dispersées (motifs répétés de 40 à 100 pb) dans le génome des espèces du CMTB, appelées MIRU-VNTR (pour Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats) ou minisatellites par analogie aux génomes eucaryotes. Le nombre de répétitions de ces séquences répétées en tandem est variable d'une souche à l'autre. Après l'amplification par PCR des loci MIRU-VNTR d'intérêt à l'aide d'amorces couplées à un fluorophore puis la séparation des produits PCR par électrophorèse capillaire, un électrophorégramme ou profil de pics de différentes tailles est observé en fonction du nombre de répétitions présent à chaque locus amplifié pour la souche analysée (Supply et al., 2006).

Le spoligotypage (ou spoligotyping pour « spacer oligonucleotide genotyping ») ou « typage par oligonucléotides des espaceurs du locus DR » utilise le locus DR (pour Direct Repeat) comme cible. Ce locus se compose de multiples copies de séquences répétées et identiques séparées les unes des autres par des séquences uniques nommées espaceurs. Ces espaceurs varient d'une souche à l'autre par leur longueur, leur nombre et leur séquence. Ainsi, après l'amplification par PCR des espaceurs à l'aide d'un seul couple d'amorces complémentaires des séquences répétées puis l'hybridation des produits PCR sur une membrane où ont été fixés préalablement les 43 espaceurs les plus fréquents, une empreinte d'hybridation est observée qui reflète la présence ou l'absence des espaceurs au niveau de la région DR de la souche analysée (Kamerbeek et al., 1997).

Le spoligotypage est actuellement la seule des méthodes de typage qui a été appliquée à de l'ADNa (Taylor et al., 1999; Mays et al., 2001; Rothschild et al., 2001; Fletcher et al., 2003b; Zink et al., 2003b; Zink et al., 2007; Hershkovitz et al., 2008). En effet, les méthodes de typage IS6110-RFLP et MIRU-VNTR ne sont pas utilisables sur des substrats anciens en raison de la nécessité d'une quantité initiale de matériel génétique importante, qui impose généralement une étape de culture, et de la taille des fragments d'ADN générés par ces analyses qui n'est pas compatible avec l'état de dégradation de l'ADNa (Zink et al., 2002).

Etant basé sur l'analyse de fragments dont la taille est inférieure à 60 pb, le spoligotypage est plus adapté pour l'analyse d'ADN dégradé (Donoghue et al., 2004). Néanmoins, son utilisation ne fait pas l'unanimité au sein de la communauté des paléomicrobiologistes (Donoghue et al., 2009a; Wilbur et al., 2009). Cette méthode ne serait pas reproductible. En effet, deux études récentes ont démontré que les spoligotypes obtenus à partir d'un même échantillon n'étaient pas forcément identiques (Zink et al., 2003b; Hershkovitz et al., 2008). Il s'agit là probablement d'un problème méthodologique résultant de l'amplification non homogène des différents espaceurs qui dépend fortement de la qualité de l'ADN. A cela s'ajoute le fait que cette technique n'est pas adaptée pour les études phylogénétiques, ce qui la rend donc, selon certains auteurs, très peu utile pour les études d'ADNa (Wilbur et al., 2009).

I.2.2.4 Analyses phylogénétiques

Les méthodes de typage MIRU-VNTR et de spoligotypage ont également été utilisées pour des analyses phylogénétiques réalisées sur des isolats contemporains (Sola et al., 2003; Brudey et al., 2006; Wirth et al., 2008). Néanmoins, des études récentes ont démontré que bien que ces méthodes de typage soient très utiles pour les études épidémiologiques, elles le sont beaucoup moins pour les études phylogénétiques (Comas et al., 2009; Reed et al., 2009). Il en ressort que d'autres polymorphismes plus stables, les LSP (pour Large Sequence Polymorphisms) ou RD (pour Regions of Difference) qui sont des délétions ou encore les SNP sont de bien meilleurs marqueurs phylogénétiques que les MIRU-VNTR ou le locus DR. La plupart des études phylogénétiques réalisées sur des échantillons contemporains publiées à ce jour se sont d'ailleurs basées sur l'analyse de SNP (Sreevatsan et al., 1997; Baker et al., 2004; Filliol et al., 2006; Gutacker et al., 2006; Hershberg et al., 2008) ou de LSP (Brosch et al., 2002; Mostowy et al., 2002). Par ailleurs, des ensembles d'une dizaine de SNP ont été proposés pour différencier les espèces du CMTB, et particulièrement les isolats de *M. tuberculosis*, en 6 groupes phylogénétiques distincts ou lignées, appelés SCG (pour SNP Cluster Groups) (Filliol et al., 2006; Alland et al., 2007). Six lignées de *M. tuberculosis* peuvent également être caractérisées par l'étude de certains LSP et les lignées définies par ces deux types de marqueurs sont concordantes (Gagneux et al., 2006; Alland et al., 2007). Ces lignées présentent la particularité d'être associées à des populations humaines et à des régions géographiques différentes (Hirsh et al., 2004; Gagneux et al., 2006; Reed et al., 2009). Il existe également une corrélation entre ces lignées phylogénétiques et les groupes génotypiques définis par les méthodes de typage (Gagneux et Small, 2007).

L'utilisation de marqueurs informatifs d'un point de vue phylogénétique est beaucoup moins fréquente dans les études d'ADNa. Cependant, elle va probablement devenir incontournable dans les années à venir puisque beaucoup d'espairs sont placés dans ces études pour mieux comprendre l'évolution du CMTB (Smith et al., 2009; Wilbur et al., 2009). Dans les études publiées à ce jour, les marqueurs phylogénétiques utilisés sont deux SNP, *gyrA*⁹⁵ et *katG*⁴⁶³, qui permettent de définir 3 groupes phylogénétiques, appelés PGG (pour Principal Genetic Groups) (Sreevatsan et al., 1997; Fletcher et al., 2003a; Fletcher et al., 2003b; Mays et Taylor, 2003). Le principe de la détermination de ces groupes génétiques au moyen de ces deux SNP est présenté sur la figure 6.

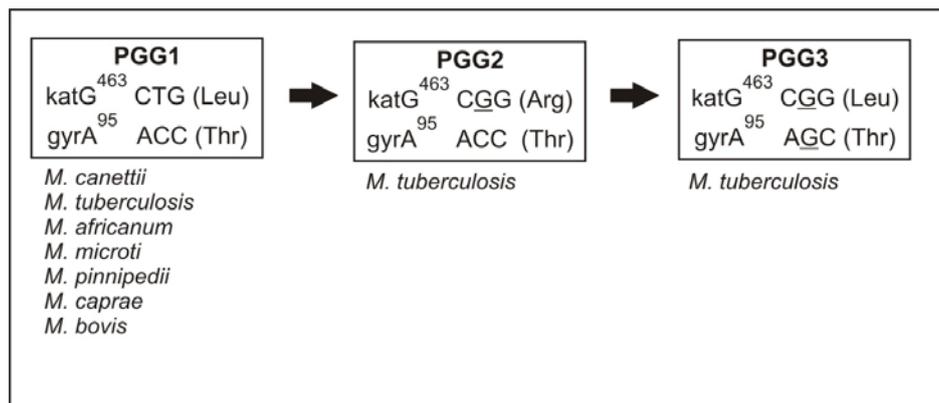


Figure 6 : Détermination des groupes phylogénétiques PGG grâce aux SNP *gyrA*⁹⁵ et *katG*⁴⁶³ (d'après Sreevatsan et al., 1997).

Ces SNP ont été analysés par l'amplification par PCR des régions d'ADN d'intérêt puis au séquençage des amplicons. La délétion *TbD1*, qui permet de différencier un groupe de *M. tuberculosis* dit « ancestral » de celui dit « moderne » (Brosch et al., 2002), a également déjà été étudiée à partir d'échantillons anciens (Brosch et al., 2002; Taylor et al., 2005; Taylor et al., 2007; Hershkovitz et al., 2008). La présence ou l'absence de cette délétion est vérifiée grâce à l'amplification par PCR de la région d'ADN d'intérêt puis à l'électrophorèse sur gel d'agarose des amplicons.

II. Problématique et approche adoptée

Si les rappels fondamentaux font état des difficultés auxquelles se heurtent les études d'ADNa, de nombreux progrès techniques ont été réalisés au cours des dernières années. Par exemple, l'utilisation d'une étape d'amplification par PCR des fragments d'ADN contenant les marqueurs d'intérêt est aujourd'hui incontournable puisqu'elle permet de s'affranchir de la faible quantité des molécules d'ADNa. En revanche, la nature dégradée et l'état fragmenté des molécules d'ADNa restent des facteurs qui limitent le type d'information pouvant être obtenu à partir de leurs analyses. La plupart des études d'ADNa publiées à ce jour ont porté sur l'analyse de polymorphismes de l'ADNmt en raison du plus grand nombre de copies de ce génome dans la cellule. Cependant, l'hérédité exclusivement maternelle de l'ADNmt fait que ces analyses sont uniquement informatives pour déterminer la lignée maternelle d'un sujet ancien. Ces observations nous ont amenés à nous intéresser à des marqueurs moléculaires qui peuvent être analysés à partir de molécules d'ADN fragmentées et qui sont localisés aussi bien sur l'ADNmt que sur l'ADNnu. Il s'agit des polymorphismes ponctuels de séquence ou SNP.

L'objectif principal de ce travail de thèse a été de déterminer ce que pouvait apporter l'analyse des polymorphismes ponctuels à l'étude des populations humaines anciennes, et plus particulièrement à celle de populations sibériennes.

II.1 Pourquoi les SNP ?

Peu d'études d'ADNa ont fait appel aux SNP à ce jour, pourtant ces polymorphismes semblent être des cibles de choix pour les études d'ADNa en raison des caractéristiques suivantes :

- Les SNP consistent en des variations de la séquence d'ADN portant sur un seul nucléotide. Compte tenu de la taille réduite de leur région informative, ces polymorphismes peuvent être analysés à partir de fragments d'ADN de petite taille (50 à 150 pb), ce qui est particulièrement adapté à l'état de dégradation de l'ADNa. Ainsi, leur étude permettrait d'augmenter les chances de succès de l'analyse de substrats très dégradés.
- Les SNP présentent un très faible taux de mutation qui les rend stables au cours du temps (de l'ordre de 10^{-8} - 10^{-9} mutations par base et par génération pour les SNP contre 10^{-3} pour les

STR chez l'Homme) (Brinkmann et al., 1998; Wang et al., 1998). Compte tenu de leur faible taux de mutation, l'étude des SNP permet de retracer des événements anciens.

- Les SNP sont les polymorphismes les plus abondants du génome humain et du génome de micro-organismes pathogènes comme ceux responsables de la tuberculose. Ils sont neutres ou phénotypiquement informatifs.

- Il existe des bases de données des SNP qui sont en libre accès sur internet. Par exemple, dbSNP qui est hébergée par le site du National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/snp>). Ces bases de données regroupent différentes informations très utiles (e.g. pour un SNP donné : la position chromosomique, les génotypes possibles, les fréquences alléliques dans différentes populations ou encore la séquence d'ADN encadrant le SNP d'intérêt).

- La majorité des SNP sont des polymorphismes bi-alléliques puisqu'ils correspondent généralement à des événements mutationnels uniques et donc non récurrents en raison de leur faible taux de mutation. Leur analyse et l'interprétation des résultats sont donc relativement aisées puisqu'il n'y a que deux allèles possibles pour un locus donné.

- De nombreuses méthodologies de typage des SNP ont été développées au cours de ces dernières années et la plupart de ces méthodes permettent de typer simultanément plusieurs SNP (Sobrino et al., 2005; Ragoussis, 2009).

II.2 Les méthodes de typage des SNP

Il n'existe pas de méthode universelle de typage des SNP puisque chacune possède des avantages et des inconvénients qui sont dépendants du type d'application, en particulier du nombre d'échantillons à analyser et du nombre de SNP à typer.

La plupart de ces méthodes comprend une étape préalable qui consiste en l'amplification par PCR du ou des fragments d'ADN contenant le ou les SNP d'intérêt. Cette réaction d'amplification est d'autant plus importante lorsque la matrice de départ est de l'ADNa en raison de la très faible quantité d'ADN endogène (de l'ordre du pg). Deux étapes successives se distinguent ensuite quelque soit la méthode de typage des SNP utilisée : (i) la discrimination allélique, qui permet de générer des produits spécifiques à chaque allèle et (ii) la détection des allèles.

Les différentes méthodes de typage décrites dans la littérature sont basées sur des réactions chimiques différentes pour l'étape de discrimination allélique (hybridation de sondes spécifiques, clivage ou liaison enzymatique, utilisation d'enzymes de restriction ou encore réaction d'extension d'amorces) et sur diverses technologies et plateformes de détection (électrophorèse capillaire, puces à ADN, transfert d'énergie par résonance entre molécules fluorescentes, luminescence ou encore spectrométrie de masse) (Sobrino et al., 2005; Ragoussis, 2009).

Deux méthodes ont retenu notre attention dans ce travail : l'une basée sur une détection par spectrométrie de masse MALDI-TOF (pour Matrix Assisted Laser Desorption/Ionization-Time Of Flight), l'autre par électrophorèse capillaire.

II.2.1 Le typage des SNP par spectrométrie de masse MALDI-TOF

II.2.1.1 Les intérêts

La spectrométrie de masse MALDI-TOF est la seule méthode de détection directe d'une propriété intrinsèque des produits à analyser, qui est leur rapport masse/charge. En effet, contrairement aux autres méthodes, elle ne nécessite pas l'utilisation de fluorochromes ou de tout autre marquage qui peut induire des variations d'une analyse à l'autre et augmente le coût de celles-ci.

Au vu des données de la littérature, la détection d'analytes par spectrométrie de masse MALDI-TOF semble être une méthode sensible, rapide, à haut-débit et qui présente une capacité de multiplexage des SNP élevée en raison de la large gamme de masses pouvant être détectée avec précision (Bray et al., 2001; Paracchini et al., 2002; Mengel-Jorgensen et al., 2004; Tost et Gut, 2005).

Cette méthode a déjà fait l'objet de développements au sein du laboratoire où a été réalisé ce travail de thèse. En effet, dans le cadre du travail de thèse d'Elizabet Petkovski des SNP autosomaux susceptibles d'être utilisés pour l'identification génétique des individus ont été typés par spectrométrie de masse MALDI-TOF (Petkovski et al., 2005; Petkovski, 2006). De plus, des SNP du chromosome Y avaient également pu être analysés à partir d'échantillons anciens grâce à cette méthode (Petkovski et al., 2006).

II.2.1.2 Le principe

Plusieurs technologies de typage de SNP par spectrométrie de masse MALDI-TOF, qui sont basées sur des réactions de discrimination allélique différentes, sont décrites dans la littérature (Tost et Gut, 2002). Les technologies les plus couramment utilisées reposent sur une réaction d'extension d'amorce allèle spécifique ou PEX (pour Primer EXtension). Le principe d'un tel typage est présenté sur la figure 7.

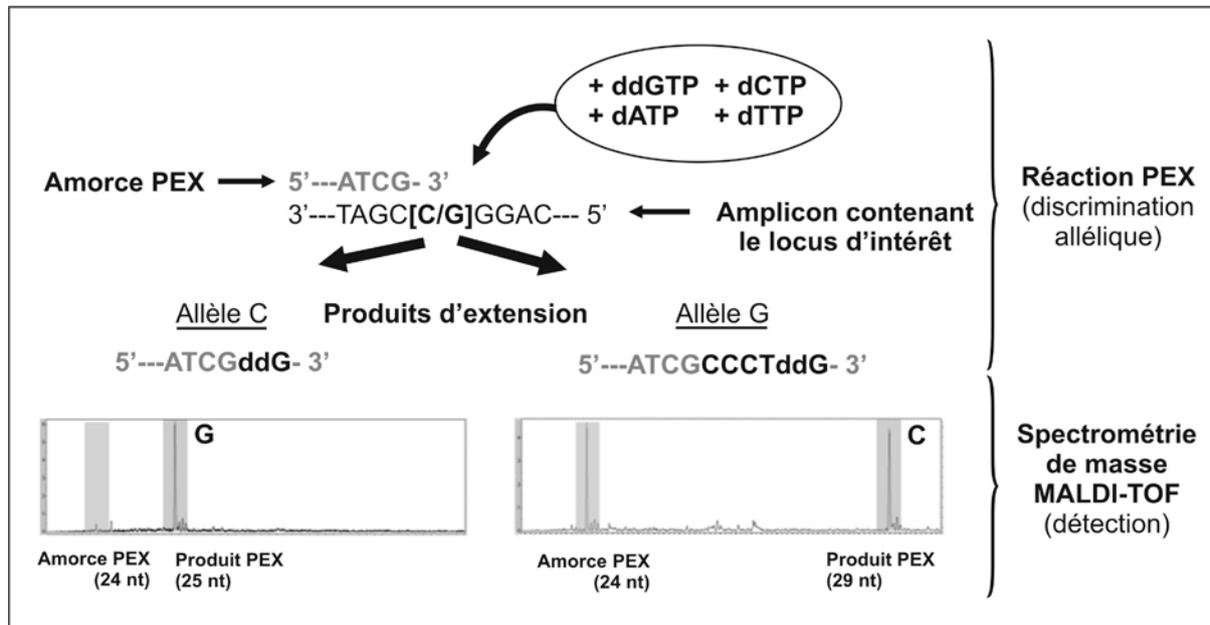


Figure 7 : Principe de l'analyse d'un SNP par spectrométrie de masse MALDI-TOF.

La réaction PEX nécessite au préalable la conception d'une amorce PEX qui s'hybride spécifiquement à la séquence en amont du SNP d'intérêt de manière à ce que le nucléotide à l'extrémité 3' de cette amorce soit complémentaire du nucléotide localisé directement en amont du SNP ciblé. Lors de l'étape d'extension, une ou plusieurs bases peuvent être ajoutées à l'extrémité 3' de cette amorce PEX. L'étape d'extension est réalisée en présence d'une ADN polymérase spécifique qui incorpore indifféremment des désoxynucléotides triphosphates (dNTP) et des didésoxy nucléotides triphosphates (ddNTP) présents dans le milieu réactionnel. Ces dNTP et ddNTP sont choisis en fonction de la séquence du fragment d'ADN d'intérêt. Etant donné que les ddNTP ne possèdent pas le groupement 3'-OH qui est nécessaire pour former la liaison 3'-5' phosphodiester entre deux nucléotides successifs, la réaction d'extension de l'amorce PEX se termine après l'incorporation d'un ddNTP. Dans l'exemple de la figure 7, l'amorce PEX (24 nt) va pouvoir s'hybrider en amont du SNP d'intérêt (substitution C/G). L'un des 4 dNTP (ici le dGTP) du mélange réactionnel est

remplacé par un ddNTP (le ddGTP) qui va bloquer l'élongation lorsqu'il sera incorporé par l'ADN polymérase. Dans le cas de l'allèle C, seul le ddGTP sera incorporé générant un produit d'extension de 25 nt. Alors que dans le cas de l'allèle G, 4 dNTP seront ajoutés avant l'incorporation du ddGTP générant un produit d'extension de 29 nt. Pour un SNP donné, les produits d'extension possibles présentent des tailles et donc des masses différentes selon l'allèle présent au locus donné. Après purification, ces produits pourront donc être séparés par spectrométrie de masse MALDI-TOF et, selon la masse des produits détectés, il sera possible de déduire les allèles présents au locus d'intérêt.

Le spectromètre de masse MALDI-TOF permet de séparer les molécules constitutives d'un échantillon à analyser (analytes) selon leur rapport masse/charge (m/z). Son principe est schématisé sur la figure 8.

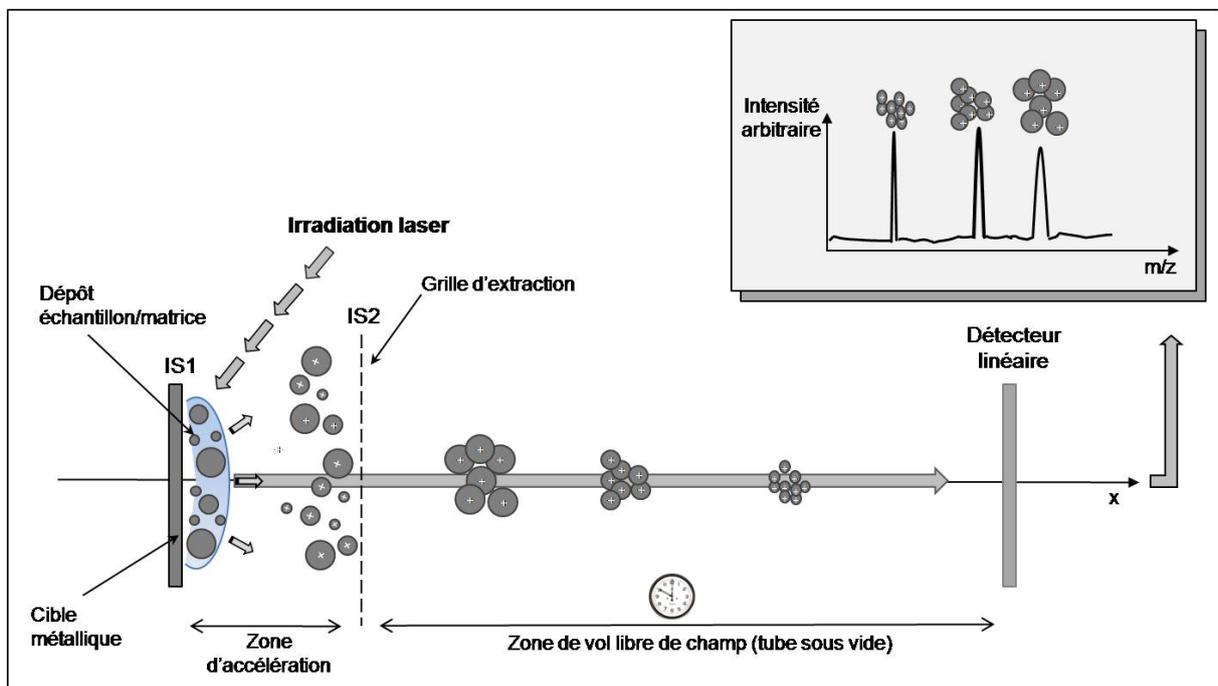


Figure 8 : Principe d'un spectromètre de masse MALDI-TOF.

La technique d'ionisation MALDI repose sur l'irradiation par un faisceau laser d'un mélange de matrice et d'échantillon co-cristallisés sur une surface métallique appelée cible. La matrice absorbant à la longueur d'onde émise par le laser est excitée et transfère l'énergie nécessaire pour ioniser les analytes. Outre l'ionisation, elle assure également une fonction protectrice contre la dégradation des analytes. Parallèlement à cette ionisation dite douce, l'irradiation du mélange cristallin donne lieu à la désorption des ions caractéristiques des analytes. Les ions formés, généralement monochargés, sont focalisés puis accélérés grâce à l'application d'une

différence de potentiel entre la cible (IS1) et la grille d'extraction (IS2) avant d'entrer dans le tube de vol qui est une zone libre de champ. Ils se déplacent ensuite sous vide dans le tube de vol en fonction de la vitesse acquise dans la zone d'accélération avant d'atteindre le détecteur placé à l'autre extrémité du tube du vol. Les ions les plus légers (de rapport m/z plus petit) voleront plus vite et arriveront les premiers au détecteur. Le temps de vol étant proportionnel au rapport m/z et les ions étant monochargés, il est donc possible d'en déduire la masse des ions. Le résultat est présenté sous la forme d'un spectre de masse.

Afin de limiter la quantité d'ADN de départ, plusieurs SNP peuvent être typés simultanément grâce à l'utilisation d'une série de deux réactions multiplexes (PCR et PEX). Dans un premier temps, les fragments d'ADN comprenant les loci d'intérêts sont amplifiés au sein d'une réaction de PCR multiplexe puis les amplicons sont soumis, après purification, à une réaction PEX multiplexe. Il suffit de s'arranger lors de la conception des amorces PEX pour que les différents produits d'extension présentent des masses différentes afin de pouvoir différencier aisément les produits d'extension des différents SNP analysés sur les spectres de masse.

Depuis le travail de thèse d'Elizabet Petkovski, des améliorations techniques quant au typage des SNP par spectrométrie de masse MALDI-TOF ont été décrites dans la littérature, ce qui nous a conduit à tenter de nouveaux développements. En particulier, l'introduction d'une molécule photoclivable dans l'amorce PEX permet de réduire la taille des produits d'extension qui sont alors détectés dans une gamme de masse comprise entre 1000 et 3000 Da au lieu de 4000 à 9000 Da sans photoclivage (Wenzel et al., 2003). Cette méthode, appelée GenoSNIP, doit permettre d'augmenter la sensibilité et la résolution de la détection par spectrométrie de masse MALDI-TOF.

II.2.2 Le typage des SNP par miniséquençage à l'aide du kit SNaPshot™

II.2.2.1 Les intérêts

Notre attention s'est également portée sur la méthode de miniséquençage à l'aide du kit SNaPshot™ (AB) puisqu'elle nécessite l'utilisation d'un séquenceur automatique, appareil qui est généralement présent dans tous les laboratoires de biologie moléculaire et que nous possédons au sein du laboratoire.

De plus, la fiabilité, la sensibilité, la robustesse, et la capacité élevée de multiplexage de cette méthode de typage de SNP ont été décrites dans de nombreuses études (Gutacker et al., 2002; Brandstatter et al., 2003; Sanchez et al., 2003; Quintans et al., 2004; Vallone et al., 2004;

Alessandrini et al., 2005; Brion et al., 2005a; Brion et al., 2005b; Brandstatter et al., 2006; Grignani et al., 2006; Gutacker et al., 2006; Onofri et al., 2006; Sanchez et al., 2006; Kristensen et al., 2007; Alvarez-Iglesias et al., 2008).

Finalement, son efficacité à partir d'échantillons d'ADN dégradés a également déjà été démontrée (Larcombe et al., 2005; Endicott et al., 2006; Burger et al., 2007; Mosquera-Miguel et al., 2009) et un protocole détaillant les différentes étapes nécessaires pour le développement de réactions multiplexes à partir de molécules d'ADN dégradées a récemment été publié (Sanchez et Endicott, 2006).

II.2.2.2 Le principe

Le typage de SNP par miniséquencage au moyen du kit SNaPshot™ (AB) repose sur une étape de discrimination allélique correspondant à une réaction d'extension d'amorce d'une seule base. Cette réaction est appelée SBE (pour Single Base Extension) ou réaction de miniséquencage et son principe est schématisé sur la figure 9.

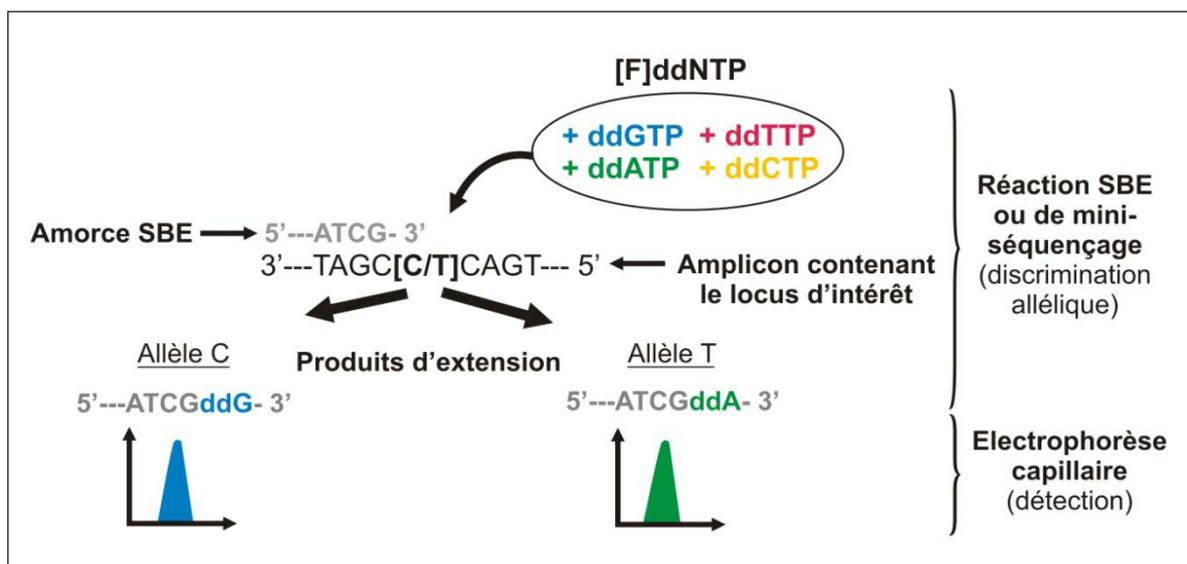


Figure 9 : Principe de l'analyse d'un SNP par miniséquencage à l'aide du kit SNaPshot™.

La réaction de miniséquencage, tout comme la réaction PEX, nécessite la conception d'une amorce SBE qui s'hybride spécifiquement en amont du SNP d'intérêt. A la différence d'un mélange réactionnel de PCR classique, la réaction de miniséquencage a lieu en présence de ddNTP marqués par un fluorophore ([F]ddNTP) différent au lieu des dNTP (*e.g.* mélange réactionnel du kit SNaPshot™ (AB): [dR6G]ddATP, [dTAMRA™]ddCTP, [dR110]ddGTP et [dROX™]ddTTP). Lors de l'étape d'extension, l'ADN polymérase va incorporer un seul

[F]ddNTP à l'extrémité 3' de l'amorce SBE, celui qui est complémentaire au nucléotide présent au locus d'intérêt. Pour un locus SNP donné, les produits d'extension générés présentent la même taille mais peuvent différer de part le [F]ddNTP incorporé qui dépend de l'allèle présent au locus d'intérêt. Ces produits sont ensuite soumis à une électrophorèse capillaire où ils vont migrer selon leur taille et la fluorescence émise sera détectée. Pour un locus donné, chaque allèle sera représenté sur l'électrophorégramme par un pic de couleur différente mais de même taille.

Il est également possible de multiplexer les réactions PCR et SBE afin d'analyser plusieurs SNP simultanément (Figure 10). Le multiplexage est rendu possible grâce à l'ajout d'un oligonucléotide (*e.g.* une séquence poly(C)) à l'extrémité 5' des amorce SBE permettant de faire varier la taille de chaque amorce SBE et donc de différencier aisément les produits d'extension de chaque SNP sur l'ectrophorégramme généré après leur migration par électrophorèse capillaire.

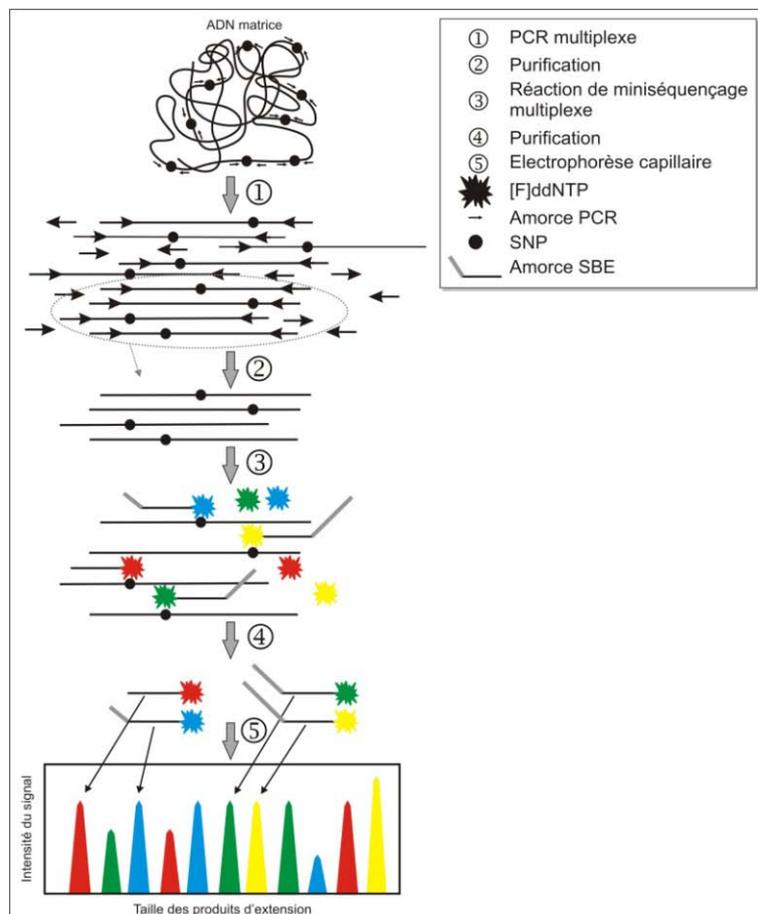


Figure 10 : Les différentes étapes d'une série de réactions multiplexes pour l'analyse simultanée de plusieurs SNP par miniséquénage à l'aide du kit SNaPshot™ (d'après Sanchez et al., 2006).

II.3 Les populations sibériennes étudiées

L'histoire des populations sibériennes est l'une des moins bien connue au monde. Pourtant, elle est l'une des plus intéressantes de part la situation géographique de la Sibérie à la jonction des mondes européens, asiatiques et américains. En effet, la Sibérie s'étend d'Ouest en Est des montagnes de l'Oural à l'Océan Pacifique et entre l'Océan Arctique au Nord et les frontières avec le Kazakhstan la Mongolie et la Chine au Sud (Figure 11).



Figure 11 : Carte géographique de la Sibérie.

La localisation des échantillons anciens ayant fait l'objet des analyses génétiques présentées dans ce mémoire est indiquée par des numéros (1 : échantillons anciens de la région de Krasnoyarsk, 2 : échantillons yakoutes anciens).

La Sibérie est généralement divisée en 3 grandes régions (d'Ouest en Est) : la Sibérie Occidentale dont les plaines s'étendent des montagnes de l'Oural au fleuve Ienisseï, la Sibérie Centrale dont le plateau couvre la zone allant du Ienisseï à la Lena et la Sibérie Orientale qui correspond pour sa quasi-totalité à la Yakoutie (plus grande République autonome de la Fédération de Russie). Les nombreuses populations autochtones de Sibérie, qui possèdent leurs propres langues et leurs modes de vie traditionnels, évoluent dans un contexte

écologique très particulier notamment marqué par un climat continental rigoureux avec des hivers longs et froids. Cet élément constitue un avantage pour les études d'ADNa puisqu'il permet une excellente conservation des sépultures, des corps et de l'ADN.

Des sujets issus de deux populations sibériennes anciennes ont fait l'objet d'investigations dans ce travail : les uns provenant de la région de Krasnoyarsk (Sud de la Sibérie Centrale) et les autres provenant de Yakoutie Centrale (Sibérie Orientale) (Figure 11).

Les sujets anciens provenant de la région de Krasnoyarsk ont été exhumés à partir de tombes surmontées de tumuli funéraires, caractéristiques de la culture dite « kourgane » (tumulus en russe). Cette culture étant considérée par certains comme liée à l'origine des Indo-Européens, il est donc intéressant d'étudier l'origine de ces sujets anciens. Lors du commencement de ce travail de thèse, des prélèvements appartenant à une trentaine de ces sujets venaient de faire l'objet d'investigations génétiques au moyen de marqueurs conventionnels des études d'ADNa (*i.e.* région HVR I de l'ADNmt, STR autosomaux et STR du chromosome Y). Cependant, ces analyses se sont révélées insuffisantes pour retracer l'origine de ces spécimens du Sud Sibérien. Etant donné que certains SNP semblent être des marqueurs intéressants pour une telle application, l'une des thématiques de ce travail a été d'étudier des SNP sur les échantillons anciens du Sud Sibérien afin de tenter d'apporter de nouveaux éléments sur leur origine.

L'étude de l'histoire des populations yakoutes présente également un intérêt puisque leur mode de vie est particulier. En effet, les Yakoutes pratiquent l'élevage du bétail et des chevaux alors que leurs ethnies voisines sont des chasseurs-cueilleurs et des éleveurs de rennes. A cela s'ajoute le fait qu'ils ne parlent pas la même langue. Plus de soixante tombes yakoutes gelées ont été mises à jour lors des Missions Archéologiques Françaises en Sibérie Orientale (MAFSO) dirigées par le Professeur Eric Crubézy (Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse, FRE2960, CNRS, Toulouse). Alors que l'éthnogénèse yakoute a été étudiée dans le travail de thèse de Sylvain Amory (Amory, 2007), l'état sanitaire de ces sujets anciens a fait l'objet d'investigations dans ce travail de thèse. En effet, le nombre de sujets, mis au jour lors de trois dernières campagnes de fouilles de la MAFSO, présentant des lésions osseuses évocatrices d'une atteinte tuberculeuse soulève l'éventualité d'une épidémie de tuberculose qui aurait pu sévir en Yakoutie Centrale pendant plus d'une centaine d'années. Etant donné qu'il est possible d'analyser des SNP pour identifier les agents

pathogènes de la tuberculose, une autre thématique de ce travail a été d'étudier des SNP sur ces échantillons anciens afin d'étudier l'état sanitaire des sujets yakoutes anciens.

II.4 La stratégie méthodologique adoptée

Pour aborder les deux thématiques de recherche précédemment énoncées, il a fallu en premier lieu définir, au vu des données de la littérature, des SNP susceptibles de répondre à notre questionnement. Pour étudier l'origine des sujets anciens provenant de la région de Krasnoyarsk, notre attention s'est portée sur les différents types de SNP du génome humain : les SNP du chromosome Y qui permettent de caractériser les haplogroupes Y, les SNP de la région codante de l'ADNmt qui permettent de compléter l'étude de la région HVR I de l'ADNmt et donc de caractériser les haplogroupes mitochondriaux et les SNP autosomaux qui permettent d'estimer l'origine biogéographique et certains phénotypes pigmentaires. Pour étudier l'état sanitaire des sujets yakoutes anciens, notre attention s'est portée sur des SNP permettant de différencier les agents responsables de la tuberculose au niveau de l'espèce et sur des SNP phylogénétiquement informatifs.

Par la suite, il a également fallu choisir une méthode de typage des SNP. Nous avons testé en parallèle les deux méthodes qui avaient retenu notre attention et qui ont été précédemment décrites : la méthode GenoSNIP basée sur une détection par spectrométrie de masse MALDI-TOF et la méthode de typage par miniséquençage à l'aide du kit SNaPshotTM (AB) basée sur une détection par électrophorèse capillaire. Les développements ont été réalisés sur des échantillons d'ADN contemporains.

La méthode qui nous a paru être la plus adaptée a ensuite été optimisée puis appliquée aux différents échantillons anciens.

III. Matériel et Méthodes

III.1 Echantillons biologiques

Etant donné que les échantillons anciens représentent une source très limitée et non renouvelable d'ADN endogène, des échantillons biologiques contemporains ont été utilisés pour les étapes de développement et d'optimisation des méthodes de typage des SNP.

III.1.1 Echantillons contemporains

Des prélèvements de cellules buccales sur cyto-brosses ou sanguines sur papier buvard réalisés sur 27 individus Européens (France), 3 Africains (Togo) et 6 Asiatiques (Yakoutie et Mongolie) ont été utilisés pour la mise au point des réactions de typage des SNP du génome humain. Conformément à la législation française, le consentement éclairé de ces 36 individus a été recueilli par écrit préalablement à la réalisation des prélèvements. Il a également été demandé à ces personnes de définir la couleur de leurs yeux (bleu, vert, gris ou brun), celle de leurs cheveux (roux/auburn, blond, brun clair ou foncé ou noir) et leur type de peau (claire, mate ou foncée).

Soixante-et-onze échantillons bactériens ont été utilisés pour le développement des réactions de typage des SNP du génome des bactéries du CMTB. Ils ont été fournis par le Professeur Yves Piémont^(†) de l'Institut de Bactériologie de Strasbourg, le Professeur Henri Dabernat du Laboratoire Bactériologie-Hygiène de Toulouse et le Docteur Nicolas Veziris du Centre National de Référence des Mycobactéries et de la Résistance des Mycobactéries aux Antituberculeux de Paris.

Cinquante-six échantillons ont été préparés à partir de cultures bactériennes des différentes espèces du CMTB incluant 3 souches de références (*M. tuberculosis* H37Rv ATCC 27294, *M. bovis* CIP 102426 et *M. bovis* BCG CIP 105226) et 53 isolats cliniques de *M. tuberculosis* (n=35), *M. bovis* (n=6), *M. bovis* BCG (n=4), *M. africanum* (n=4), *M. canettii* (n=1), *M. caprae* (n=1), *M. microti* (n=1) et *M. pinnipedii* (n=1).

Quinze échantillons provenant de cultures de bactéries du genre *Mycobacterium* mais n'appartenant pas au CMTB (MOTT, *Mycobacteria other than tuberculosis*) comprenant *M. fortuitum* (n=2), *M. kansasii* (n=2), *M. abscessus* (n=1), *M. avium* (n=3), *M. chelonae* (n=2) et *M. gordonae* (n=1) et de 4 bactéries n'appartenant pas au genre *Mycobacterium* (*Nocardia nova*, *Corynebacterium amycolatum*, *Staphylococcus aureus* et *Escherichia coli*) ont été utilisés comme témoins négatifs.

Les échantillons bactériens correspondaient à des extraits d'ADN purifiés et quantifiés pour les 3 souches de référence du CMTB et à des extraits cellulaires obtenus après thermolyse pour toutes les autres espèces bactériennes.

III.1.2 Echantillons anciens

Les échantillons anciens provenant de la région de Krasnoyarsk (Sud de la Sibérie Centrale) analysés dans ce travail ont été prélevés sur 25 sujets issus de tombes appartenant à différentes cultures de l'Age du Bronze Moyen à l'Age du Fer (II^{ème} millénaire avant J.-C. - IV^{ème} siècle après J.-C.) et correspondent à des fragments d'os longs (Tableau 2).

Echantillon	Culture	Période	Sexe	Haplogroupe mitochondrial supposé
S07	Andronovo	1800-1400 av. J.-C.	M	U4
S08	Andronovo	1800-1400 av. J.-C.	F	Z
S09	Andronovo	1800-1400 av. J.-C.	F	T1
S10	Andronovo	1800-1400 av. J.-C.	M	U2e
S11	Andronovo	1800-1400 av. J.-C.	F	T4
S13	Andronovo	1800-1400 av. J.-C.	F	H
S14	Andronovo	1800-1400 av. J.-C.	F	U4
S15	Andronovo	1800-1400 av. J.-C.	F	K2b
S16	Andronovo	1800-1400 av. J.-C.	M	U5a1
S18	Karasuk	1400-800 av. J.-C.	F	U5a1
S19	Karasuk	1400-800 av. J.-C.	F	U4*
S21	Tagar	800 av. J.-C.-100 ap. J.-C.	F	T3
S22	Tagar	800 av. J.-C.-100 ap. J.-C.	F	T3
S23	Tagar	800 av. J.-C.-100 ap. J.-C.	F	T3
S24	Tagar	800 av. J.-C.-100 ap. J.-C.	M	I
S25	Tagar	800 av. J.-C.-100 ap. J.-C.	M	G2a
S26	Tagar	800 av. J.-C.-100 ap. J.-C.	M	C
S28	Tagar	800 av. J.-C.-100 ap. J.-C.	M	F1b
S29	Tagar	800 av. J.-C.-100 ap. J.-C.	M	H
S32	Tagar	800 av. J.-C.-100 ap. J.-C.	M	H
S33	Tashtyk	800 av. J.-C.-100 ap. J.-C.	F	C
S34	Tashtyk	800 av. J.-C.-100 ap. J.-C.	M	HV
S35	Tashtyk	800 av. J.-C.-100 ap. J.-C.	F	H
S36	Tashtyk	800 av. J.-C.-100 ap. J.-C.	F	N9A
S37	Tashtyk	100-400 ap. J.-C.	F	T1

Tableau 2 : Liste des échantillons anciens provenant de la région de Krasnoyarsk.

Le profil génétique (STR autosomaux), l'haplotype mitochondrial (polymorphismes de la région HVR I de l'ADNmt) et l'haplotype Y (STR du chromosome Y) de chacun de ces sujets avaient déjà été obtenus lors d'analyses génétiques antérieures réalisées au sein du laboratoire

d'anthropologie moléculaire. Un haplogroupe mitochondrial avait déjà pu être déduit à partir de l'haplotype grâce à la méthode du « near-matching » (Tableau 2). Ces données génétiques ont été publiées (cf. Publications, Keyser et al., 2009).

Les échantillons anciens provenant de Yakoutie (Sibérie Orientale) analysés dans ce travail ont été prélevés sur des sujets issus de 7 tombes individuelles mises au jour lors de différentes fouilles de la MAFSO et correspondent à des vertèbres (Tableau 3). Il s'agit de vertèbres lombaires ou thoraciques isolées ou fusionnées présentant des lésions osseuses lytiques au niveau du corps vertébral (Figure 12). Une tuberculose a été évoquée par les anthropologues chez ces sujets suite à l'observation de ces lésions. De plus, d'autres lésions de type infectieux notamment au niveau des côtes (abcès froid tuberculeux, Kous Tcharbyt), de la hanche (coxite, Bouogaryma 2) ou du poumon (pleurésie, Kyys Ounouoga) ont été observées chez plusieurs de ces sujets. Néanmoins, seules les vertèbres ont fait l'objet d'analyses génétiques dans ce travail de thèse.

MAFSO	Tombe	Datation	Echantillons analysés
2005	Batta Tcharana	XVII ^{ème}	3 vertèbres
2005	Kous Tcharbyt	XVIII ^{ème}	2 vertèbres
2005	Bouogaryma 2	XVIII ^{ème}	1 vertèbre
2005	Okhtobout 2	XVIII ^{ème}	1 vertèbre
2005	Kyys Ounouoga	XVIII ^{ème}	1 vertèbre
2006	Odjuluun 2	XVIII ^{ème}	1 vertèbre
2007	Atakh	XVIII ^{ème}	4 vertèbres soudées

Tableau 3 : Liste des échantillons anciens provenant de Yakoutie.

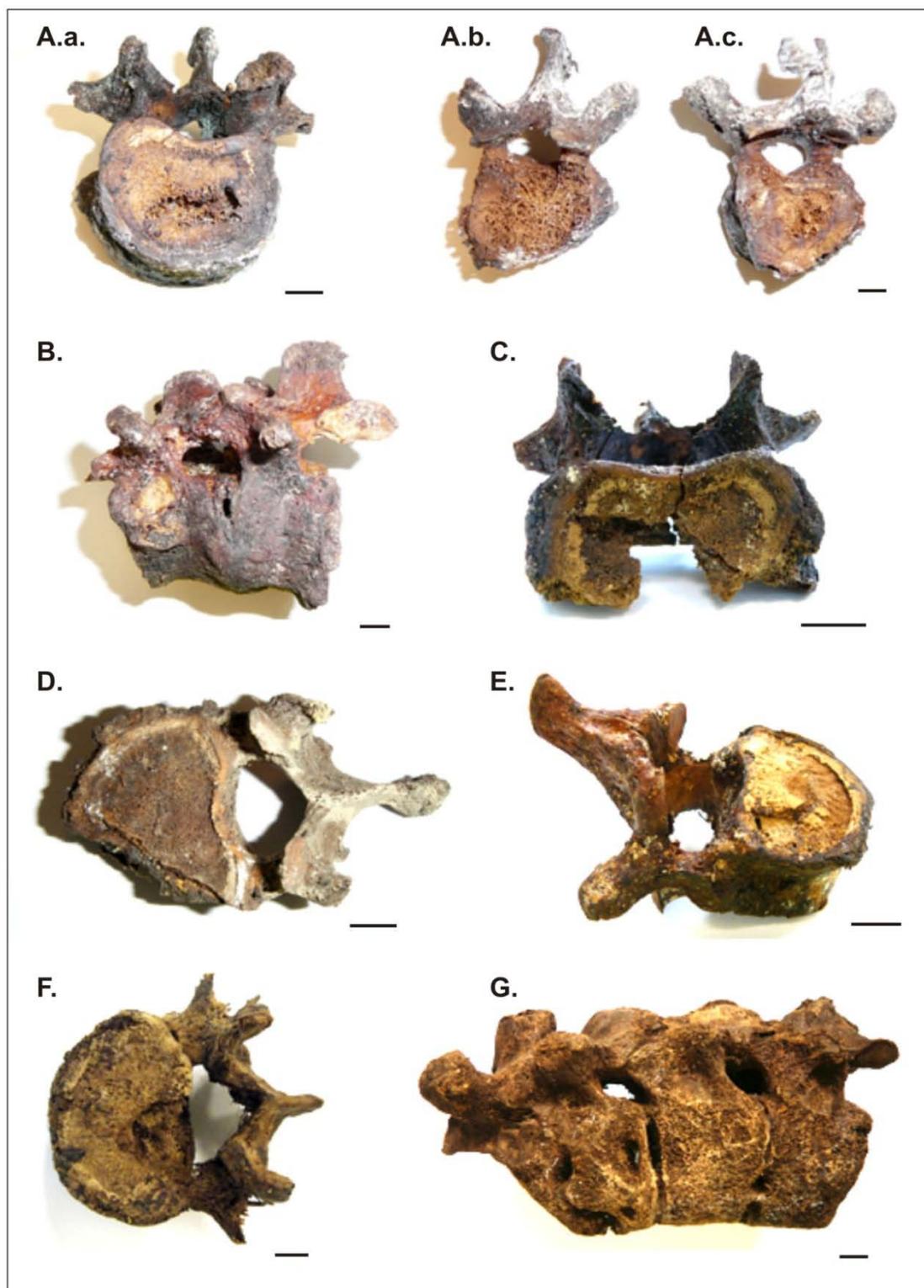


Figure 12 : Photos des échantillons yakoutes anciens.

Les vertèbres présentées sur cette figure ont fait l'objet d'analyses génétiques visant à détecter la présence d'agents pathogènes de la tuberculose. Elles proviennent de sujets yakoutes anciens exhumés à partir des tombes Batta Tcharana (A.), Kous Tcharbyt (B.), Bouogaryma 2 (C.), Okhtobout 2 (D.), Kyys Oounougha (E.), Odjuluun 2 (F.) et Atakh (G.) mises au jour lors de différentes campagnes de fouilles de la MAFSO.

III.2 Marqueurs moléculaires

III.2.1 Génome humain

III.2.1.1 SNP du chromosome Y

Des SNP du chromosome Y ont sélectionnés sur la base de données de la littérature de telle façon à permettre la détermination des haplogroupes Y des échantillons humains anciens étudiés au laboratoire (*i.e.* échantillons provenant de Sibérie mais également de Mongolie et d'Amérique du Sud). Les 13 SNP sélectionnés à cet effet sont présentés sur la figure 13. Ils définissent des haplogroupes Y fréquemment retrouvés dans les populations humaines actuelles sibériennes, mongoliennes et amérindiennes (Underhill et al., 2001; Karafet et al., 2002; Lell et al., 2002; Zegura et al., 2004; Derenko et al., 2006).

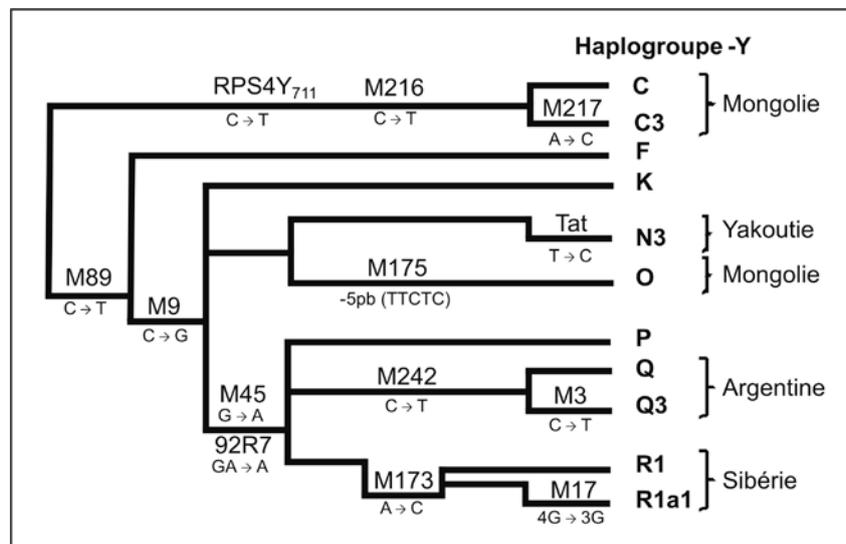


Figure 13 : Arbre phylogénétique présentant les SNP du chromosome Y sélectionnés.

Les haplogroupes définis par ces SNP ainsi que les régions géographiques où ils sont fréquemment retrouvés sont indiqués à l'extrémité de chaque branche. La nomenclature utilisée pour les SNP et les haplogroupes est celle définie par le Consortium du Chromosome Y (The Y Chromosome Consortium, 2002). Les marqueurs RPS4Y₇₁₁ et Tat sont aussi connus, respectivement, sous les noms M130 et M46.

III.2.1.2 SNP de la région codante de l'ADNmt

Des SNP de la région codante de l'ADNmt ont été choisis à partir de données de la littérature, de la base de données MITOMAP (www.mitomap.org/) et du site internet PhyloTree.org

(<http://www.phylotree.org/tree/main.htm>) afin de confirmer et/ou d'affiner les haplogroupes mitochondriaux des échantillons anciens provenant de la région de Krasnoyarsk (Tableau 2). Les 27 SNP sélectionnés à cet effet sont présentés sur la figure 14. Ils caractérisent des haplogroupes ou des sous-haplogroupes d'intérêt (Derbeneva et al., 2002; Kivisild et al., 2002; Derenko et al., 2003; Kong et al., 2003; Achilli et al., 2004; Loogvali et al., 2004; Starikovskaya et al., 2005; Brandstatter et al., 2006; Grignani et al., 2006; Mooder et al., 2006; Pakendorf et al., 2006).

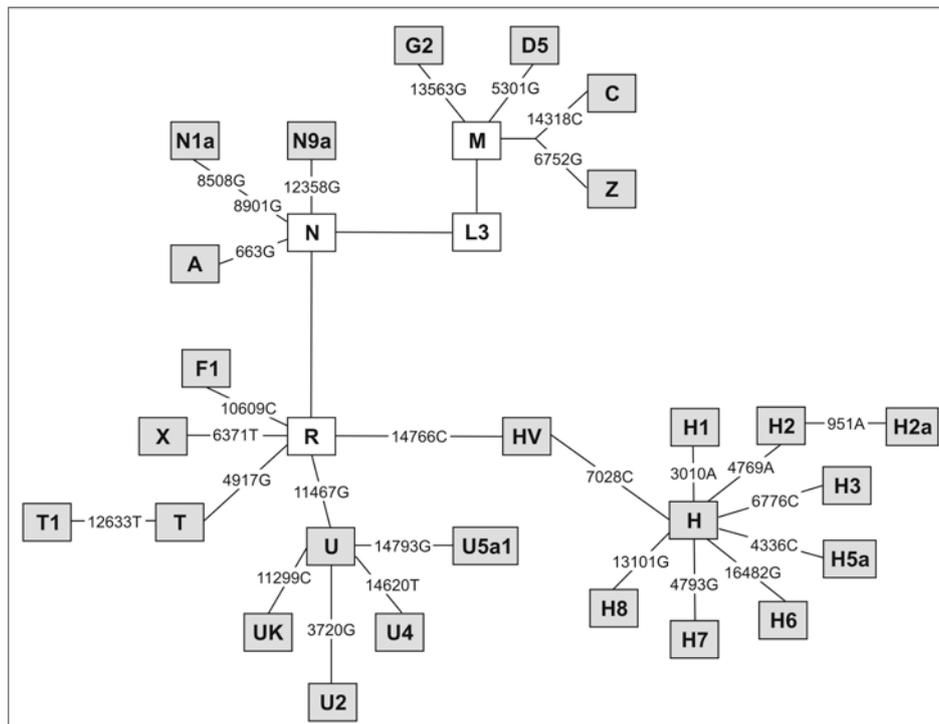


Figure 14 : Arbre phylogénétique présentant les SNP de la région codante de l'ADNmt sélectionnés. La nomenclature utilisée pour désigner les SNP correspond à leur position sur le génome mitochondrial. Pour chaque locus, seul l'allèle muté est représenté, l'allèle ancestral est celui présent dans la séquence de référence rCRS. Les haplogroupes mitochondriaux qu'ils caractérisent sont indiqués dans les encadrés.

III.2.1.3 SNP autosomaux

Des SNP autosomaux ont été sélectionnés à partir de données de la littérature afin d'apporter des informations sur l'apparence physique et l'origine biogéographique des échantillons anciens provenant de la région de Krasnoyarsk. Des informations relatives aux 10 SNP autosomaux sélectionnés à cet effet sont présentées dans le tableau 4. Six de ces SNP ont été associés à des phénotypes pigmentaires particuliers (couleur des cheveux, des yeux et de la peau) (Branicki et al., 2007; Duffy et al., 2007; Eiberg et al., 2008; Sturm et al., 2008). Quatre

de ces SNP présentent des différences significatives de fréquences alléliques entre les populations européennes et asiatiques (Graf et al., 2005; Lamason et al., 2005; Myles et al., 2007; Soejima et Koda, 2007; Yuasa et al., 2007). Ces SNP sont localisés au niveau de 6 gènes impliqués dans la pigmentation humaine (Figure 5).

Gène	N° accession sur dbSNP (rs#)	Allèles	Variation protéique	Phénotype associé
<i>MC1R</i>	rs1805007	C/T	R151C	Couleur des cheveux (roux)
	rs1805008	C/T	R160W	
<i>HERC2</i>	rs12913832	A/G	-	Couleur des yeux (bleue)
<i>OCA2</i>	rs7495174	T/C	-	Couleur des yeux, des cheveux et type de peau
	rs6497268 ou rs4778241	G/T	-	
	rs11855019 ou rs4778138	T/C	-	
	rs1545397	G/A	-	
<i>SLC45A2</i>	rs16891982	C/G	F374L	Origine européenne
<i>SLC24A5</i>	rs1426654	G/A	A111T	Origine européenne
<i>DCT</i>	rs2031526	G/A	-	Origine asiatique

Tableau 4 : Informations concernant les SNP autosomaux sélectionnés.

La nomenclature utilisée pour désigner ces SNP correspond à leur numéro d'accèsion dans la base de données dbSNP du NCBI (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

III.2.2 Génome des bactéries du CMTB

III.2.2.1 SNP

Des SNP du génome des bactéries du CMTB ont également été choisis afin de pouvoir identifier les bactéries du CMTB au niveau de l'espèce et caractériser des groupes phylogénétiques. Les 16 SNP qui ont été sélectionnés à cet effet ont été répartis en deux groupes de 8 SNP chacun.

Les 8 SNP qui composent le premier groupe sont présentés sur la figure 15. Ils permettent de différencier les 7 espèces du CMTB et les 3 groupes phylogénétiques PGG (pour Principal Genetic Groups) décrits par Sreevatsan et al (Sreevatsan et al., 1997; Frothingham et al., 1999; Kasai et al., 2000; Niemann et al., 2000; Goh et al., 2001; Huard et al., 2006).

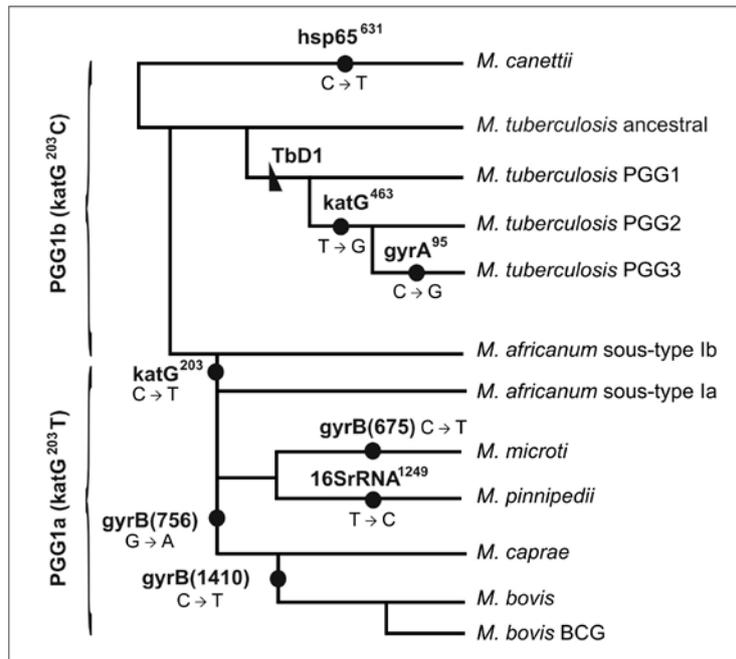


Figure 15 : Arbre phylogénétique présentant les 8 SNP sélectionnés pour la différenciation des espèces du CMTB.

La délétion TbD1 a été représentée pour différencier les espèces de *M. tuberculosis* dites anciennes et modernes (PGG2 et PGG3) mais elle n’a pas été étudiée dans ce travail.

Les 8 SNP qui composent le second groupe sont présentés dans le tableau 5. Ils permettent de caractériser les 7 groupes phylogénétiques ou SCG (pour SNP cluster group) décrits par Filliol et al (Filliol et al., 2006; Alland et al., 2007).

SCG	SNP ^a								
	1977	3352929	74092	105139	2460626	232574	311613	913274	2154724 ^b
1	G	G	C	C	C	G	T	<u>G</u>	<u>A</u>
2	G	G	C	<u>A</u>	C	G	T	C	<u>A</u>
3a	G	G	C	C	C	G	T	C	<u>A</u>
3b	G	G	C	C	C	G	T	C	C
3c	G	G	C	C	C	<u>T</u>	T	C	C
4	G	G	C	C	<u>A</u>	<u>T</u>	T	C	C
5	G	C	C	C	C	G	T	C	C
6a	<u>A</u>	C	C	C	C	G	T	C	C
6b	<u>A</u>	C	C	C	C	G	<u>G</u>	C	C
7	<u>G</u>	G	<u>T</u>	C	C	G	T	<u>G</u>	<u>A</u>

Tableau 5 : Détermination des SCG grâce aux SNP sélectionnés.

^aLa nomenclature utilisée pour désigner les SNP correspond à leurs positions sur le génome de *M. tuberculosis* H37Rv (séquence déposée sous le numéro d’accès NC_000962 sur la base de

données GenBank du NCBI). Les allèles mutés apparaissent en gras et soulignés. ^bCe SNP correspond au SNP katG⁴⁶³ qui fait également partie du premier groupe de SNP sélectionné.

III.2.2.2 Séquence d'insertion IS6110

La séquence d'insertion IS6110 a été utilisée pour vérifier la spécificité de l'analyse des SNP à partir des échantillons yakoutes anciens. En effet, le fait de détecter un fragment de 123 pb de cette séquence permet de confirmer la présence d'un agent pathogène des espèces du CMTB dans l'extrait d'ADN analysé (Eisenach et al., 1990).

III.3 Extraction et quantification de l'ADN

III.3.1 Extraction, purification et concentration de l'ADN

Etant donné que les échantillons bactériens contemporains fournis étaient des extraits d'ADN ou des extraits cellulaires, seuls les prélèvements humains contemporains et anciens ont fait l'objet d'une extraction d'ADN au sein de notre laboratoire.

III.3.1.1 Echantillons humains contemporains

L'embout des cytobrosses ou des morceaux de papier buvard supportant une tâche de sang ont été placés dans un tube contenant 400 µL de tampon de lyse (EDTA 0,5 mM, SDS 2%, Tris HCl 10 mM, acétate de sodium 0,3 M) et 10 µL de Protéinase K (20 mg/ml) puis ce mélange a été incubé sous agitation à 50°C pendant 2h. Une extraction organique classique au phénol/chloroforme/alcool isoamylique (25/24/1 ; v/v/v) a ensuite été réalisée à partir du lysat cellulaire. La phase aqueuse a été récupérée après centrifugation, puis l'ADN a été purifié sur les colonnes de silice du kit CleanMix (Talent) selon les recommandations du fabricant. Ce kit permet de purifier l'ADN grâce à son affinité pour la silice en présence de thiocyanate de guanidium. L'ADN a finalement été élué dans 400 µL d'eau ultra-pure (milli-Q, Millipore).

III.3.1.2 Echantillons humains anciens

Afin d'éliminer tous les contaminants de surface, les échantillons osseux anciens (*i.e.* fragments d'os longs de la région de Krasnoyarsk et vertèbres yakoutes) ont d'abord été décontaminés par un traitement mécanique de surface. La terre ou les fragments de tissu séché qui étaient encore présents sur les vertèbres ont été délicatement enlevés à l'aide d'une lame

de scalpel stérile. La surface des échantillons osseux a ensuite été abrasée sur environ 2-3 mm à l'aide d'une perceuse de type Dremel[®] (Breda) équipée d'un disque abrasif sous une hotte préalablement irradiée aux UV et dédiée à cet usage.

Les fragments d'os longs ont été trépanés à l'aide d'une perceuse à colonne équipée d'un trépan chirurgical dans le but d'obtenir de la poudre d'os. Une vitesse de rotation réduite a été utilisée afin de limiter tout échauffement qui pourrait endommager les molécules d'ADN. Concernant les vertèbres, seuls les corps vertébraux ont été réduits en poudre en utilisant un cryobroyeur Freezer Mill 6800 (Spex). Cette méthode a été choisie car les corps vertébraux étaient de trop petite taille et trop poreux pour être trépanés. Chaque échantillon a été placé avec un aimant cylindrique dans un tube de polycarbonate fermé par deux bouchons métalliques. L'ensemble a été immergé dans l'azote liquide et soumis à un champ magnétique qui a déplacé l'aimant d'un bout à l'autre du tube réduisant le corps vertébral en poudre. L'azote liquide permet d'éviter les échauffements susceptibles de dégrader davantage les molécules d'ADN.

Tous les échantillons anciens ont fait l'objet d'une extraction d'ADN au moyen d'un protocole publié en 2005 (Keyser-Tracqui et Ludes, 2005). Il repose sur une extraction organique au phénol-chloroforme à partir de 2 g de poudre d'os et est suivi d'une purification à l'aide de colonnes de silice du kit CleanMix (Talent). L'ADN a été élué dans 400 μ L d'eau ultra-pure puis concentré à l'aide de filtres Microcon YM-30 (Millipore) dans 40-50 μ L de tampon Tris-EDTA (TE) 1X. Tous les extraits d'ADNa ont été conservés à -20°C pour une meilleure conservation de l'ADN au cours du temps.

III.3.2 Quantification de l'ADN par PCR en temps réel

La quantité d'ADN pouvant varier considérablement d'un échantillon à l'autre et d'un substrat à l'autre, nous avons quantifié l'ADN humain présent dans tous les extraits d'ADN humains contemporains et tous les extraits d'ADN yakoutes anciens. Les extraits d'ADNa obtenus à partir des prélèvements de la région de Krasnoyarsk n'ont pas pu être tous quantifiés. En effet, des analyses génétiques avaient déjà été réalisées auparavant sur ces prélèvements et donc la quantité de matériel (extraits d'ADN et prélèvements) dont nous disposions pour ce travail était très limitée.

Tous les extraits d'ADN humain ont été quantifiés à l'aide du kit Quantifiler[®] (AB) et de l'appareil de détection ABI Prism 7000 Sequence Detection System (AB) selon les recommandations du fabricant.

Le principe de ce kit repose sur une réaction de PCR dite quantitative ou en temps réel. A l'inverse d'une PCR classique, la PCR en temps réel permet de mesurer la quantité d'ADN tout au long de la réaction grâce à l'utilisation d'un marqueur fluorescent dont la fluorescence est proportionnelle à la quantité d'amplicons produite. Le kit Quantifiler[®] utilise une sonde TaqMan qui s'hybride spécifiquement à la séquence cible humaine amplifiée (un fragment de 62 pb du gène codant pour la télomérase reverse transcriptase humaine). Cette sonde est marquée à son extrémité 5' par un fluorophore émetteur (reporter) et à son extrémité 3' par une molécule appelée suppresseur de fluorescence (quencher). Si la sonde est intacte, aucune fluorescence n'est détectée en raison de la proximité de ces deux molécules. Lors de la phase d'élongation de la réaction PCR, la sonde hybridée à sa séquence cible va être dégradée par l'ADN Polymérase grâce à son activité 5' nucléase. Le fluorophore émetteur, alors séparé du suppresseur, va pouvoir émettre de la fluorescence qui sera détectée au niveau d'une caméra. Pendant la phase exponentielle de la réaction PCR, l'augmentation du signal de fluorescence détecté est proportionnelle à la quantité d'amplicons générée. Il est possible de déterminer le nombre de molécules d'ADN cibles de départ en fonction du nombre de cycles nécessaire pour atteindre une valeur seuil de fluorescence. Un échantillon d'ADN de concentration connue sert à réaliser une gamme de dilutions qui permet de réaliser une courbe standard à partir de laquelle sont déduites les concentrations des échantillons. La quantification est donc déduite de la cinétique de réaction.

Outre le fait d'estimer la quantité d'ADN humain présent dans un extrait d'ADN donné, ce kit permet également de détecter la présence d'inhibiteurs de réactions. En effet, la réaction d'amplification par PCR est une réaction multiplexe qui permet d'amplifier la séquence cible humaine mais également une séquence d'ADN cible synthétique qui sert de contrôle interne de réaction (IPC, internal PCR control). Une deuxième sonde Taqman marquée avec un fluorophore ayant un spectre d'émission différent de celui de la première sonde et spécifique de l'IPC, est utilisée. La cinétique d'amplification de l'IPC est connue dans un milieu exempt d'inhibiteurs. Une variation de cette cinétique peut par conséquent être imputée à la présence d'inhibiteurs.

Grâce à cette étape de quantification, tous les échantillons d'ADN humain contemporain ont pu être dilués à 1ng/μL afin d'utiliser la même quantité d'ADN de départ pour tous les échantillons lors des étapes de développement et d'optimisation des réactions de typage des SNP. Les extraits d'ADN présentant des inhibiteurs ont été dilués tout en conservant une quantité d'ADN suffisante pour réaliser les analyses génétiques.

III.4 Analyse des SNP

III.4.1 Amplification par PCR des régions d'ADN contenant les SNP d'intérêt

Toute méthode de typage des SNP faisant intervenir une étape préalable d'amplification par PCR des régions d'ADN contenant les SNP d'intérêt, nous avons donc conçu des amorces PCR pour les 59 SNP qui ont fait l'objet d'analyses dans ce travail (les 43 SNP du génome humain dont 13 SNP du chromosome Y, 27 SNP de la région codante de l'ADNmt et 10 SNP autosomaux et les 16 SNP du génome des bactéries du CMTB). Le développement des réactions PCR a été conduit selon les recommandations émises par plusieurs auteurs (Rompler et al., 2006; Sanchez et Endicott, 2006).

III.4.1.1 Conception des amorces PCR

L'efficacité d'une réaction de PCR dépendant fortement du choix des amorces, nous avons conçu chaque couple d'amorces en suivant 5 étapes différentes.

1. Pour chaque SNP, la séquence nucléique comprenant le locus d'intérêt (soit environ 400 pb, 200 pb avant et 200 pb après le locus) a été retranscrite à partir de la base de données dbSNP grâce au numéro d'accèsion du SNP ou à partir de la base de données des séquences nucléiques, Genbank, en utilisant les séquences d'amorces PCR déjà publiées pour le SNP d'intérêt et le logiciel d'alignement des séquences nucléiques nBLAST. Les bases de données dbSNP et Genbank ainsi que le logiciel nBLAST sont disponibles sur le site internet du NCBI (<http://www.ncbi.nlm.nih.gov/>).
2. A partir de ces séquences, les régions qui correspondent à des répétitions connues et les régions de faible complexité ont été masquées grâce au logiciel RepeatMasker pour éviter que des amorces soient dessinées dans ces régions (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*, 1996-2004, <http://www.repeatmasker.org/>).
3. Puis, les couples d'amorces permettant d'amplifier les loci d'intérêt ont été conçus grâce au logiciel Primer3 (Rozen & Skaletsky, 2000, <http://frodo.wi.mit.edu/primer3/input.htm>). Dans le but d'amplifier de façon homogène et spécifique plusieurs loci au sein de la même réaction, les paramètres suivants ont été sélectionnés pour tous les couples d'amorces : température de fusion théorique (T_m) proche de 66 °C, composition équilibrée entre les bases A et T d'une part et G et C d'autre part, longueur des amorces comprise entre 18 et 30 nucléotides et une à trois bases G ou C parmi les 7 dernières bases à partir de l'extrémité 3' de l'amorce. Par

ailleurs, seuls les couples d'amorces générant des amplicons de taille comprise entre 80 et 150 pb ont été conservés pour que même les fragments d'ADN très dégradés puissent être amplifiés.

4. La spécificité des amorces choisies a été testée en comparant la séquence de chaque amorce à la base de données GenBank grâce au logiciel nBLAST (<http://blast.ncbi.nlm.nih.gov/>).

5. Le programme AutoDimer a finalement été utilisé pour vérifier que les amorces choisies ne formaient pas de structures secondaires telles que des dimères d'amorces ou des épingles à cheveux.

Les couples d'amorces conservés pour chaque SNP sont présentés dans le tableau joint en annexe 1.

III.4.1.2 Validation des amorces PCR

La validation des couples d'amorces PCR a consisté en l'amplification des fragments contenant les SNP d'intérêt au moyen de réactions de PCR simples (ou individuelles) puis au séquençage des amplicons pour (i) vérifier la spécificité des amorces, (ii) vérifier que les séquences d'ADN encadrant le loci d'intérêt ne présentaient pas d'autres mutations et (iii) confirmer la position du SNP d'intérêt.

Chaque région d'ADN d'intérêt a été amplifiée à partir de 1 ng d'ADN matrice (ADN humain masculin contemporain pour les SNP du génome humain et ADN de la souche de référence *M. tuberculosis* H37Rv pour les SNP du génome des bactéries du CMTB) dans un volume réactionnel final de 25 µL contenant du tampon PCR Gold 1X (AB), du MgCl₂ à 1,5 mM, 200 µM de chaque dNTP (dATP, dCTP, dGTP, dTTP), 0,6 U d'ADN Polymérase AmpliTaq Gold (AB) et 0,4 µM de chaque amorce du couple d'intérêt. Toutes les réactions d'amplification par PCR ont été réalisées dans un thermocycleur T3 ou T1 (Biometra®). Les conditions d'amplification utilisées pour les PCR simples comprenaient une étape de dénaturation initiale à 95°C pendant 5 min suivie de 33 cycles composés d'une étape de dénaturation à 95°C pendant 30 sec, d'une étape d'hybridation à 60°C pendant 30 sec et d'une étape d'extension à 72°C pendant 30 sec et finalement une étape d'extension finale à 72°C pendant 7 min.

Afin d'estimer la concentration des amplicons, 10 µL de produits PCR mélangés à 2 µL de tampon de charge Orange Loading Dye 6X (Euromedex) ont été soumis à une électrophorèse sur un gel d'agarose à 3% contenant du bromure d'éthidium (Euromedex) dans du tampon

Tris-Borate EDTA 1X. Grâce à l'utilisation du marqueur de taille Orange DNA ladder Low Range (Euromedex), la taille des amplicons générés a été observée sous lumière UV et la concentration de ces produits a été estimée. Les produits PCR restants ont été purifiés et concentrés à 2 ng/μL sur les colonnes de silice du kit NucleoSpin[®] Extract II (Macherey-Nagel) selon les recommandations du fabricant. Quatre microlitres de produits PCR à 2 ng/μL ont été séquencés à l'aide du kit ABI PRISM[®] BigDye[™] Terminator Cycle Sequencing Ready Reaction v.1.1 (AB) selon les recommandations du fabricant. Après une incubation de 15 min à -20°C en présence de 4 volumes d'éthanol 80%, les produits de séquençage ont été précipités par centrifugation à 12 000 rotations par minute (rpm) pendant 30 min. Après lavage avec 200 μL d'éthanol 70%, le culot sec a finalement été suspendu dans 10 μL d'eau ultra-pure.

Quatre microlitres de produits de séquençage purifiés, préalablement ajoutés à 16 μL d'eau ultra-pure, ont été analysés par électrophorèse capillaire sur un séquenceur automatique ABI PRISM[®] 3100 Genetic Analyzer (AB) en utilisant du polymère Performance Optimum Polymer 4 (POP-4[™], AB) selon les recommandations du fabricant. Les données brutes ont été analysées à l'aide du logiciel Sequencing Analysis v.3.7 (AB) et les séquences ont été comparées, grâce au logiciel Sequencher (GeneCodes), aux séquences de référence ayant servi à la conception des amorces.

III.4.1.3 PCR multiplexes

Afin d'utiliser une quantité minimale d'ADN de départ, des PCR permettant l'amplification simultanée de plusieurs SNP présentant un intérêt commun ont ensuite été développées sur des échantillons biologiques contemporains.

Les différentes combinaisons de SNP analysés au sein des 9 PCR multiplexes développées sont présentées dans le tableau joint en annexe 1. Dans un premier temps, toutes les PCR multiplexes ont été réalisées dans un volume réactionnel final de 50 μL contenant du tampon PCR Gold 1X (AB), du MgCl₂ à 8 mM, 400 μM de chaque dNTP, 2U d'ADN polymérase AmpliTaq Gold (AB), 0,2 μM de chaque amorce et 1 ng d'ADN humain contemporain (SNP du génome humain) ou d'ADN bactérien contemporain (SNP du génome des espèces du CMTB). Les conditions d'amplification ont été les mêmes que celles utilisées pour les réactions de PCR simples. Les concentrations des amorces ont ensuite été ajustées jusqu'à l'obtention d'une amplification homogène de chaque type d'amplicon. Les concentrations finales d'amorces utilisées dans chaque PCR multiplexe sont indiquées sur le tableau joint en

annexe 1. Les optimisations réalisées pour l'analyse des SNP sur les échantillons anciens sont détaillées dans un chapitre ultérieur.

III.4.2 Typage des SNP par spectrométrie de masse MALDI-TOF

Huit SNP autosomaux et les 8 SNP permettant de différencier les espèces du CMTB et les groupes génétiques PGG ont fait l'objet de développements sur des échantillons contemporains au moyen de la méthode GenoSNIP. Cependant, cette méthode de typage n'a pas été appliquée aux échantillons anciens.

III.4.2.1 Réaction d'extension d'amorce ou PEX

III.4.2.1.1 Conception des amorces PEX

Les amorces PEX ont été conçues avec les mêmes logiciels et en suivant la même procédure que pour le dessin des amorces PCR. Néanmoins, le choix des amorces PEX est plus limité que celui des amorces PCR. En effet, une seule amorce a été conçue par SNP de telle façon à ce qu'elle s'hybride immédiatement en amont (du côté 5') du locus d'intérêt. Pour pouvoir typer plusieurs SNP simultanément au sein de la même réaction, le T_m de ces amorces a été fixé à environ 56°C. Une molécule photoclivable (PC-o-nitrophényl-CE-phosphoramidite) qui remplace le 5^{ème} nucléotide à partir de l'extrémité 3' de chaque amorce PEX a été ajoutée. Il s'agit d'une molécule photoclivable par exposition aux rayons UV de longueur d'onde comprise entre 254 et 366 nm. Par ailleurs, des nucléotides modifiés par méthylation (*e.g.* 2'-O-méthyladénosine ou 2'-O-méthylguanosine) ont également été ajoutés afin de modifier la masse de certaines amorces (31 Da par groupement méthyl supplémentaire) pour permettre une meilleure séparation des différents produits d'extension dans les réactions multiplexes. Deux types d'amorces ont été conçus afin de pouvoir tester deux méthodes de purification des produits d'extension : la méthode *genopure* oligo (Bruker Daltonics) qui utilise des amorces simples puis la méthode *genostrep* (Bruker Daltonics) qui nécessite l'utilisation d'amorces biotinylées. Les séquences des amorces PEX conservées sont indiquées dans le tableau joint en annexe 2.

III.4.2.1.2 Validation des amorces PEX

Afin de s'assurer de la spécificité des amorces PEX et de la position des produits d'extension sur les spectres, chaque amorce PEX (non biotinylée et biotinylée) a été testée

individuellement à partir des produits d'une réaction de PCR simple ayant permis d'amplifier le fragment d'ADN contenant le SNP d'intérêt.

Les réactions d'amplification par PCR simple ont été réalisées au préalable comme précédemment décrit, si ce n'est que 5 ng d'ADN humain contemporain ou d'ADN bactérien (*M. tuberculosis* H37Rv ou *M. bovis* CIP 102426) ont été utilisés comme quantité d'ADN de départ. Les produits PCR ont été purifiés à l'aide du kit de purification d'ADN double brin *genopure* ds (Bruker Daltonics) suivant les recommandations du fabricant. Le principe de ce kit repose sur la formation de liaisons covalentes entre les billes magnétiques fournies avec le kit et les molécules d'ADN double brin. Grâce à l'utilisation d'un séparateur magnétique, l'ADN lié aux billes est séparé des composants en excès dans le milieu réactionnel après plusieurs lavages. Une fois purifiés, les produits PCR ont été élués dans 10 µL de tampon d'éluion fourni avec le kit.

Les réactions PEX simples ont été effectuées à partir de 3 µL de produits PCR purifiés dans un volume réactionnel final de 10 µL contenant du tampon HotTermiPol 1X (Solis Biodyne), du MgCl₂ à 1,5 mM, 200 µM de dNTP et ddTTP (dATP, dCTP, dGTP et ddTTP pour les SNP autosomaux et dATP, dCTP, dTTP et ddGTP pour les SNP du génome des espèces du CMTB), 2 µM d'amorce PEX et 0,5 U d'ADN Polymérase Hot TermiPol (Solis Biodyne). Les réactions PEX ont été réalisées sur un thermocycler T1 ou T3 (Biometra®) avec un programme comprenant une étape de dénaturation initiale et d'activation de l'ADN polymérase à 95°C pendant 12 min suivie de 35 cycles comprenant chacun 15 sec de dénaturation à 95°C, 1 min d'hybridation à 55°C et 30 sec d'extension à 72°C et finalement une étape d'élongation finale pendant 5 min à 72°C.

Les produits d'extension non biotinylés ont été purifiés à l'aide du kit de purification d'ADN simple brin *genopure* oligo (Bruker Daltonics) suivant les recommandations du fabricant puis ont été élués dans 10 µL de tampon d'éluion fourni avec le kit. Le principe est le même que pour le kit *genopure* ds à la seule différence que les billes magnétiques ont une affinité plus forte pour les molécules d'ADN simple brin. Les produits d'extension non biotinylés ont été exposés aux UV (312 nm) pendant une minute afin que le photoclivage puisse avoir lieu.

Les produits d'extension biotinylés ont été purifiés à l'aide de la méthode *genostrep* selon le protocole fourni par le fabricant Bruker Daltonics. Le principe de cette méthode repose sur la forte affinité de la streptavidine pour la biotine. Grâce à l'utilisation de microplaques recouvertes de streptavidine, les produits d'extension biotinylés ont été séparés des composants en excès dans le milieu réactionnel après plusieurs lavages. Ces produits ont été

élués dans 20 µL de tampon d'éluion, puis les microplaques ont été exposées aux UV (312 nm) pendant une minute afin que le photoclivage puisse avoir lieu et finalement les produits d'extension clivés ont été transférés dans des nouveaux tubes.

Les produits d'extension ont été analysés par spectrométrie de masse MALDI-TOF afin de vérifier que la taille observée de ces produits sur les spectres de masse correspondait bien à la taille attendue. Les conditions de cette analyse sont détaillées dans un chapitre ultérieur.

III.4.2.1.3 Réactions PEX multiplexes

Afin d'évaluer la capacité de multiplexage de la méthode GenoSNIP, nous avons progressivement augmenté le nombre de SNP analysé au sein des réactions PEX. Nous avons commencé par développer des réactions PEX duplexes avec les amorces non biotinylées, dans un premier temps, et celles biotinylées, dans un second temps. Au vu des résultats obtenus, seules les amorces biotinylées ont été utilisées pour la réalisation de réactions PEX triplexes.

Les réactions PCR multiplexes permettant l'amplification des régions d'ADN d'intérêt ont été réalisées au préalable comme précédemment décrit à partir de 5 ng d'ADN humain contemporain ou d'ADN bactérien (*M. tuberculosis* H37Rv ou *M. bovis* CIP 102426) puis les amplicons ont été purifiés à l'aide du kit *genopure* ds (Bruker Daltonics).

Les réactions PEX multiplexes ont été effectuées à partir de 3 µL d'amplicons purifiés dans un volume réactionnel final de 10 µL contenant du tampon HotTermiPol 1X (Solis Biodyne), du MgCl₂ à 8 mM, 400 µM de dNTP et ddTTP (dATP, dCTP, dGTP et ddTTP pour les SNP autosomaux et dATP, dCTP, dTTP et ddGTP pour les SNP bactériens), 2 µM de chaque amorce PEX d'intérêt et 1U d'ADN Polymérase Hot TermiPol (Solis Biodyne). Les conditions thermiques de réaction ont été les mêmes que celles utilisées pour les réactions de PEX simples.

Les produits des réactions PEX multiplexes ont été purifiés à l'aide du kit *genopure* oligo (produits d'extension non biotinylés) ou de la méthode *genostrep* (produits d'extension biotinylés) selon les conditions précédemment décrites.

III.4.2.2 Spectrométrie de masse MALDI-TOF

Les produits d'extension des réactions PEX simples ou multiplexes purifiés ont été analysés par spectrométrie de masse MALDI-TOF.

La qualité d'une telle analyse dépend en grande partie de la qualité du dépôt du couple « échantillon-matrice ». La méthode de la goutte séchée a été utilisée pour le dépôt de ce couple sur une cible AnchorChip™ (Bruker Daltonics, ancre de diamètre 400 µm ou 600 µm). D'abord, 0,5 à 1 µL de matrice constituée d'un mélange d'acide 3-hydroxypicolinique (3-HPA) 10 mg/mL et de sels d'ammonium 1 mg/mL (citrate de diammonium hydrogène, DAHC) ont été déposés puis ont été laissés sécher à température ambiante. Par la suite, 0,5 à 1 µL de produits d'extension purifiés ont été déposés sur la matrice cristallisée et ont également été laissés sécher à température ambiante. Le dépôt d'un échantillon a toujours été dupliqué.

L'analyse a été effectuée sur un spectromètre de masse MALDI-TOF de type BiFLEX™ IV (Bruker Daltonics) en mode linéaire positif et en utilisant les paramètres indiqués dans le tableau 6.

Paramètres de l'instrument	Valeurs pour l'analyse d'oligonucléotides (1kDa<M<4kDa)
Pression du vide	10 ⁻⁷ mbar
Puissance laser	52-80%
Nombres de tirs	120
Spectromètre	
Source ionique 1 (IS1)	19kV
Source ionique 2 (IS2)	17kV
Lentille	9kV
Extraction d'ion pulsée	200 ns
Polarité	Positive
Détecteur	Linéaire
Filtrage d'ions de basse masse	700 Da
Détection	
Gamme de masse	basse : 0,9kDa - 6kDa
Gain électrique	26x
Lissage en temps réel	oui
Paramétrage	
Fréquence laser	5 Hz
Atténuateur de laser	minimun 40/gamme 15
Digitaliseur	127 cnt

Tableau 6 : Paramètres appliqués pour l'analyse par spectrométrie de masse MALDI-TOF.

L'instrument a été calibré grâce à l'utilisation des produits de référence Oligonucleotide calibration standard Low Mass Weight (Bruker Daltonics) dont les masses sont comprises entre 1000 et 4000 Da.

Le logiciel flexControl v.3.0 (Bruker Daltonics) a été utilisé pour l'acquisition des spectres de masses et le logiciel flexAnalysis v.3.0 (Bruker Daltonics) pour l'analyse des spectres.

III.4.3 Typage des SNP par miniséquençage à l'aide du kit SNaPshot™

Si seulement 16 SNP ont fait l'objet de développements sur les échantillons contemporains au moyen de la méthode GenoSNIP, la méthode de typage des SNP par miniséquençage à l'aide du kit SNaPshot™ a été développée pour l'analyse de tous les SNP sélectionnés dans ce travail de thèse puis a été appliquée pour l'étude des échantillons anciens. Le développement et l'optimisation des réactions de miniséquençage ont été conduits selon les recommandations émises par plusieurs auteurs (Rompler et al., 2006; Sanchez et Endicott, 2006).

III.4.3.1 Développement des réactions à partir des échantillons contemporains

III.4.3.1.1 Réactions de miniséquençage ou SBE

III.4.3.1.1.1 Conception des amorces SBE

Etant donné qu'une amorce SBE doit s'hybrider immédiatement en amont du SNP d'intérêt, tout comme l'amorce PEX, la séquence de ces deux types d'amorces peut être identique. C'est pourquoi, la plupart des amorces SBE utilisées dans ce travail présentaient la même séquence que les amorces PEX pour un SNP donné. Néanmoins, la conception de ces deux types d'amorces a différé par l'ajout d'une séquence oligonucléotidique à l'extrémité 5' de certaines amorces SBE. L'ajout d'une telle séquence permet de faire varier la taille des amorces SBE utilisées au sein d'une même réaction multiplexe pour que les produits d'extension puissent être facilement distingués sur les électrophorogrammes. Les amorces SBE qui devaient être utilisées au sein de la même réaction ont été conçues pour qu'elles présentent une taille qui varie de 20 à 90 nt et qu'elles diffèrent les unes des autres par au moins 6 nt. Pour ce faire, nous avons ajouté une séquence « neutre » ainsi qu'une séquence additionnelle poly-(C) si nécessaire à l'extrémité 5' de certaines amorces SBE. Cette séquence « neutre » ne correspond à aucune séquence nucléotidique humaine connue (Lindblad-Toh et al., 2000) et a déjà été utilisée avec succès à cet effet dans une autre étude (Sanchez et al., 2003). Les séquences des amorces SBE conservées pour ce travail sont indiquées dans le tableau joint en annexe 3.

III.4.3.1.1.2 Validation des amorces SBE

Afin de s'assurer de la spécificité des amorces SBE et de la position des produits d'extension sur l'électrophorégramme, chaque amorce SBE a été testée individuellement à partir des produits d'une réaction de PCR simple ayant permis d'amplifier le fragment d'ADN contenant le SNP d'intérêt.

Les réactions d'amplification par PCR simple ont été réalisées comme précédemment décrit. Les produits PCR ont été purifiés sur les colonnes de silice du kit NucleoSpin® Extract II (Macherey-Nagel) selon les recommandations du fabricant et ont finalement été élués dans 50 µL de tampon d'éluion fourni avec le kit.

Chaque réaction de miniséquençage simple a été réalisée grâce au kit ABI PRISM SNaPshot™ Multiplex (AB) dans un volume réactionnel final de 10 µL contenant 4 µL de mélange réactionnel SNaPshot™ Multiplex Ready (AB), 0,2 µM d'amorce SBE d'intérêt et 3 µL de produits PCR purifiés. Les conditions d'extension utilisées ont comporté 25 cycles constitués d'une étape de dénaturation à 96°C pendant 10 sec suivie d'une étape d'hybridation à 50°C pendant 5 sec et une étape d'extension à 60°C pendant 30 sec.

Les produits d'extension ont ensuite été incubés à 37°C pendant 1h après l'ajout d'1U de phosphatase alcaline de crevette ou SAP (shrimp alkaline phosphatase, Amersham Biosciences). Ce traitement est important afin d'éviter la migration simultanée des [F]ddNTP non incorporés et des produits d'extension lors de l'électrophorèse capillaire, co-migration qui génère du bruit de fond sur l'électrophorégramme. La SAP permet de déphosphoryler les [F]ddNTP résiduels, ce qui altère leur migration qui n'interfère plus celle des produits d'extension. Cette enzyme a finalement été inactivée par chauffage à 75°C pendant 15 min. Les produits d'extension ont ensuite été soumis à une électrophorèse capillaire afin de vérifier que la taille observée de ces produits sur l'électrophorégramme correspondait bien à la taille attendue (*i.e.* taille de l'amorce SBE + 1 nt). Les conditions d'électrophorèse capillaire sont détaillées dans un chapitre ultérieur.

III.4.3.1.1.3 Réactions de miniséquençage multiplexes

Au total, 8 réactions de miniséquençage multiplexes ont été développées (tableau joint en annexe 3).

Les réactions PCR multiplexes permettant l'amplification des régions d'ADN d'intérêt ont été réalisées au préalable comme précédemment décrit puis les amplicons ont été purifiés à l'aide du kit NucleoSpin Extract II (Macherey-Nagel).

Dans un premier temps, chaque réaction de miniséquençage multiplexe a été réalisée dans un volume réactionnel final de 10 µL contenant 4 µL de mélange SNaPshot™ Multiplex Ready Mix (AB), du sulfate d'ammonium 20 mM, 3 µL de produits purifiés de la PCR multiplexe correspondante (les produits des deux PCR multiplexes pour les SNP du chromosome Y ont été poolés) et 0,2 µM de chaque amorce SBE d'intérêt. Les conditions thermiques de réaction ont été les mêmes que celles utilisées pour les réactions de miniséquençage simples. Dans un second temps, la concentration de chaque amorce SBE a été ajustée en fonction de l'intensité du signal obtenue sur l'électrophorégramme (*i.e.* lorsque l'intensité du signal était trop faible, la concentration de l'amorce SBE a été augmentée et inversement) dans le but d'obtenir une réaction homogène (*i.e.* des pics de même hauteur pour chaque SNP sur l'électrophorégramme). Les concentrations finales des amorces SBE utilisées dans les réactions de miniséquençage multiplexes sont indiquées dans le tableau joint en annexe 3.

Afin de limiter l'effet néfaste des [F]ddNTP résiduels lors de la migration par électrophorèse capillaire, les produits des réactions de miniséquençage multiplexes ont fait l'objet d'un traitement à la SAP selon les conditions précédemment définies.

III.4.3.1.2 Electrophorèse capillaire

Deux microlitres de produits d'extension purifiés ont été ajoutés à 17,6 µL de formamide Hi-Di™ (AB) et 0,4 µL de standard de taille interne 120-LIZ™ GeneScan (AB). Le mélange a été dénaturé par chauffage à 95°C pendant 5 min puis en le refroidissant rapidement sur un bloc réfrigéré. Les produits d'extension ont ensuite été séparés selon leur taille et la fluorescence a été détectée par électrophorèse capillaire sur un séquenceur automatique ABI PRISM® 3100 Genetic Analyzer (AB) en utilisant du polymère POP-4™ (AB) (tension d'injection 2kV, 22 sec et tension de séparation 15 kV, 1000 sec, 60°C).

Les données brutes générées par le séquenceur automatique ont été analysées à l'aide du logiciel GeneMapper ID v.3.2.1 (AB) qui a également permis d'assigner automatiquement le nom des allèles.

III.4.3.1.3 Interprétation des données

Les haplogroupes Y et mitochondriaux ainsi que les phénotypes pigmentaires (*i.e.* couleur des yeux, des cheveux et de la peau) des individus ont été estimés à partir des génotypes obtenus pour les SNP d'intérêt et des données des références bibliographiques utilisées lors de leur sélection.

L'origine géographique de ces individus a été estimée à partir des génotypes obtenus pour les 4 SNP présentant des différences de fréquences alléliques entre les populations (AIM) sélectionnés dans ce travail et à l'aide du logiciel STRUCTURE 2.2 (en utilisant les paramètres par défaut) (Pritchard et al., 2000). Ce logiciel permet d'assigner un individu à l'une des populations de référence grâce aux génotypes obtenus pour différents AIM. Les 3 populations du projet HapMap, *i.e.* une population africaine (Nigéria), une population européenne (citoyens américains originaires du Nord et de l'Ouest de l'Europe) et une population asiatique (Japon et Chine), soit 251 individus ont été utilisées comme populations de référence. Les génotypes de chacun de ces individus pour les 4 loci AIM sélectionnés dans ce travail ont été obtenus à partir de la base de données dbSNP du NCBI.

L'identification des bactéries du CMTB au niveau de l'espèce ainsi que la caractérisation des SCG ont été réalisées à partir des génotypes obtenus pour les SNP d'intérêt et des références bibliographiques utilisées lors de leur sélection.

III.4.3.2 Optimisation des réactions pour les échantillons anciens

Au vu des résultats obtenus sur les échantillons contemporains, les réactions d'amplification par PCR et de miniséquençage multiplexes ont été optimisées puis appliquées aux échantillons anciens. Les modifications apportées sont décrites ci-dessous.

1. Les réactions PCR multiplexes ont été réalisées à partir d'un volume d'extrait d'ADNa variant de 5 à 15 μ L et le nombre de cycles PCR a été augmenté de 33 à 37 cycles.
2. Les produits PCR ont été purifiés sur les colonnes silices du kit NucleoSpin[®] Extract II (Macherey-Nagel) selon les recommandations du fabricant et ont finalement été élués dans 20 μ L de tampon d'éluion fourni avec le kit au lieu de 50 μ L pour les échantillons contemporains.
3. Les réactions de miniséquençage multiplexes ont été réalisées à partir de 3 μ L de produits PCR purifiés comme pour les échantillons contemporains mais le nombre de cycles de réaction d'extension a été augmenté de 25 à 30 cycles.

4. Les SNP qui n'ont pas pu être analysés au sein des séries de réactions multiplexes ont été amplifiés puis typés au sein de réactions individuelles.

Les différents ensembles de SNP du génome humain (SNP du chromosome Y, SNP de la région codante de l'ADNmt et SNP autosomaux) ont été analysés au moyen des réactions optimisées à partir des échantillons anciens provenant de la région de Krasnoyarsk.

Seuls les 4 SNP permettant de différencier les espèces *M. bovis* et *M. tuberculosis* et de caractériser les 3 groupes génétiques PGG (*i.e.* gyrB(1410), katG²⁰³, katG⁴⁶³ et gyrA⁹⁵) ont été analysés au moyen du miniséquençage à partir des échantillons yakoutes anciens.

III.5 Analyse de la séquence IS6110

La séquence IS6110 a été analysée au moyen d'une amplification par PCR nichée suivie du séquençage classique des amplicons. Il s'agit de la méthode de référence des études d'ADNa pour diagnostiquer une tuberculose. L'intérêt de la PCR nichée par rapport à une PCR simple est d'augmenter la spécificité de la réaction grâce à l'utilisation de deux couples d'amorces (externes et internes). Dans un premier temps, un fragment d'ADN de 123 pb de la séquence IS6110 est généré grâce à l'utilisation du couple d'amorces externes IS-1F (5'-CTCGTCCAGCGCCGCTTCGG-3') et IS-1R (5'-CCTGCGAGCGTAGGCGTTCGG-3') (Eisenach et al., 1990). Dans un second temps, un fragment interne de 92 pb est obtenu à partir des amplicons de la première amplification grâce à l'utilisation du couple d'amorces internes IS-2F (5'-TTCGGACCACCAGCACCTAA-3') et IS-2R (5'-TCGGTGACAAAGGC CACGTA-3') (Taylor et al., 1996). Les deux réactions d'amplification ont été validées à partir d'extraits d'ADN de *M. tuberculosis* H37Rv et *M. bovis* CIP 102426 puis les amplifications ont été réalisées à partir des échantillons yakoutes anciens. L'ADN d'un échantillon humain contemporain a été utilisé comme témoin négatif.

La première réaction d'amplification (IS-1) a été effectuée dans un volume final de 50 µL contenant 1 ng d'ADN contemporain ou 5 µL d'extrait d'ADNa, du tampon PCR Gold 1X (AB), du MgCl₂ à 1,5 mM, 0,2 mg/mL de BSA, 200 µM de chaque dNTP (dATP, dGTP, dCTP, dTTP), 0,5 µM de chaque amorce externe (IS-1F et IS-1R), 1,25 U d'ADN Polymérase AmpliTaq Gold (AB) et de l'eau ultra-pure en quantité suffisante pour atteindre un volume final de 50 µL. La PCR a été réalisée dans un thermocycleur T3 ou T1 (Biometra®) en utilisant les conditions d'amplification suivantes : 5 min de dénaturation initiale à 95°C

suivies de 40 cycles comprenant chacun une étape de dénaturation à 95°C pendant 1 min, une étape d'hybridation à 66°C pendant 1 min et une étape d'élongation à 72°C pendant 1 min et finalement 7 min d'élongation finale à 72°C.

La seconde réaction d'amplification (IS-2) a été effectuée à partir d'1 µL de produits PCR issus de la première réaction (IS-1). Les mêmes conditions réactionnelles que pour la première PCR ont été utilisées excepté que les amorces internes (IS-2F et IS-2R) ont remplacé les amorces externes et que seulement 25 cycles PCR ont été programmés avec une température d'hybridation modifiée à 58°C.

Le séquençage des amplicons a été réalisé selon le protocole précédemment décrit pour la validation des amorces PCR. Les séquences obtenues ont été comparées à la séquence de référence de l'élément *IS6110* récupérée sur la base de données Genbank du NCBI (N° d'accèsion : X17348.1).

III.6 Précautions prises lors de l'analyse des échantillons anciens

Lors des différentes étapes du traitement des échantillons anciens, de nombreuses précautions ont été prises pour ne pas contaminer ces échantillons avec de l'ADN exogène.

1. Toutes les étapes en amont des amplifications par PCR ont été réalisées dans un laboratoire d'ADNa (pré-PCR) comprenant deux parties distinctes, l'une pour le stockage et le pré-traitement (décontamination et réduction des os en poudre d'os) des prélèvements biologiques anciens, l'autre pour les extractions d'ADNa et la préparation des mélanges réactionnels pour les réactions d'amplification. Ces dernières étapes ont été effectuées sous une hotte à flux laminaire équipée de lampes à UV. Les paillasse, le matériel et les réactifs présents dans ce laboratoire d'ADNa ont été régulièrement décontaminés au moyen d'une irradiation aux UV, d'un autoclavage ou d'un traitement à la Javel ou à l'éthanol 70%. Toutes les étapes réalisées en aval des amplifications par PCR ont été réalisées dans un laboratoire d'ADN contemporain (post-PCR). Le cheminement des personnes, des échantillons et du matériel entre ces deux laboratoires ne s'est toujours fait que dans le sens pré-PCR vers post-PCR.

2. Les personnes qui ont manipulé les échantillons anciens ont toujours porté une double paire de gants et une blouse. Un masque facial a également été porté lors des étapes de pré-traitement des prélèvements anciens. De plus, le profil génétique (STR autosomaux) de ces personnes, ainsi que celui de toute personne ayant eu accès au laboratoire d'ADNa, a été établi. Les profils génétiques obtenus à partir des échantillons anciens sont toujours comparés

à ceux de la base de données du laboratoire afin de vérifier qu'il n'y a pas eu de contamination des échantillons anciens par une personne du laboratoire.

3. Des contrôles négatifs (eau ultra-pure à la place de l'ADN) ont été utilisés pour les étapes d'extraction d'ADN, d'amplification par PCR et pour les réactions d'extension (PEX et miniséquençage). Des contrôles positifs (ADN contemporain) ont également été utilisés mais ils ont été ajoutés dans le laboratoire d'ADN contemporain juste avant de placer les tubes dans le thermocycleur afin d'éviter de contaminer les échantillons anciens. De plus, les contrôles positifs et les échantillons anciens étaient toujours placés dans des blocs ou des thermocycleurs différents.

4. Pour chaque échantillon, les amplifications par PCR ont été réalisées plusieurs fois à partir d'un même extrait d'ADN et d'extraits différents dès que possible.

IV. Résultats

IV.1 Analyse des SNP par spectrométrie de masse MALDI-TOF

La méthode GenoSNIP, qui a été choisie pour le typage des SNP par spectrométrie de masse MALDI-TOF, a été développée pour l'analyse de 8 SNP autosomaux à partir de plusieurs extraits d'ADN humain contemporain et pour l'analyse des 8 SNP permettant de différencier les espèces du CMTB et les 3 groupes génétiques PGG à partir des extraits d'ADN de *M. tuberculosis* H37Rv et *M. bovis* CIP 102426. Deux méthodes de purification des produits d'extension ont été testées : la méthode *genopure* oligo (Bruker Daltonics) puis la méthode *genostrep* (Bruker Daltonics).

IV.1.1 Résultats obtenus avec la méthode de purification « *genopure* oligo »

Lors de l'utilisation de la méthode de purification des produits d'extension *genopure* oligo (Bruker Daltonics), des difficultés ont été rencontrées dès le développement des réactions PEX simples. Des exemples de spectres obtenus à partir de réactions PEX simples sont présentés sur la figure 16.

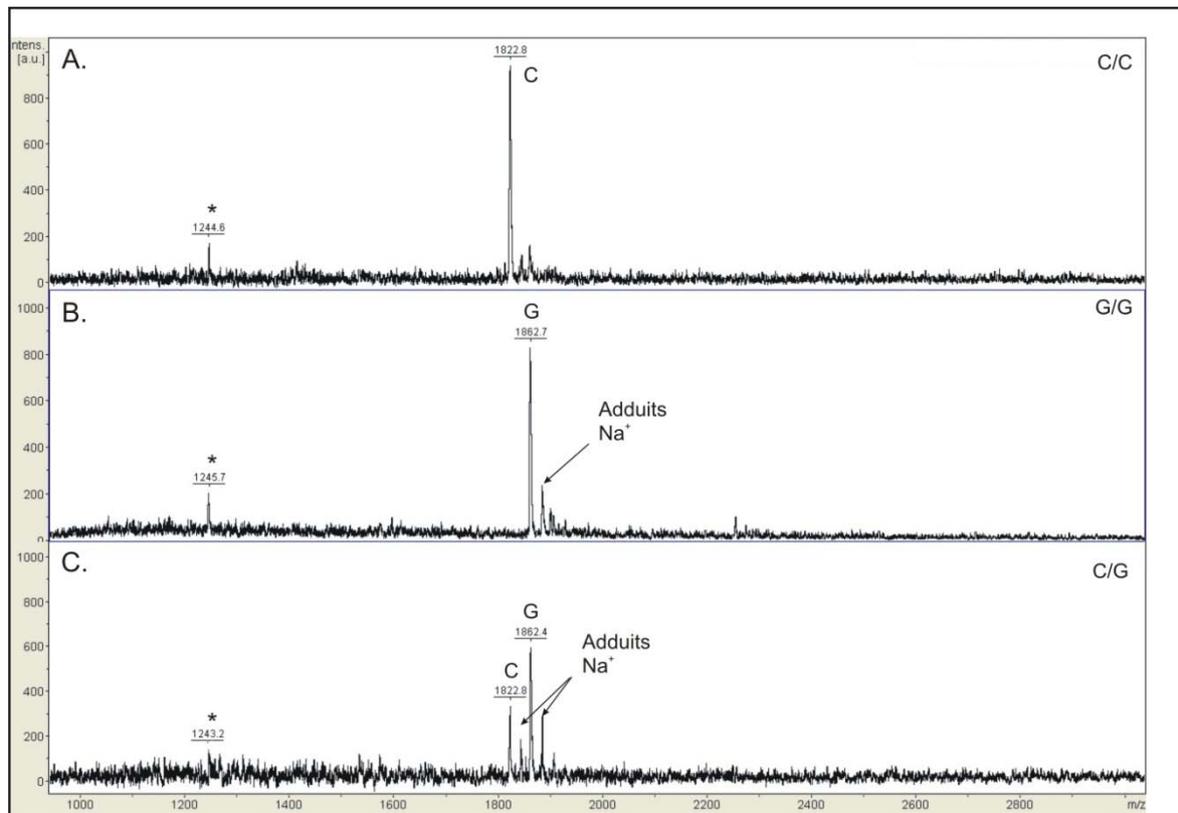


Figure 16 : Spectres obtenus pour l'analyse du locus rs16891982 à partir de 3 extraits d'ADN humain contemporains qui présentent des génotypes homozygote C/C (A.), homozygote G/G (B.) et hétérozygote C/G (C.).

Les pics normalement attendus étaient bien observés et les masses observées étaient toujours très proches voire identiques aux masses théoriques (cf. Annexe 2). Ces pics correspondaient à l'amorce PEX (représentée par une étoile sur la figure 16), qui était toujours présente en excès dans le milieu réactionnel de la réaction PEX, et aux produits d'extension. Un génotype homozygote était bien caractérisé sur les spectres par deux pics (amorce + un type de produits d'extension correspondant à un allèle) et un génotype hétérozygote par 3 pics (amorce + deux types de produits d'extension correspondant aux deux allèles). Sur la figure 16, le pic correspondant à l'amorce PEX n'est pas très visible sur le spectre C, ce qui s'explique par le fait que la plupart des amorces PEX ont pu être utilisées lors de la réaction PEX. Grâce à l'incorporation de la molécule photoclivable dans la séquence des amorces, ces pics étaient bien observés dans la gamme de masse allant de 1 000 à 3 000 Da. Cependant, outre les pics attendus, plusieurs pics parasites ont également observés (*e.g.* adduits Na^+ sur le spectre C de la figure 16).

Des réactions PEX duplexes (*i.e.* pour le typage simultané de 2 SNP) ont néanmoins été effectuées mais la présence de ces pics parasites a été augmentée, ce qui a chargé les spectres et rendu l'assignation des allèles parfois difficiles. Un exemple de spectre obtenu à partir d'une réaction PEX duplexe est présenté sur la figure 17 (spectre A). Comme l'illustre cette figure, l'origine des pics parasites a pu être déterminée dans la plupart des cas grâce à leur masse. Les pics observés dans la gamme de masse allant de 3 000 à 7 000 Da correspondaient vraisemblablement à des produits non clivés ainsi qu'aux fragments résiduels des produits clivés (Figure 18, A.). Néanmoins, ces pics n'ont pas gêné l'assignation des allèles puisqu'ils n'étaient pas présents dans la gamme de masse d'intérêt. En revanche, des produits secondaires ont été observés dans la gamme de masse allant de 1 000 à 3 000 Da. Il s'agissait d'adduits Na^+ ou K^+ , de produits dépurinés ou encore de produits dichargés. Au vu de ces résultats, le degré de multiplexage de cette méthode de typage de SNP n'a pas été incrémenté. Par la suite, nous avons testé la méthode de purification *genostrep* (Bruker Daltonics).

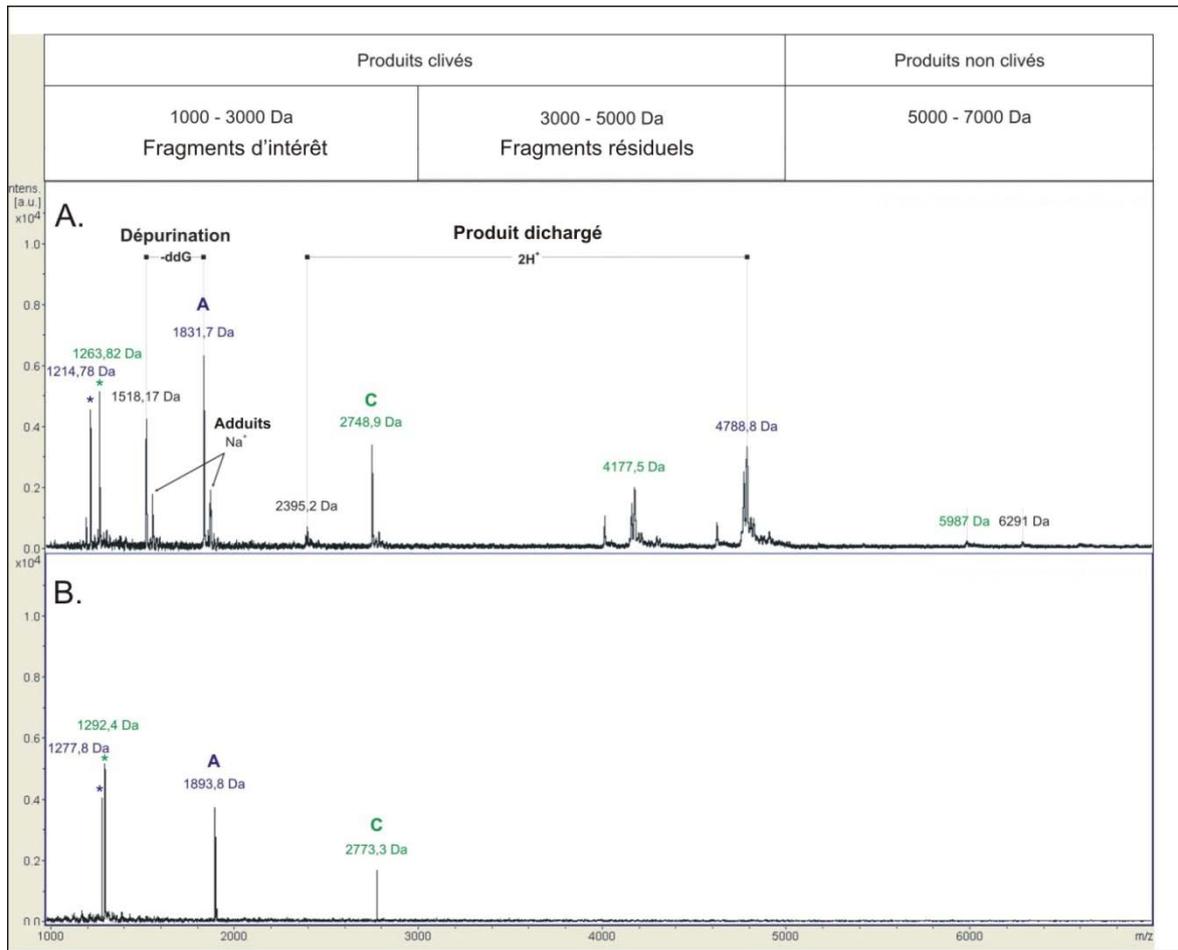


Figure 17 : Spectres obtenus pour l'analyse des loci *gyrA*⁹⁵ et *katG*⁴⁶³ à partir de l'extrait d'ADN de *M. tuberculosis* H37Rv.

A. Spectre obtenu à partir de l'analyse par spectrométrie de masse MALDI-TOF des produits d'extension d'une réaction PEX duplexe réalisée en présence des amorces PEX non biotinylées spécifiques des loci *gyrA*⁹⁵ et *katG*⁴⁶³. Avant l'analyse, les produits ont été purifiés à l'aide de la méthode *genopure* oligo (Bruker Daltonics).

B. Spectre obtenu à partir de l'analyse par spectrométrie de masse MALDI-TOF des produits d'extension d'une réaction PEX duplexe réalisée en présence des amorces PEX biotinylées spécifiques des loci *gyrA*⁹⁵ et *katG*⁴⁶³. Avant l'analyse, les produits ont été purifiés à l'aide de la méthode *genostrep* (Bruker Daltonics).

Les masses des pics relatifs au locus *gyrA*⁹⁵ sont indiquées en vert et celles de ceux relatifs au locus *katG*⁴⁶³ en bleu. Les pics correspondant aux amorces PEX en excès sont marqués par une étoile. Les pics correspondant aux produits d'extension attendus sont marqués par l'allèle auxquels ils correspondent.

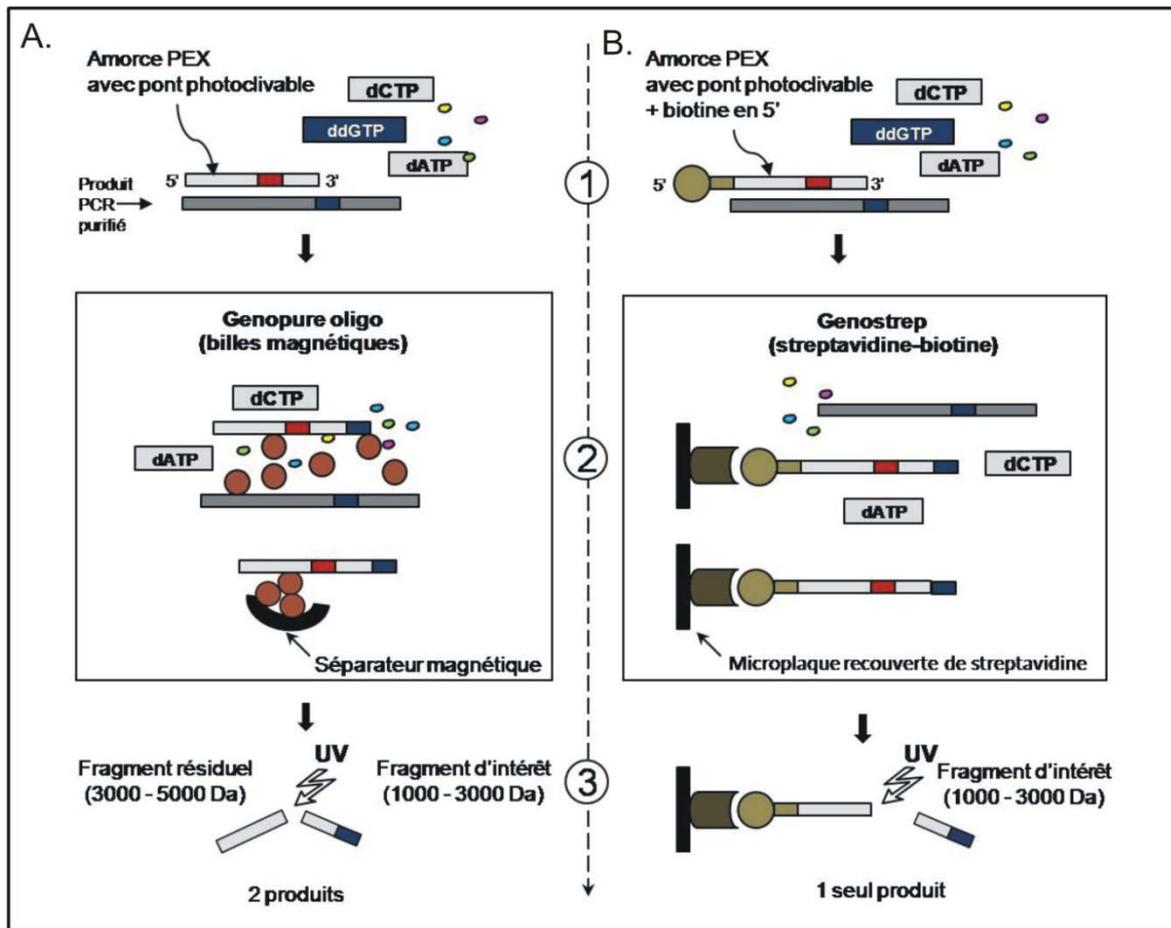


Figure 18 : Comparaison des méthodes de purification des produits d'extension « *genopure oligo* » (A.) et « *genostrep* » (Bruker Daltonics) (B.).

IV.1.2 Résultats obtenus avec la méthode de purification « *genostrep* »

Aucune difficulté n'a été rencontrée pour le développement des réactions de PEX simples lorsque la méthode *genostrep* a été utilisée pour la purification des produits d'extension. Cependant, les difficultés sont apparues lors de la mise au point des réactions multiplexes. Dans un premier temps, des réactions PEX duplexes ont été développées puis des triplexes. Certes, cette méthode de purification s'est révélée être très efficace puisque seuls les amorces et les produits attendus étaient observés sur les spectres dans la gamme de masse allant de 1 000 à 7 000 Da (Figure 17, spectre B). Contrairement à la méthode *genopure oligo* où deux fragments sont conservés après l'étape de photoclivage, la méthode de purification *genostrep* permet de ne garder que le fragment d'intérêt des produits d'extension clivés ainsi que les amorces clivées (Figure 18, B.). Cependant, certains produits d'extension attendus n'ont pas été observés sur certains spectres et ce même en répétant les réactions et en modifiant

certaines conditions de la réaction PEX (*e.g.* augmentation de la concentration des amorces PEX et du $MgCl_2$). Le multiplexage des réactions PEX s'est révélé être une étape extrêmement délicate puisque l'introduction d'une amorce PEX dans le mélange réactionnel modifie à chaque fois l'équilibre réactionnel, équilibre qui n'est pas contrôlable.

Par ailleurs, il est apparu que l'intensité des pics n'était pas forcément la même sur les spectres de masse résultants de deux dépôts réalisés dans des conditions identiques. Ce phénomène avait déjà été observé dans le travail de thèse d'Elisabeth Petkovski. L'étape de cristallisation est déterminante ; or c'est une étape qui est également difficilement contrôlable et, de ce fait, deux dépôts réalisés dans des conditions identiques peuvent cristalliser de manière différente.

Finalement, nous n'avons pas poursuivi nos développements portant sur l'analyse des SNP au moyen de la méthode GenoSNIP en raison du manque apparent de reproductibilité des dépôts mais surtout à cause de l'optimisation des réactions multiplexes qui s'est révélée être une étape très longue et complexe.

IV.2 Analyse des SNP par miniséquençage à l'aide du kit SNaPshotTM

Si de nombreux problèmes ont été soulevés lors du développement des réactions avec la méthode GenoSNIP, en revanche aucune difficulté n'a été rencontrée lors du développement des réactions avec la méthode de miniséquençage. Des réactions multiplexes ont été développées avec succès à partir des échantillons contemporains pour l'analyse des différents ensembles de SNP sélectionnés dans ce travail. Ces réactions ont été optimisées puis appliquées aux échantillons anciens d'intérêt. Les résultats obtenus sont présentés ci-dessous pour chaque ensemble de SNP sélectionné.

IV.2.1 Analyse des SNP du génome humain

IV.2.1.1 SNP du chromosome Y

Les 13 SNP du chromosome Y sélectionnés ont été analysés à partir de deux réactions de PCR multiplexes suivie d'une réaction de miniséquençage multiplexe à l'aide du kit SNaPshotTM (AB).

Les réactions multiplexes ont été développées avec succès à partir de l'analyse de 5 échantillons contemporains issus d'individus masculins d'origines géographiques européenne

et asiatique. Comme attendu pour chacun de ces échantillons, les électrophorogrammes obtenus ne présentaient qu'un seul pic pour chaque SNP, à l'exception du locus 92R7 qui en présentait parfois deux. En effet, pour ce marqueur, l'allèle muté A était bien représenté par un seul pic alors que l'allèle ancestral G était toujours associé à un variant A. D'après les données de la littérature, ce phénomène peut s'expliquer par la localisation du marqueur 92R7 dans une région dupliquée (Sanchez et al., 2004).

L'utilisation d'un échantillon contemporain féminin, comme témoin négatif de réaction, a permis de démontrer que les amplifications étaient bien spécifiques du chromosome Y. En effet, aucun pic n'a été observé sur l'électrophorogramme résultant de l'analyse de cet échantillon féminin sauf pour le locus Tat où un pic a été observé; ceci s'explique par la localisation du marqueur Tat sur le gène USP9Y du chromosome Y, qui est homologue au gène USP9X du chromosome X.

Etant donné que ces SNP ont été sélectionnés dans l'intention de déterminer les haplogroupes Y des échantillons anciens provenant de la région de Krasnoyarsk, ces séries de réactions multiplexes ont été appliquées aux 10 échantillons masculins sibériens anciens. Les résultats obtenus sont présentés dans le tableau 7.

Echantillon	M89	M9	RPS4Y ₇₁₁	M216	M217	Tat	M175	M45	92R7	M242	M3	M173	M17	Haplo-groupe Y
S07	C	C	<u>T</u>	<u>T</u>	A	T	non délété	G	GA	C	C	-	-	C(xC3)
S10	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	-	C	<u>C</u>	<u>3G</u>	R1a1
S16	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	C	C	<u>C</u>	<u>3G</u>	R1a1
S24	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	-	C	<u>C</u>	<u>3G</u>	R1a1
S25	<u>T</u>	<u>G</u>	C	C	-	T	non délété	<u>A</u>	<u>A</u>	-	C	<u>C</u>	<u>3G</u>	R1a1
S26	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	C	C	<u>C</u>	<u>3G</u>	R1a1
S28	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	C	C	<u>C</u>	<u>3G</u>	R1a1
S29	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S32	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	C	C	<u>C</u>	<u>3G</u>	R1a1
S34	<u>T</u>	<u>G</u>	C	C	A	T	non délété	<u>A</u>	<u>A</u>	-	C	-	-	P

Tableau 7 : Allèles obtenus pour les 13 SNP du chromosome Y d'intérêt analysés à partir des 10 échantillons masculins anciens provenant de la région de Krasnoyarsk.

Les allèles mutés sont soulignés et les haplogroupes qu'ils définissent sont indiqués dans la dernière colonne du tableau.

De manière générale, la plupart des marqueurs ont pu être analysés. Le marqueur M242 est celui pour lequel le typage a le moins bien fonctionné, ce qui est logique puisqu'il s'agit du

marqueur amplifié à partir du plus grand amplicon (159 pb). Seul un échantillon n'a pu être typé pour ces marqueurs, il s'agit de l'échantillon S29. Cela peut s'expliquer par la mauvaise préservation de l'ADN dans cet échantillon au vu des résultats de la quantification en temps réel des deux extraits d'ADN testés (0 pg/μL et 2 pg/μL).

Parmi les 9 échantillons anciens typés avec succès, 7 ont été assignés à l'haplogroupe R1a1, un à l'haplogroupe majeur C (S07) et le dernier à l'haplogroupe majeur P (S34). L'haplogroupe Y de l'échantillon S34 n'a pu être déterminé de façon plus précise à l'aide des SNP du chromosome Y puisque les loci M173 et M17 n'ont pu être amplifiés. Cependant, l'échantillon S34 partageait le même haplotype Y que l'échantillon S24 (analyse des STR du chromosome Y réalisée antérieurement à ce travail de thèse) et appartenait donc à la même lignée paternelle que cet échantillon. Etant donné que l'échantillon S24 a été affilié à l'haplogroupe R1a1 grâce à l'analyse des SNP, il a été possible d'en déduire que l'échantillon S34 présentait également cet haplogroupe. Un exemple d'électrophorégramme obtenu à partir de l'analyse d'un échantillon ancien affilié à l'haplogroupe R1a1 est présenté sur la figure 19.

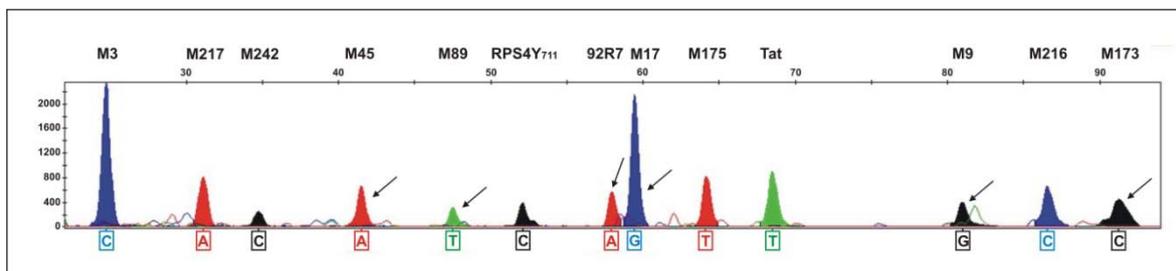


Figure 19 : Electrophorégramme obtenu pour l'analyse de 13 SNP du chromosome Y à partir d'un échantillon ancien affilié à l'haplogroupe Y R1a1.

Cet électrophorégramme a été obtenu à partir de 450 pg d'ADN de l'échantillon S26 grâce au logiciel GeneMapper ID (AB) et représente l'intensité de fluorescence en RFU (pour Relative Fluorescence Units) en fonction de la taille des produits de la réaction de miniséquençage en nucléotides (nt). Les pics correspondant aux allèles mutés ayant permis l'affiliation de cet échantillon à l'haplogroupe R1a1 sont indiqués par des flèches.

IV.2.1.2 SNP de la région codante de l'ADNmt

Les 27 SNP de la région codante de l'ADNmt sélectionnés ont été analysés à partir de 4 réactions de PCR multiplexes suivie de 4 réactions de miniséquençage multiplexes à l'aide du kit SNaPshotTM (AB).

Les réactions multiplexes ont été développées avec succès à partir de l'analyse de 10 échantillons contemporains issus d'individus d'origines européenne et asiatique. De façon

similaire à l'analyse des marqueurs du chromosome Y, des électrophorégrammes présentant un seul pic pour chaque SNP ont été obtenus pour chaque échantillon analysé.

Etant donné que ces SNP ont été sélectionnés dans l'intention de confirmer ou d'affiner les haplogroupes mitochondriaux des échantillons anciens provenant de la région de Krasnoyarsk (haplogroupes qui avaient été déduits des haplotypes mitochondriaux obtenus lors d'analyses antérieures à ce travail), les séries de réactions multiplexes d'intérêt ont donc été appliquées aux 25 échantillons anciens provenant du Sud Sibérien. Les allèles obtenus pour ces SNP à partir des échantillons anciens, ainsi que les haplogroupes qu'ils définissent, sont présentés dans le tableau joint en annexe 4. Un exemple d'électrophorégramme obtenu à partir de l'analyse d'un échantillon ancien affilié à l'haplogroupe U5a1 est présenté sur la figure 20.

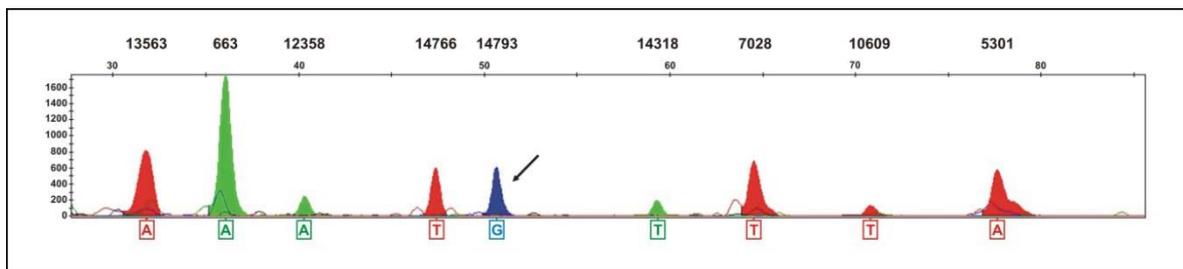


Figure 20 : Electrophorégramme obtenu pour l'analyse de 9 SNP de la région codante de l'ADNmt à partir d'un échantillon ancien affilié à l'haplogroupe mitochondrial U5a1.

Cet électrophorégramme a été obtenu à partir d'un extrait d'ADN de l'échantillon S16. Le pic correspondant à l'allèle muté ayant permis l'affiliation de cet échantillon à l'haplogroupe U5a1 est indiqué par une flèche.

Seuls deux échantillons, S07 et S24, n'ont pu être assignés à un haplogroupe mitochondrial sur la base de l'analyse des SNP de la région codante de l'ADNmt. Pour tous les autres échantillons, les données génétiques obtenues à partir de l'analyse de ces SNP sont concordantes avec celles obtenues antérieurement à partir de l'analyse des polymorphismes de la région non codante de l'ADNmt. En effet, pour la plupart des échantillons, les haplogroupes déterminés à partir des SNP de la région codante de l'ADNmt sont identiques à ceux déduits à partir des haplotypes. De plus, il a été possible d'affiner l'assignation pour 5 échantillons (S13, S27, S29, S32 et S35) qui avaient été initialement affiliés à l'haplogroupe H à partir de leurs haplotypes. Grâce au typage des SNP d'intérêt, trois de ces échantillons (S27, S29, S35) ont été affiliés à l'haplogroupe U, l'échantillon S13 à l'haplogroupe H6 et l'échantillon S32 à l'haplogroupe H5a.

IV.2.1.3 SNP autosomaux

Les 10 SNP autosomaux sélectionnés ont été analysés à partir d'une réaction de PCR multiplexe suivie d'une réaction de miniséquençage multiplexe à l'aide du kit SNaPshot™ (AB).

Les réactions multiplexes ont été mises au point à partir de 36 échantillons contemporains issus d'individus d'origines géographiques européenne, africaine et asiatique. Les traits pigmentaires (*i.e.* couleur des yeux, des cheveux et de la peau) et l'origine géographique indiqués par ces individus ainsi que les génotypes obtenus pour les 10 SNP d'intérêt sont indiqués sur le tableau joint en annexe 5. Des électrophorégrammes obtenus à partir d'échantillons issus d'individus de différentes origines (européenne, asiatique et africaine) sont présentés sur la figure 21.

Les deux SNP localisés sur le gène MC1R (rs1805007 et rs1805008) ont été choisis en raison de leur association à la couleur des cheveux roux. Selon la littérature, les génotypes « hétérozygote pour les deux loci » et « homozygote T/T pour l'un des deux locus » seraient associés à la couleur rousse des cheveux (Branicki et al., 2007). Parmi les 6 individus ayant les cheveux roux/auburn, seul un individu a été trouvé homozygote T/T pour l'un des deux SNP du gène MC1R. En revanche, les 5 autres individus sont tous hétérozygotes pour l'un des deux variants et aucun autre individu ne présente de polymorphisme pour ces deux loci.

Les 3 SNP localisés dans l'intron 1 du gène OCA2 (rs7495174, rs4778241 et rs2778138) ont été choisis en raison de leur implication dans la couleur des yeux, des cheveux et du type de peau. En fait, il s'agit de la combinaison des allèles pour les 3 loci (dans l'ordre suivant : rs7495174, rs4778241 et rs2778138) ou diplotype, qui présente un intérêt. Plus particulièrement, le diplotype TGT/TGT est retrouvé très fréquemment chez les Européens aux yeux bleus ou verts, aux cheveux clairs et à la peau claire ou mate (Duffy et al., 2007). Dans ce travail, ce diplotype a été retrouvé uniquement chez des individus d'origine européenne qui, de plus, avaient majoritairement les yeux bleus ou verts, les cheveux clairs (blonds, roux/auburn, bruns clairs) et la peau claire.

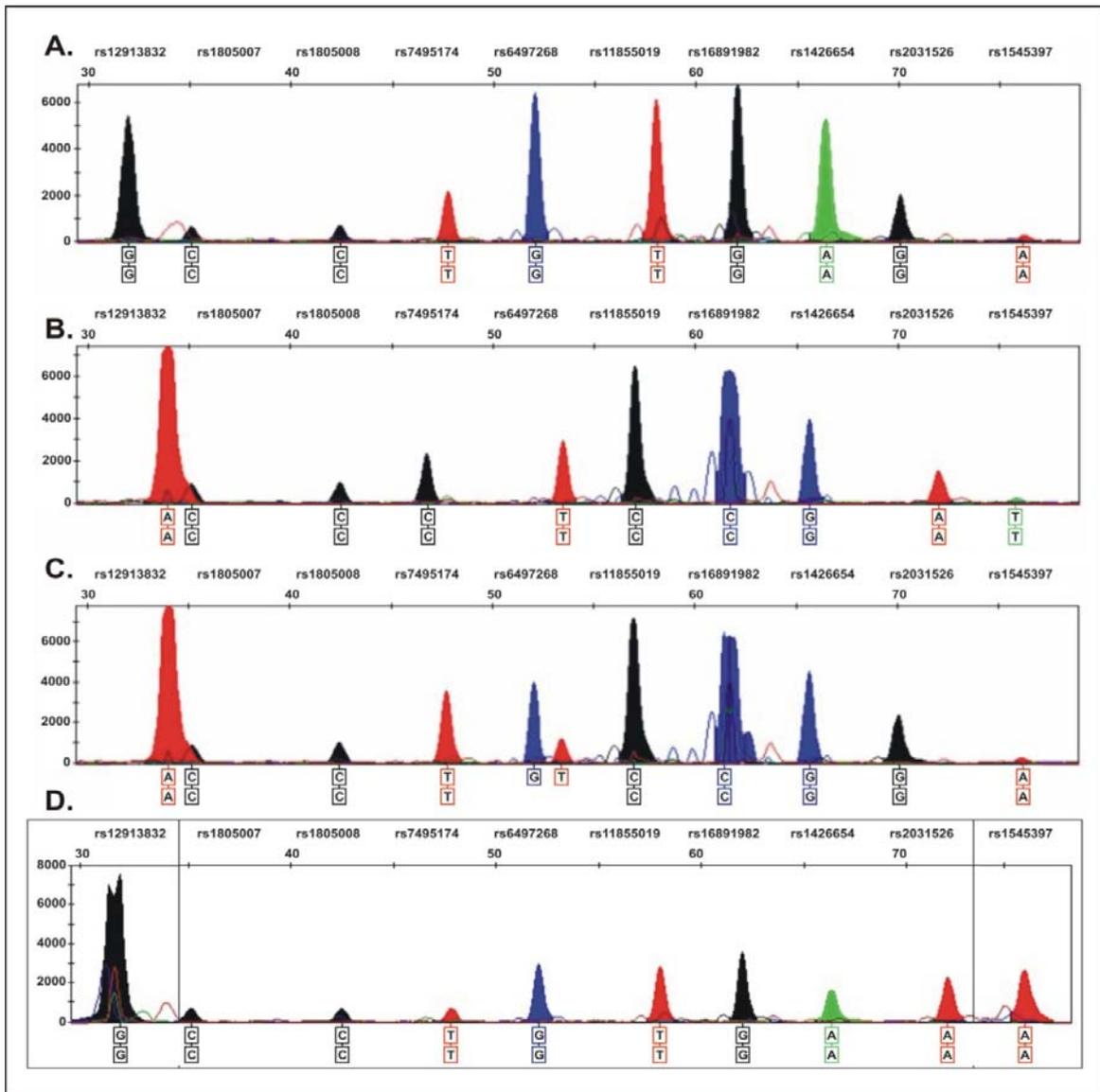


Figure 21 : Electrophorégrammes obtenus pour l’analyse de 10 SNP autosomaux à partir d’1 ng d’ADN extrait à partir d’échantillons contemporains issus d’origines géographiques européenne (A.), asiatique (B.) et africaine (C.) et d’un échantillon ancien (D.).

L’électrophorégramme D est un assemblage réalisé à partir de 3 électrophorégrammes obtenus grâce à 2 séries de réactions simples (rs12913832 et rs1545397) et une série de réactions multiplexes (8 autres loci).

Le SNP localisé au niveau du gène *HERC2* (rs12913832) a été sélectionné en raison de son association avec la couleur bleue/non bleue des yeux. Deux études récentes ont démontré que l’état homozygote G/G induirait une couleur bleue (ou verte) des yeux (*i.e.* absence de pigment bruns) (Eiberg et al., 2008; Sturm et al., 2008). Tous les individus présentant des yeux bleus ou verts sont bien homozygote G/G pour ce marqueur. Tous les individus ayant les yeux bruns sont hétérozygotes G/A ou homozygotes A/A.

Les 4 SNP restants (rs16891982, rs1426654, rs2031526 et rs1545397) localisés sur les gènes SLC45A2, SLC24A5, DCT et OCA2 ont été sélectionnés en raison des différences de fréquences alléliques qu'ils présentent entre les différentes populations humaines. Les origines géographiques des échantillons contemporains estimées à partir des génotypes obtenus pour ces 4 loci à l'aide du logiciel STRUCTURE ont toutes été concordantes avec les renseignements indiqués par les individus eux-mêmes.

Les résultats obtenus à partir des échantillons contemporains nous ont donc montré qu'il était possible d'estimer certains traits pigmentaires et l'origine géographique d'un individu à partir des génotypes obtenus pour ces 10 loci et des différentes fréquences alléliques décrites dans la littérature.

Nous avons donc analysé ces 10 loci à partir des 25 échantillons anciens provenant de la région de Krasnoyarsk dans le but d'obtenir de nouvelles informations sur leur apparence physique et leur origine. Les génotypes obtenus à partir de ces échantillons sont indiqués dans le tableau joint en annexe 6. Un électrophorégramme obtenu à partir d'un échantillon ancien est présenté sur la figure 21 (électrophorégramme D). Parmi les 25 échantillons analysés, 8 ont permis l'obtention de génotypes complets pour les 10 loci. Il est important de noter que dans la majorité des cas, tous les loci n'ont pu être analysés à partir d'une seule réaction de PCR multiplexe ; il a donc fallu réaliser des PCR simples pour pouvoir typer les loci manquants. Cependant, nous avons privilégié l'analyse des 4 SNP permettant d'estimer l'origine biogéographique (rs16891982, rs1426654, rs2031526 et rs1545397) et de celui permettant de déterminer la couleur des yeux (rs12913832) pour plusieurs échantillons où la quantité d'extrait d'ADN disponible était très limitée.

Un génotype pour les deux loci du gène MC1R a été obtenu pour 11 échantillons anciens. Ils n'étaient ni homozygote T/T pour l'un des deux marqueurs ni hétérozygote pour les deux marqueurs.

Un diplotype pour les 3 loci de l'intron 1 du gène OCA2 a été obtenu pour 12 échantillons anciens. La majorité d'entre eux (n=8) étaient homozygotes pour les 3 loci et présentaient le diplotype TGT/TGT. Les diploypes TGT/TGC (n=2), TGT/TTT (n=1) et TTT/CTC (n=1) ont également été observés. Ces résultats suggèrent que la plupart des échantillons anciens avaient un phénotype pigmentaire typiquement européen : yeux bleus ou verts, cheveux clairs et peau claire ou mate.

La couleur bleue des yeux a été confirmée par l'analyse du locus rs12913832. Un génotype pour ce locus a été obtenu pour 23 des 25 échantillons anciens analysés. La plupart présentait

le génotype G/G (n=15) suggérant que plus de 60 % des échantillons avaient probablement les yeux bleus. Les autres échantillons étaient hétérozygotes A/G ou homozygotes A/A et présentaient donc probablement des pigments bruns dans leurs iris.

Un génotype pour les 4 SNP présentant des différences de fréquence allélique entre les populations a été obtenu pour tous les échantillons anciens. L'analyse de ces génotypes à l'aide du logiciel STRUCTURE a révélée que tous les échantillons anciens sauf 3 (S07, S08 et S32) étaient plus proches des populations humaines actuelles européennes que des populations asiatiques (Figure 22). A l'inverse, deux échantillons, S07 et S08, étaient probablement plus proches des populations asiatiques actuelles. L'échantillon S32 présentait apparemment des caractéristiques génétiques communes à ces deux populations.

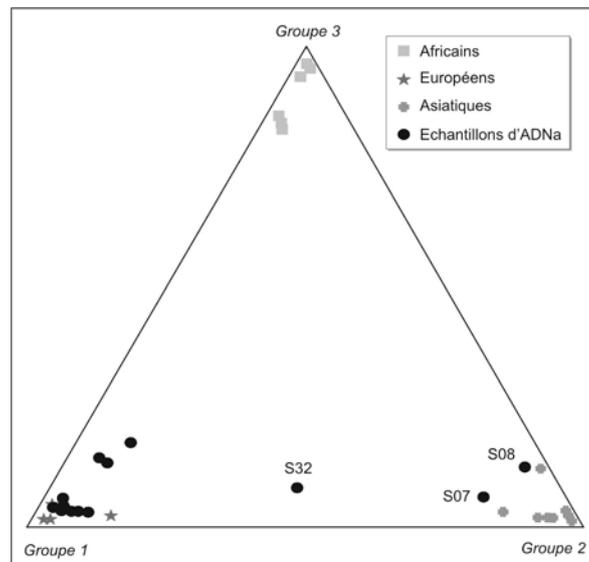


Figure 22 : Diagramme triangulaire généré à l'aide du logiciel STRUCTURE représentant les affinités génétiques entre différentes populations humaines actuelles (européenne, asiatique et africaine) et les échantillons anciens provenant de la région de Krasnoyarsk.

IV.2.2 Analyse des SNP du génome des agents pathogènes de la tuberculose

IV.2.2.1 SNP permettant d'identifier les bactéries du CMTB au niveau de l'espèce et de caractériser les groupes génétiques PGG

Les 8 SNP permettant d'identifier les agents responsables de la tuberculose (bactéries du CMTB) au niveau de l'espèce et de caractériser les 3 groupes génétiques PGG ont été

analysés à partir d’une réaction de PCR multiplexe suivie d’une réaction de miniséquençage multiplexe à l’aide du kit SNaPshotTM (AB).

Les réactions multiplexes ont été développées avec succès à partir d’extraits d’ADN des souches de référence *M. tuberculosis* H37Rv et *M. bovis* CIP 102426, puis elles ont été validées à partir de 54 extraits cellulaires de cultures bactériennes d’espèces du CMTB. Tous les loci ont pu être typés pour ces échantillons mis à part le locus *gyrA*⁹⁵ pour l’échantillon *M. canettii*. Les allèles observés pour ces échantillons sont indiqués dans le tableau 8.

Echantillons (n=56)	<u>hsp65⁶³¹</u>	<u>katG⁴⁶³</u>	<u>gyrA⁹⁵</u>	<u>katG²⁰³</u>	<u>ARNr16S¹²⁴⁹</u>	<u>gyrB(675)</u>	<u>gyrB(756)</u>	<u>gyrB(1410)</u>	PGG
<i>M. canettii</i> (n=1)	<u>T</u>	T	-	<u>C</u>	T	C	G	C	PGG1b
<i>M. tuberculosis</i> (n=7) et <i>M. africanum</i> (n=1)	C	T	C	<u>C</u>	T	C	G	C	PGG1b
<i>M. tuberculosis</i> (n=23)	C	<u>G</u>	C	<u>C</u>	T	C	G	C	PGG2
<i>M. tuberculosis</i> (n=6 dont H37Rv)	C	<u>G</u>	<u>G</u>	<u>C</u>	T	C	G	C	PGG3
<i>M. africanum</i> (n=3)	C	T	C	T	T	C	G	C	PGG1a
<i>M. pinnipedii</i> (n=1)	C	T	C	T	<u>C</u>	C	G	C	PGG1a
<i>M. microti</i> (n=1)	C	T	C	T	T	<u>T</u>	G	C	PGG1a
<i>M. caprae</i> (n=1)	C	T	C	T	T	C	<u>A</u>	C	PGG1a
<i>M. bovis</i> (n=7) et <i>M. bovis</i> BCG (n=5)	C	T	C	T	T	C	<u>A</u>	<u>T</u>	PGG1a

Tableau 8 : Allèles obtenus pour les 8 SNP permettant d’identifier les espèces du CMTB et de caractériser les 3 groupes génétiques PGG à partir des 56 échantillons bactériens contemporains testés. Les allèles caractérisant une espèce du CMTB ou un groupe génétique PGG sont soulignés.

Comme indiqué dans le tableau 8, 9 combinaisons alléliques différentes pour ces 8 SNP ont été observées à partir des 56 échantillons analysés ; ce qui est concordant avec les données de la littérature (Huard et al., 2006). Les PGG ont pu être déterminés pour tous les échantillons grâce aux génotypes pour les loci *katG*⁴⁶³, *gyrA*⁹⁵ and *katG*²⁰³ et aux données de la littérature (Sreevatsan et al., 1997; Frothingham et al., 1999) (cf. Figure 6). La souche de référence *M. tuberculosis* H37Rv a bien été affiliée au groupe phylogénétique PGG 3. Les échantillons cliniques de *M. tuberculosis* ont été affiliés aux groupes PGG 1b, 2 ou 3, l’échantillon clinique *M. canettii* au groupe PGG 1b et toutes les autres espèces (sauf un *M. africanum*) au groupe PGG 1a. Les allèles caractéristiques d’une espèce (soulignés dans le tableau 8) ont bien été identifiés uniquement chez les espèces concernées (*i.e.* l’allèle T au locus *hsp65*⁶³¹ chez *M. canettii*, l’allèle C au locus *ARNr16S*¹²⁴⁹ chez *M. pinnipedii*, l’allèle T au locus *gyrB*(675) chez *M. microti*, l’allèle A au locus *gyrB*(756) chez *M. caprae* et *M. bovis* et

l'allèle T pour le locus *gyrB*(1410) pour *M. bovis*). Par conséquent, cette série de réactions multiplexes a permis d'identifier correctement chaque espèce du CMTB. Seules les espèces *M. tuberculosis* PGG1b et *M. africanum* PGG1b n'ont pu être différenciées. Par ailleurs, la spécificité de cette série de réactions multiplexes a été démontrée puisque les différents loci n'ont pas été amplifiés à partir des témoins négatifs (*i.e.* 11 MOTT et 4 bactéries n'appartenant pas au genre *Mycobacterium*). Un exemple d'électrophorégramme obtenu pour l'analyse de ces 8 SNP à partir de chaque espèce du CMTB testée est présenté sur la figure 23.

Ces SNP ont été sélectionnés dans l'intention de pouvoir détecter la présence éventuelle d'un pathogène de la tuberculose dans les extraits d'ADNa yakoutes et en déterminer l'espèce. Cependant pour ces échantillons anciens, notre intérêt s'est porté uniquement sur les loci permettant de différencier les principaux agents pathogènes de la tuberculose chez l'Homme (*M. tuberculosis* et *M. bovis*). De ce fait, seuls 4 loci (*katG*²⁰³, *katG*⁴⁶³, *gyrA*⁹⁵ et *gyrB*(1410)) ont fait l'objet d'une amplification simultanée au sein d'une PCR multiplexe à partir des échantillons anciens et d'une analyse par miniséquençage.

Parmi les 7 échantillons anciens testés, seuls ceux provenant des tombes Batta Tcharana et Atakh ont pu être analysés avec succès pour les 4 loci. Deux loci ont également pu être typés à partir de l'échantillon provenant de la tombe Odjuluun 2. Les génotypes obtenus pour ces échantillons étaient identiques. Ils ne correspondaient pas à une espèce *M. bovis* puisqu'ils présentaient l'allèle ancestral pour le marqueur *gyrB*(1410). En revanche, l'allèle muté a été observé pour les marqueurs *katG*²⁰³ et *katG*⁴⁶³, ce qui est caractéristique des espèces *M. tuberculosis* du groupe phylogénétique PGG2 (Sreevatsan et al., 1997; Frothingham et al., 1999). Il semblerait donc que les sujets des tombes Batta Tcharana, Atakh et Odjuluun 2 aient été infectés par une bactérie *M. tuberculosis* du groupe phylogénétique PGG 2. Des électrophorégrammes obtenus pour l'analyse des marqueurs *katG*²⁰³, *katG*⁴⁶³, *gyrA*⁹⁵ et *gyrB*(1410) à partir des échantillons anciens provenant des tombes Batta Tcharana, Atakh et Odjuluun 2 sont présentés sur la figure 24.

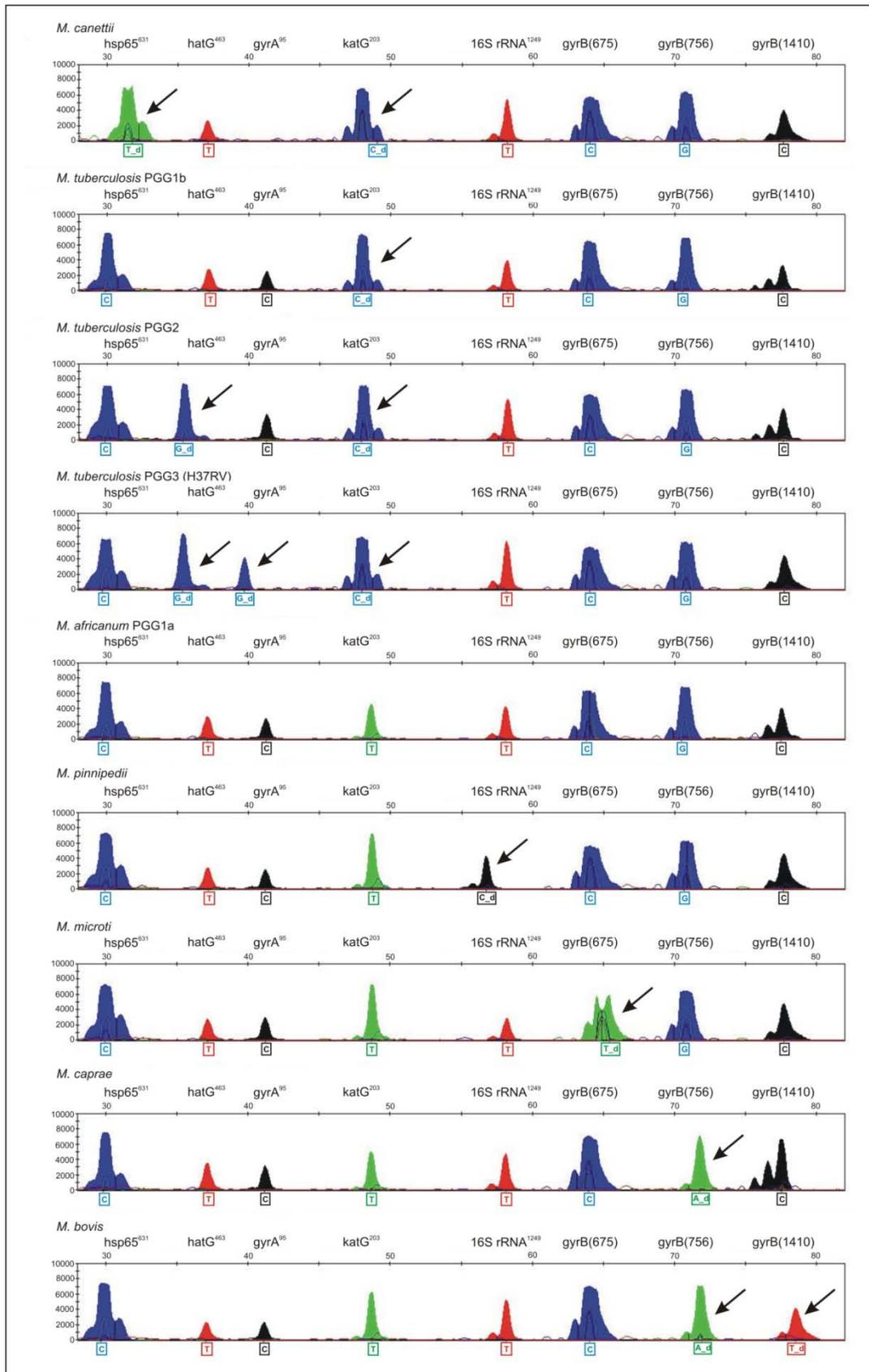


Figure 23 : Electrophorégrammes obtenus pour l'analyse des 8 SNP permettant de différencier les espèces du CMTB et de caractériser les 3 groupes génétiques PGG à partir d'échantillons contemporains correspondant aux 7 espèces du CMTB.

Les pics correspondant aux allèles caractérisant une espèce du CMTB ou un groupe génétique PGG sont indiqués par des flèches.

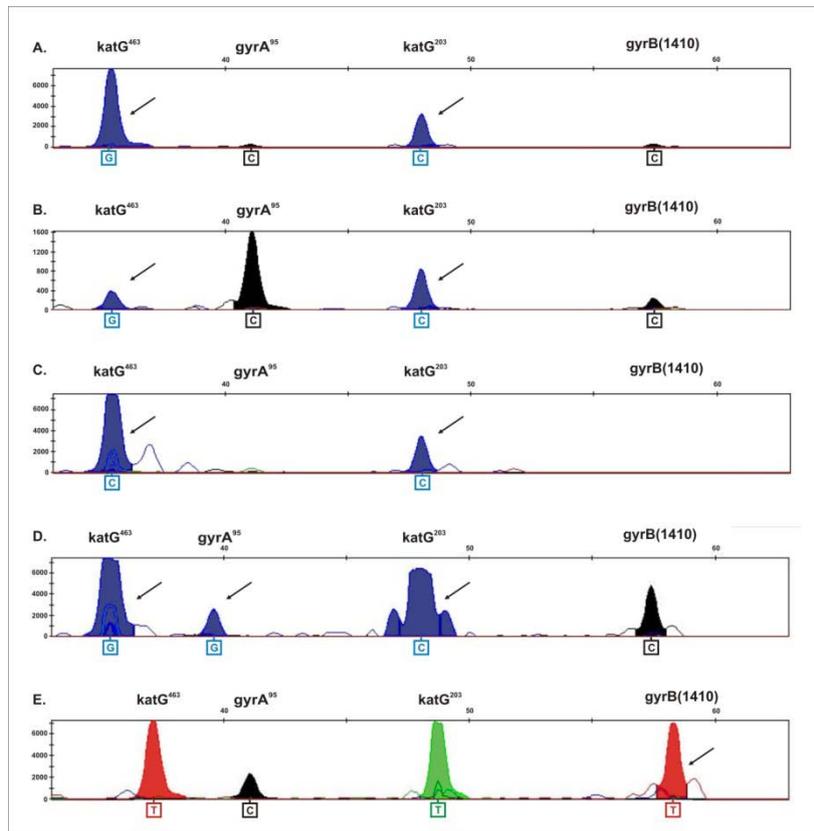


Figure 24 : Electrophorégrammes obtenus pour l’analyse des 4 SNP permettant de différencier les espèces *M. tuberculosis* et *M. bovis* et de caractériser les 3 groupes génétiques PGG à partir des échantillons yakoutes anciens provenant des tombes Batta Tcharana (A.), Atakh (B.) et Odjuluun 2 (C.) et des contrôles positifs *M. tuberculosis* H37Rv (D.) et *M. bovis* CIP 102426 (E.).

IV.2.2.2 SNP permettant la caractérisation des groupes phylogénétiques SCG

Les 8 SNP permettant de caractériser les 7 groupes phylogénétiques SCG ont été analysés à partir d’une réaction de PCR multiplexe suivie d’une réaction de miniséquençage multiplexe à l’aide du kit SNaPshot™ (AB).

Les réactions multiplexes ont été développées puis validées à partir des extraits d’ADN des souches de référence *M. tuberculosis* H37Rv et *M. bovis* CIP 102426 et de 35 extraits cellulaires *M. tuberculosis* cliniques. Tous les loci ont pu être typés avec succès pour ces échantillons (Tableau 9).

Echantillons	1977	3352929	74092	105139	2460626	232574	311613	913274	2154724 (katG ⁴⁶³)	gyrA ⁹⁵	katG ²⁰³	PGG ^a	SCG ^b
<i>M. tuberculosis</i> (n=2)	G	G	C	C	C	G	T	<u>C</u>	A	C	<u>C</u>	1b	1
<i>M. tuberculosis</i> (n=3)	G	G	C	<u>A</u>	C	G	T	C	A	C	<u>C</u>	1b	2
<i>M. tuberculosis</i> (n=2)	G	G	C	C	C	G	T	C	A	C	<u>C</u>	1b	3a
<i>M. tuberculosis</i> (n=9)	G	G	C	C	C	G	T	C	<u>C</u>	C	<u>C</u>	2	3b
<i>M. tuberculosis</i> (n=14)	G	<u>C</u>	C	C	C	G	T	C	<u>C</u>	C	<u>C</u>	2	5
<i>M. tuberculosis</i> (n=5)	<u>A</u>	<u>C</u>	C	C	C	G	T	C	<u>C</u>	<u>G</u>	<u>C</u>	3	6a
<i>M. tuberculosis</i> H37Rv	<u>A</u>	<u>C</u>	C	C	C	G	<u>G</u>	C	<u>C</u>	<u>G</u>	<u>C</u>	3	6b
<i>M. bovis</i> CIP 102426	G	G	<u>T</u>	C	C	G	T	<u>G</u>	A	C	T	1a	7

Tableau 9 : Allèles obtenus pour les 8 SNP permettant de caractériser les 7 groupes génétiques SCG à partir de 37 échantillons bactériens contemporains.

Les allèles caractérisant un groupe génétique sont soulignés. Les allèles pour les loci katG⁴⁶³, katG²⁰³ et gyrA⁹⁵ ont été déterminés grâce à la série de réactions multiplexes permettant d'identifier les espèces du CMTB et les PGG.

Comme indiqué dans le tableau 9, 8 combinaisons alléliques différentes pour ces 8 SNP ont été observées à partir des 37 échantillons contemporains analysés, ce qui est concordant avec les résultats publiés (Filliol et al., 2006; Alland et al., 2007). Chaque échantillon analysé a pu être affilié à un groupe phylogénétique SCG au vu des résultats de ces deux études. Les échantillons de référence *M. tuberculosis* H37Rv et *M. bovis* ont bien été affiliés, respectivement, au groupe SCG 6b et SCG 7. Par ailleurs, pour chaque échantillon, les PGG et les SCG sont concordants (*i.e.* SCG 1, 2, 3a, 7 → PGG 1 ; SCG 3b, 5 → PGG 2 et SCG 6a, 6b → PGG 3). Par ailleurs, cette série de réactions multiplexes est bien spécifique des espèces du CMTB puisque les différents loci n'ont pas été amplifiés à partir des témoins négatifs (*i.e.* 11 MOTT et 4 bactéries n'appartenant pas au genre *Mycobacterium*). Un exemple d'électrophorégramme obtenu à partir des espèces de référence de *M. tuberculosis* et *M. bovis* est présenté sur la figure 25.

Cette série de réactions multiplexes a été appliquée aux 7 échantillons yakoutes anciens mais la plupart des loci n'ont pu être analysés à partir d'une seule réaction PCR multiplexe. Des optimisations sont encore à réaliser avant de pouvoir caractériser le pathogène *M. tuberculosis* qui a apparemment infecté les sujets des tombes Batta Tcharana, Atakh et Odjuluun 2.

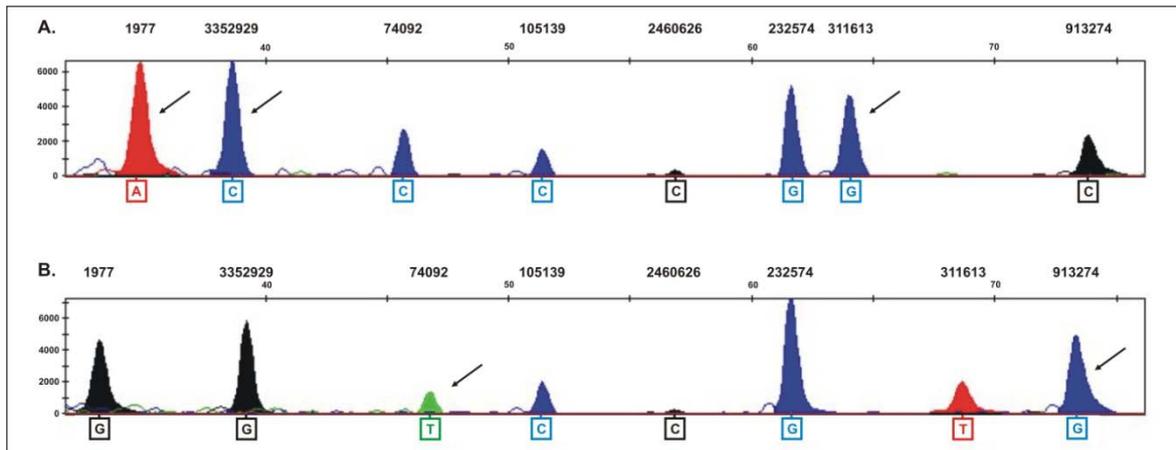


Figure 25 : Electrophorégrammes obtenus pour l'analyse des 8 SNP permettant de différencier les 7 groupes génétiques SCG à partir des échantillons *M. tuberculosis* H37Rv affilié au groupe SCG 6b (A.) et *M. bovis* CIP 102426 affilié au groupe SCG 7 (B.).

Les pics correspondant aux allèles ayant permis de déterminer les SCG pour ces deux échantillons sont indiqués par des flèches.

IV.3 Analyse de la séquence d'insertion IS6110

L'analyse de la séquence d'insertion IS6110 a été réalisée en complément de l'analyse des SNP sur les échantillons yakoutes anciens afin de pouvoir confirmer le diagnostic de la tuberculose. La méthode de référence des études d'ADNa, *i.e.* PCR nichée et séquençage des produits d'amplification, a été utilisée dans ce travail.

L'utilisation de témoins d'amplification positifs (extraits d'ADN des souches de référence *M. tuberculosis* H37Rv et *M. bovis* CIP 102426) et négatifs (eau et extrait d'ADN d'un échantillon humain contemporain) a permis de confirmer la spécificité des réactions de PCR. En effet, des amplicons de 123 pb pour la réaction de PCR IS-1 et 92 pb pour la réaction de PCR IS-2 ont été visualisés après leur migration par électrophorèse sur gel d'agarose pour les témoins positifs. Le séquençage de ces amplicons a révélé qu'il s'agissait bien des séquences attendues. A l'inverse, aucun produit d'amplification n'a été observé pour les témoins négatifs.

La séquence IS6110 a fait l'objet d'une détection sur les 7 échantillons yakoutes anciens présentant des altérations osseuses évocatrices d'une atteinte tuberculeuse. L'importance de l'utilisation d'une réaction de PCR nichée a pu être observée lors de ces analyses. En particulier pour l'échantillon provenant de la tombe Bouogaryma 2, des produits

d'amplification aspécifiques d'environ 450 pb ont été générés lors de la réaction de PCR IS-1 à partir de différents extraits d'ADN. Ils ont pu être séquencés mais leur séquence ne correspond à aucune séquence connue. Il pourrait s'agir d'un phénomène de recombinaison aléatoire appelé parfois « jumping PCR ». Néanmoins, lorsque ces produits PCR ont été utilisés comme matrice pour la réaction de PCR IS-2, aucun produit d'amplification n'a été observé après leur migration par électrophorèse sur gel d'agarose. Un gel d'agarose où ont été déposés les amplicons générés à partir des PCR IS-1 et IS-2 utilisant deux extraits d'ADNa des échantillons provenant des tombes Bouogaryma 2 et Batta Tcharana est représenté sur la figure 26.

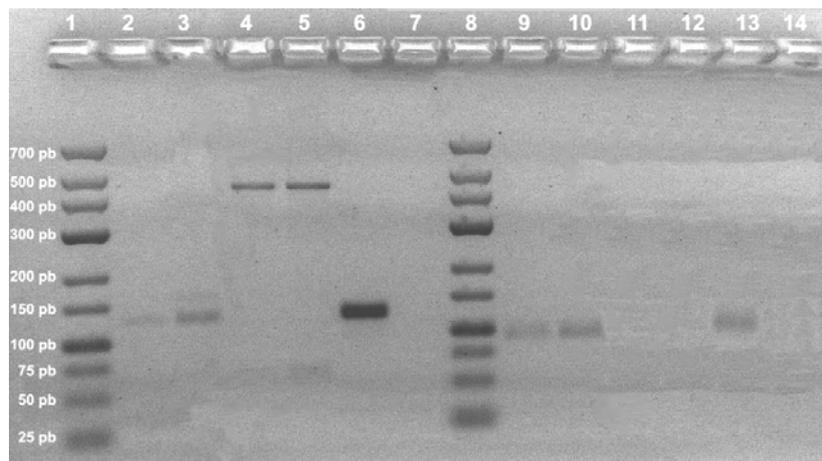


Figure 26 : Photo d'un gel d'agarose obtenu après la migration par électrophorèse de différents produits d'amplification des réactions de PCR IS-1 et IS-2 pour les échantillons Bouogaryma 2 et Batta Tcharana.

Sont représentés : le marqueur de taille (puits 1 et 8), les produits d'amplification de réactions PCR IS-1 réalisées à partir d'échantillons anciens provenant des tombes Batta Tcharana (puits 2 et 3) et Bouogaryma 2 (puits 4 et 5) et des contrôles positif *M. tuberculosis* H37Rv (puits 6) et négatif (puits 7) et les produits d'amplification de réactions PCR IS-2 réalisées à partir d'échantillons anciens provenant des tombes Batta Tcharana (puits 9 et 10) et Bouogaryma 2 (puits 11 et 12) et des contrôles positif *M. tuberculosis* (puits 13) et négatif (puits 14).

Sur les 7 échantillons anciens testés, seuls ceux provenant des tombes Batta Tcharana, Atakh et Odjuluun 2 ont permis l'obtention d'une séquence de 92 pb spécifique de la séquence IS6110. Le diagnostic de la tuberculose a donc pu être confirmé pour les sujets concernés. Ces résultats sont concordants avec ceux obtenus lors de l'analyse des 4 SNP.

Il n'est pas possible d'exclure le diagnostic de la tuberculose pour les 4 autres sujets analysés puisque le fait de ne pas avoir détecté la séquence IS6110 n'induit pas forcément l'absence

d'agents pathogènes de la tuberculose dans les 4 échantillons anciens concernés. En effet, de l'ADN de ces agents pathogènes était peut-être présent dans certains de ces échantillons mais il est possible qu'il n'est pu être détecté en raison de sa quantité trop infime ou de son état trop dégradé.

IV.4 Validation des résultats obtenus à partir des échantillons anciens

Comme dans toutes les études d'ADNa, de nombreuses précautions ont été prises afin d'éviter de contaminer les échantillons anciens analysés et de pouvoir détecter l'éventuelle présence d'ADN contaminants (cf. Matériel et Méthodes). Malgré le fait que tous les critères d'authenticité décrits dans le tableau 1 n'aient pu être observés pour des raisons techniques et matérielles, l'authenticité des résultats obtenus à partir des échantillons analysés dans ce travail est attestée par différents points.

1. Aucune amplification n'a été observée à partir des témoins négatifs d'extraction et d'amplification.
2. Tous les résultats ont été obtenus à partir d'au moins deux réactions d'amplification indépendantes réalisées à partir d'un même extrait d'ADN ou d'extraits différents quand cela était possible.
3. L'état de préservation de l'ADN endogène des échantillons anciens a été estimé par l'établissement de profils génétiques (STR autosomaux humains).

Pour les échantillons anciens provenant de Krasnoyarsk, des profils génétiques avaient été obtenus à partir de chaque échantillon antérieurement à cette étude. Ces profils n'ont pas mis en évidence de mélanges d'ADN et étaient tous différents les uns des autres ainsi que des profils des personnes du laboratoire. Des phénomènes fréquemment rencontrés lorsque de l'ADN dégradé est utilisé comme matrice, tel qu'une corrélation inverse entre l'efficacité de l'amplification par PCR et la taille des fragments amplifiés, ont été observés.

Pour les échantillons yakoutes anciens, les profils génétiques obtenus à partir de chaque nouvel extrait d'ADN étaient concordants avec ceux obtenus à partir d'autres prélèvements réalisés sur les mêmes sujets lors du travail de thèse de Sylvain Amory (2007).

L'obtention de profils génétiques de bonne qualité a donc attesté de la bonne préservation de l'ADN dans ces deux types d'échantillons sibériens anciens. Les conditions environnementales rencontrées en Sibérie ont vraisemblablement fortement contribué à l'état de préservation exceptionnel de l'ADN. En effet, le climat froid et continental ainsi que le

pergélisol de Yakoutie (sol gelé en permanence), constituent un environnement favorable pour la conservation de l'ADN.

4. Une concordance des résultats a été observée à partir des différents marqueurs analysés. D'une part, pour les échantillons anciens provenant de la région de Krasnoyarsk, les haplogroupes mitochondriaux déduits de l'analyse des polymorphismes de la région codante de l'ADNmt (haplotype mitochondrial) ont été confirmés par l'analyse des SNP de la région codante de l'ADNmt. D'autre part, pour les échantillons yakoutes anciens, les résultats obtenus à partir de l'analyse des 4 SNP et de la séquence IS6110 sont concordants puisqu'elles ont toutes les deux mis en évidence la présence d'un agent pathogène de la tuberculose dans les échantillons anciens provenant des tombes Batta Tcharana, Atakh et Odjuluun 2.

V. Discussion

Depuis une vingtaine d'années, des scientifiques tentent de décrypter l'histoire de l'Homme et l'évolution du génome humain en réalisant des analyses génétiques à partir de tissus humains anciens et en comparant ces données génétiques anciennes à celles obtenues à partir d'individus actuels. Cependant, ces études d'ADNa se heurtent à de nombreuses difficultés techniques liées au fait que les molécules d'ADN sont instables au cours du temps. En particulier, l'état fragmenté de l'ADNa restreint la taille des amplicons pouvant être générés grâce à la PCR à partir de cette matrice, ce qui limite le type d'information pouvant être obtenu à partir de l'analyse de ces molécules anciennes.

Les polymorphismes ponctuels de séquence ou SNP semblent être des marqueurs génétiques particulièrement bien adaptés pour étudier des molécules d'ADN fragmentées en raison de la petite taille de leurs régions informatives portant sur 1 nucléotide. Bien qu'il s'agisse des polymorphismes les plus abondants du génome humain et qu'ils soient localisés aussi bien sur l'ADNnu que sur l'ADNmt, seuls les polymorphismes ponctuels de la région non codante de l'ADNmt sont largement étudiés à partir de substrats biologiques anciens. Or la région non codante de l'ADNmt ne représente qu'une infime partie du génome humain et son analyse ne permet que de définir un haplotype mitochondrial qui sert à retracer l'histoire évolutive d'un individu du côté maternel uniquement.

L'objectif de ce travail de thèse a été de déterminer la pertinence de l'analyse d'autres SNP pour l'étude de l'histoire des populations humaines anciennes. Dans cette perspective, nos travaux ont principalement porté sur l'analyse de différents SNP du génome humain (27 SNP de la région codante de l'ADNmt, 13 SNP localisés sur le chromosome Y et 10 SNP autosomaux) à partir d'échantillons humains anciens provenant de la région de Krasnoyarsk (Sud de la Sibérie Centrale) afin de déterminer si l'étude de ces marqueurs permettait d'apporter de nouvelles informations sur l'origine de ces sujets de l'Age du Bronze et de l'Age du Fer. Des SNP du génome de micro-organismes pathogènes (4 SNP spécifiques du génome des agents responsables de la tuberculose) ont également été analysés à partir d'échantillons humains anciens provenant de Yakoutie (Sibérie Orientale) afin de déterminer s'il était possible d'appréhender l'état sanitaire d'une population du passé grâce à la biologie moléculaire.

V.1 Méthodes utilisées pour l'analyse des SNP

Avant de pouvoir étudier les SNP de génomes anciens, il a fallu choisir la méthode la plus appropriée à l'analyse de ces loci. Celle sélectionnée se devait d'être adaptée à l'étude de molécules d'ADN dégradées et présentes en faible quantité. La capacité d'analyse simultanée de plusieurs SNP pour réduire la quantité d'ADN initiale a donc été un critère de choix déterminant.

Au vu de données de la littérature, notre choix s'est porté sur deux méthodes de typage des SNP dont les performances sont décrites comme équivalentes et ce même à partir d'échantillons contenant peu d'ADN (Lessig et al., 2005). Il s'agit, d'une part, de la méthode GenoSNIP basée sur une détection des allèles par spectrométrie de masse MALDI-TOF et, d'autre part, de la méthode de miniséquençage à l'aide du kit SNaPshotTM (AB) basée sur une détection par électrophorèse capillaire. Ces deux méthodes ont fait l'objet de développements réalisés en parallèle sur des échantillons d'ADN contemporains mais, au vu des résultats obtenus, seul le miniséquençage a été appliqué pour l'analyse d'échantillons anciens pour des raisons de temps.

V.1.1 GenoSNIP/spectrométrie de masse MALDI-TOF

La spectrométrie de masse MALDI-TOF est une méthode de détection directe des allèles puisqu'elle permet de différencier les produits d'extension en fonction de leur masse. Contrairement aux autres méthodes de typage des SNP, et notamment au miniséquençage, la méthode GenoSNIP ne nécessite pas l'utilisation d'un marquage (*e.g.* marquage fluorescent) qui peut induire des variations d'une analyse à l'autre. Par ailleurs, grâce à l'introduction d'une molécule photoclivable dans l'amorce PEX, les produits d'extension sont détectés dans une gamme de masse réduite (1 000 à 3 000 Da au lieu de 5 000 à 9 000 Da) ce qui permet d'augmenter la résolution et la sensibilité de la détection (Sauer et al., 2003; Wenzel et al., 2003).

Les résultats obtenus dans ce travail ont montré que la méthode GenoSNIP est une méthode précise (*i.e.* les masses observées des produits correspondaient toujours aux masses théoriques) et fiable (*i.e.* les génotypes obtenus à partir de l'analyse des spectres étaient ceux attendus). Néanmoins, nous nous sommes très vite heurtés à plusieurs difficultés lors de la mise au point des réactions avec cette méthode.

En effet, il nous a fallu tester deux méthodes de purification des produits PEX associées à la méthode GenoSNIP et commercialisées par la société Bruker Daltonics. Alors que l'une est basée sur l'affinité de l'ADN pour des billes magnétiques (*genopure oligo*), l'autre est basée sur l'affinité de la biotine pour la streptavidine (*genostrep*) et nécessite par conséquent l'utilisation d'amorces PEX biotinylées (Figure 18). La méthode de purification *genopure oligo*, utilisée dans un premier temps à partir de réactions PEX simples et duplexes, s'est révélée être beaucoup moins efficace que la méthode *genostrep* utilisée dans un second temps. En effet, en plus des pics attendus, de nombreux pics parasites ont été observés sur les spectres rendant leur interprétation difficile (*e.g.* produits secondaires du photoclivage, adduits entre les produits d'extension et les sels Na⁺ ou K⁺ ou encore produits dichargés). A l'inverse, seuls les pics attendus ont été observés sur les spectres générés à partir de l'analyse des produits d'extension (de réactions PEX simples, duplexes voire même triplexes) purifiés au moyen de la méthode *genostrep*. Ces résultats ont démontré que la purification des produits d'extension est une étape très importante puisqu'elle conditionne la qualité de l'analyse des spectres. En effet, étant donné que la spectrométrie de masse n'est pas une méthode de détection spécifique des acides nucléiques, la présence dans le dépôt de composants des réactions PCR et PEX tels que les stabilisants, le glycérol et surtout les sels qui forment souvent des liaisons avec l'ADN (adduits), est probablement à l'origine de pics parasites visibles sur les spectres. Il est donc indispensable de procéder à des étapes de purification rigoureuses lorsque la spectrométrie de masse est utilisée comme méthode de détection des produits d'extension. En revanche, ces étapes de purification sont moins importantes lorsque l'électrophorèse capillaire est utilisée comme méthode de détection ou toute autre méthode basée sur une détection de la fluorescence puisque seuls les produits marqués sont détectés.

S'il a été possible de trouver une méthode de purification efficace, d'autres difficultés rencontrées lors de la mise au point des réactions multiplexes ont subsisté. Cette étape s'est révélée être longue et fastidieuse. La difficulté de l'homogénéisation des réactions multiplexes résulte du fait que la hauteur des pics sur les spectres n'est pas proportionnelle à la quantité des produits présents dans l'échantillon analysé. Il est donc nécessaire de tester différentes quantités de chaque composant réactionnel jusqu'à obtenir une réaction homogène. Cependant, même en modifiant les conditions réactionnelles, certains produits de réactions triplexes n'ont jamais pu être observés.

Etant donné que les résultats obtenus grâce à la méthode de miniséquençage ont été rapidement concluants, seules des réactions duplexes et triplexes ont été développées au moyen de la méthode GenoSNIP. Ces réactions n'ont finalement pas été appliquées aux échantillons anciens puisqu'il a été possible d'analyser un plus grand nombre de SNP à partir d'une réaction d'amplification par PCR au moyen du miniséquençage.

V.1.2 Miniséquençage à l'aide du kit SNaPshotTM/électrophorèse capillaire

Notre attention s'est portée sur la méthode de miniséquençage à l'aide du kit SNaPshotTM (AB) d'une part parce que l'efficacité de cette méthode pour l'analyse de SNP avait déjà été démontrée à partir d'échantillons anciens (Larcombe et al., 2005; Endicott et al., 2006; Sanchez et al., 2006; Burger et al., 2007) et d'autre part parce qu'elle nécessite l'utilisation d'un séquenceur automatique, appareil dont nous disposons au laboratoire.

Cette méthode a été mise au point pour l'analyse de tous les SNP d'intérêt à partir des échantillons contemporains. Aucune difficulté particulière n'a été rencontrée lors du développement des différentes réactions multiplexes. Ces réactions ont donc été optimisées puis appliquées avec succès aux échantillons anciens.

Ainsi, le miniséquençage à l'aide du kit SNaPshotTM (AB) s'est révélé être une méthode fiable, reproductible, sensible et parfaitement adaptée à l'analyse d'ADN dégradé. De plus, il s'agit d'une méthode facile à utiliser et qui nécessite peu d'optimisation. En effet, l'utilisation du kit SNaPshotTM (AB) facilite la mise au point des réactions de miniséquençage multiplexes puisque ce kit contient un mélange réactionnel optimisé auquel les amorces de miniséquençage et les amplicons sont ajoutés. Les réactions multiplexes ont été facilement optimisées en utilisant le protocole publié par Sanchez et Endicott (2006). En effet, étant donné que l'intensité de fluorescence (hauteur) des pics sur un électrophorégramme est proportionnelle à la quantité des produits d'extension dans l'échantillon analysé, les concentrations des amorces PCR et SBE ont été ajustées jusqu'à ce que l'intensité de fluorescence soit la même pour tous les pics. Des réactions multiplexes homogènes ont été obtenues à partir de l'analyse des échantillons contemporains. En revanche, des variations de l'intensité du signal entre les différents pics d'un électrophorégramme ont souvent été observées lors de l'analyse des échantillons anciens mais cela s'explique par les dommages de l'ADN qui peuvent affecter l'efficacité de l'amplification de certains fragments d'ADN au sein d'une réaction multiplexe. Des réactions simples ont donc été réalisées pour les loci qui n'ont pas pu être analysés à partir de réactions multiplexes.

Le seul inconvénient que pourrait présenter le miniséquençage est qu'il ne s'agit pas d'une méthode de typage à haut débit d'analyse (*e.g.* un maximum de 96 échantillons en 30 min environ peut être détecté par électrophorèse capillaire alors que 384 échantillons peuvent être détectés en 45 min grâce à la spectrométrie de masse). Si cela n'est pas un inconvénient pour les études d'ADNa puisque le nombre d'échantillons anciens à analyser reste toujours relativement limité, cela pourrait être plus problématique pour des applications nécessitant l'analyse d'un plus grand nombre d'échantillons.

V.2 Apport de l'analyse des SNP à l'histoire du peuplement de la Sibérie

Actuellement, le peuplement de la Sibérie reste l'un des moins bien connus au monde. Jusqu'à présent, les vestiges archéologiques n'ont pu fournir que des indications très sommaires sur l'histoire des populations sibériennes. En effet, les hypothèses basées sur les données archéologiques souffrent soit de l'indigence de celles-ci et/ou de leurs datations (les datations C14 sont relativement rares), soit des méthodes utilisées (craniologie et typologie raciale) et/ou des présupposés mis en œuvre (successions de populations faisant table rase des sujets antérieurs notamment). Dans le même temps, le peuplement de la Sibérie reste un cas d'école parmi les plus intéressants au monde, d'une part parce qu'il se situe à la jonction des mondes européen, asiatique et américain, d'autre part parce qu'il comporte encore de nombreuses populations autochtones qui possèdent leurs langues propres, certaines rattachées à différentes familles linguistiques, et qui poursuivent encore leur mode de vie traditionnel (*e.g.* agriculteurs, éleveurs nomades, semi-nomades ou sédentaires ou encore chasseurs-cueilleurs) (Crubézy et Alexeev, 2007). Par ailleurs, ces populations évoluent dans un contexte écologique très particulier, mêlant steppes du sud et de la région péri-Baïkal à celles de la taïga plus au nord marqué par des hivers longs et froids et la présence du pergélisol (sol gelé en permanence), ce qui est favorable pour la bonne conservation de l'ADN dans les restes humains sibériens anciens.

Au commencement de ce travail de thèse, une étude d'ADNa avait déjà été initiée dans le but de déterminer l'origine de sujets sibériens anciens mis au jour dans la région de Krasnoyarsk, au Sud de la Sibérie Centrale (Figure 11). Ces sujets avaient été exhumés à partir de tombes surmontées de tumuli funéraires et, d'après les fouilleurs, étaient porteurs de différentes cultures de l'Age du Bronze et du Fer (Andronovo, Karasuk, Tagar et Tashtyk). L'intérêt

d'étudier l'origine de ces sujets réside dans le fait que certaines de ces cultures sembleraient être liées à la culture des « kourganes », culture considérée par certains comme liée à l'origine des Indo-Européens (Gimbutas, 1970). Ces échantillons avaient fait l'objet d'investigations génétiques au niveau des marqueurs couramment utilisés dans les études d'ADNa : les STR autosomaux (profil génétique), les STR du chromosome Y (haplotype Y) et la région HVR I de l'ADNmt (haplotype mitochondrial) (cf. article Keyser et al., 2009 reproduit dans la partie Publications). Si ces analyses ont révélé la bonne préservation de l'ADN dans ces échantillons, elles se sont révélées insuffisantes pour retracer l'origine et l'histoire des populations du Sud Sibérien. Nous avons donc souhaité approfondir ces travaux au moyen de l'analyse de 3 types de SNP du génome humain localisés au niveau du chromosome Y, de la région codante de l'ADNmt et des autosomes.

V.2.1 Analyse des SNP du chromosome Y

L'analyse des SNP du chromosome Y a révélé que tous les sujets masculins anciens étudiés appartenaient à l'haplogroupe R1a1, à l'exception du sujet S07 qui a été affilié à l'haplogroupe majeur C (cf. articles Bouakaze et al., 2007 et Keyser et al., 2009 reproduits dans la partie Publications). Alors que l'haplogroupe C présente une distribution géographique limitée aux populations actuelles d'Asie du Sud et du Nord-Est, d'Océanie et d'Amérique, l'haplogroupe R1a1 est très répandu dans l'ensemble de l'Eurasie (Figure 27).

L'haplogroupe R1a1 est retrouvé dans les populations actuelles d'Europe Centrale et de l'Est (surtout en Pologne, Hongrie et Ukraine), de la partie européenne de la Russie (surtout dans la région Volga-Oural), de l'Asie du Sud (surtout en Inde et au Pakistan) et du Sud de la Sibérie (Semino et al., 2000; Derenko et al., 2006; Fechner et al., 2008). En revanche, il est très rarement retrouvé en Europe de l'Ouest et en Asie de l'Est. Cette vaste répartition géographique pourrait être liée, selon certains auteurs, à la diffusion des langues indo-européennes et, de ce fait, l'haplogroupe R1a1 pourrait correspondre à une signature génétique de l'expansion des Indo-Européens (Zerjal et al., 1999 ; Semino et al., 2000; Kharkov et al., 2004). Néanmoins, l'origine de cet haplogroupe constitue encore aujourd'hui un sujet de controverses. D'après certaines études, il serait apparu en Europe de l'Est, dans une région de l'actuelle Ukraine, après le dernier maximum glaciaire soit il y a 10 000-15 000 ans (Semino et al., 2000 ; Passarino et al., 2001). Par la suite, il se serait répandu à partir des steppes pontiques du nord de la Mer Caspienne et de la Mer Noire jusqu'au reste de l'Eurasie (Europe, Asie Centrale et Inde) grâce aux migrations des populations de la culture Kourgane

du III^{ème} millénaire au I^{er} millénaire avant J.-C (Rosser et al., 2000; Wells et al., 2001). La domestication du cheval aurait été le principal vecteur de cette expansion. Selon la théorie de Marija Gimbutas, les porteurs de la culture Kourgane auraient été les Proto-Indo-Européens, locuteurs de la langue commune à toutes les langues indo-européennes actuelles (Gimbutas, 1970). Une partie des Proto-Indo-Européens aurait migré vers l'Ouest de l'Europe (expansion des Indo-Européens) alors que l'autre partie aurait migré vers l'Est de l'Eurasie (expansion des Indo-Iraniens). Cependant, des études récentes ont démontré qu'il existait probablement un autre foyer d'origine de cet haplogroupe en Asie, près de l'Inde (Mirabal et al., 2009; Sharma et al., 2009; Underhill et al., 2009). Tous les auteurs ne s'accordent pas sur l'origine du foyer primaire de l'expansion des Indo-Européens. Si nos résultats ne permettent pas de préciser l'origine des lignées paternelles des échantillons étudiés (asiatique ou européenne), les résultats obtenus à partir de l'ADNmt semblent indiquer une origine plutôt européenne.

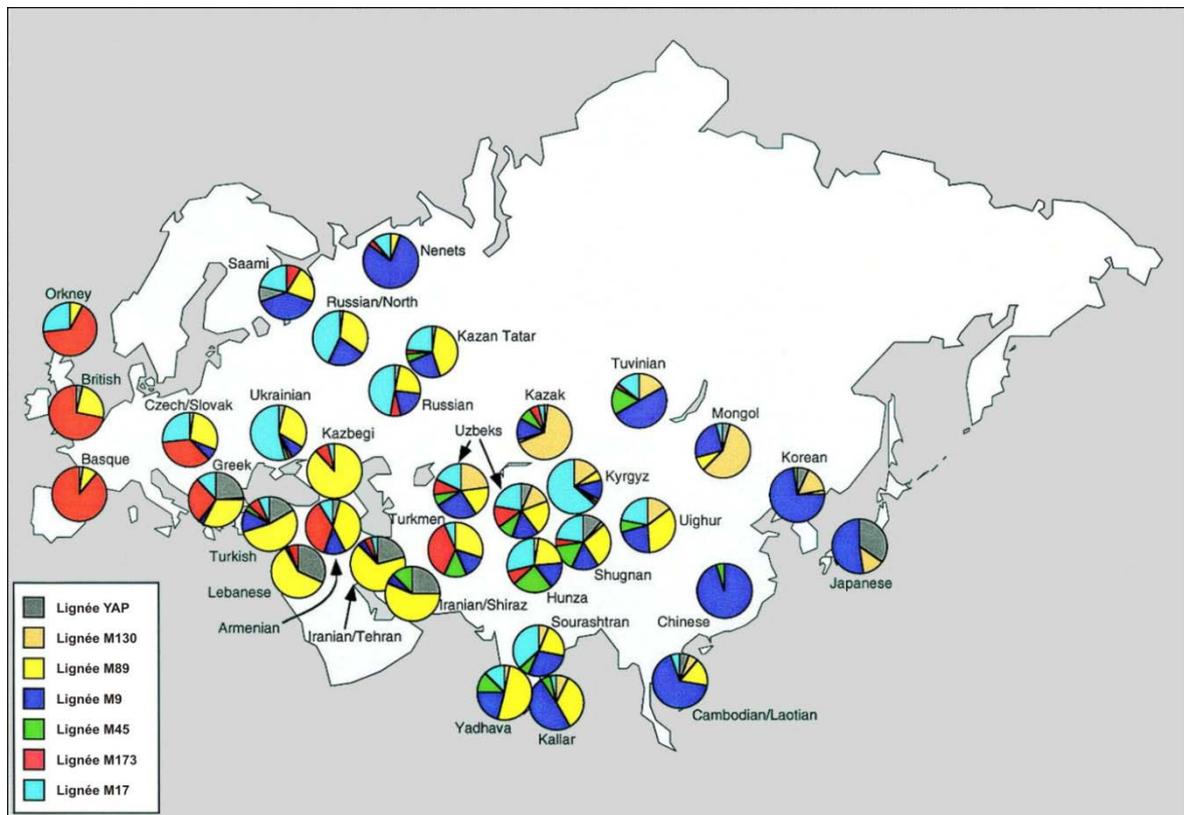


Figure 27 : Répartition géographique actuelle de l'haplogroupe Y R1a1 (M17).
(Modifiée d'après Wells et al., 2001)

V.2.2 Analyse des SNP de la région codante de l'ADNmt

L'analyse des SNP de la région codante de l'ADNmt a permis de confirmer l'appartenance des sujets sibériens anciens à 15 haplogroupes mitochondriaux différents (cf. articles Keyser et al., 2008 et Keyser et al., 2009 reproduits dans la partie Publications). Dix-neuf échantillons anciens ont été affiliés à des haplogroupes mitochondriaux fréquemment retrouvés dans les populations actuelles ouest-eurasiennes (U, UK, U2, U4, U5a1, T, T1, H5a, H6 et HV) et les 6 échantillons restants à des haplogroupes retrouvés dans les populations actuelles est-eurasiennes (Z, G2, C, F1 et N9a). Les lignées ouest-eurasiennes comptent pour 90% des échantillons datés de l'Age du Bronze (n=9/10) et 67% de ceux datés de l'Age du Fer (n=10/15). Bien qu'un certain biais d'échantillonnage ne puisse être écarté en raison du nombre restreint d'échantillons analysés, ces résultats laissent apparaître une structuration temporelle des haplogroupes de cette population sibérienne ancienne : une composante majeure européenne se distingue au cours de l'Age du Bronze et une composante asiatique commence à apparaître à partir de l'Age du Fer.

Des résultats similaires avaient été obtenus à partir d'une étude d'ADNa portant sur des échantillons de l'Age du Bronze provenant de l'actuel Kazakhstan, Asie Centrale (Lalueza-Fox, 2004). Cette étude avait montré que des sujets datant des XIII^{ème}-XVII^{ème} siècles avant J.-C. appartenaient à des lignées maternelles ouest-eurasiennes (notamment aux haplogroupes majeurs U et H) alors que les Kazakhs actuels présentent des haplogroupes mitochondriaux typiquement est-eurasiatiques. Des lignées est-eurasiennes n'avaient été observées qu'à partir d'échantillons datant des XVIII^{ème} ou XVII^{ème} siècles avant J.-C.

V.2.3 Analyse des SNP autosomaux

L'origine européenne semble également confirmée par l'analyse des SNP autosomaux qui, en plus de donner des indications sur l'aspect physique de ces sujets sibériens anciens, a livré des informations quant à leur origine biogéographique (cf. articles Bouakaze et al., 2009 et Keyser et al., 2009 reproduits dans la partie Publications).

En effet, l'analyse de 4 SNP présentant des différences de fréquences alléliques entre les populations a permis d'estimer une origine biogéographique européenne pour la majorité des échantillons analysés. Seuls les génomes nucléaires de deux échantillons, S07 et S08, présentaient clairement une composante asiatique majeure, ce qui est concordant avec l'analyse de marqueurs uniparentaux. En effet, les échantillons S07 et S08 ont été,

respectivement, affiliés à un haplogroupe Y asiatique (C) et à un haplogroupe mitochondrial asiatique (Z).

L'hypothèse d'une origine européenne pour ces sujets anciens a également été confirmée par l'analyse de SNP autosomaux associés à certains phénotypes pigmentaires. Il semble important de rappeler que l'identification de tels marqueurs phénotypiques et des gènes impliqués dans la pigmentation chez l'Homme constitue un domaine de recherche très récent et donc en constante évolution. En 2007, le gène OCA2 a été décrit comme ayant un rôle majeur dans la couleur des yeux d'un individu et pouvant également être associé à la couleur des cheveux et de la peau (Duffy et al., 2007). En 2008, il a été démontré que la mutation rs12913832, située dans une région régulatrice de l'expression du gène OCA2 (au niveau du gène HERC2 qui est localisé en amont du gène OCA2), pourrait expliquer à elle seule la couleur bleue des yeux d'un individu (Eiberg et al., 2008; Sturm et al., 2008). Au vu des résultats de l'étude de ce polymorphisme et de 3 SNP du gène OCA2, il semble que plus de 60% des sujets sibériens anciens avaient les yeux bleus ou verts. Les résultats de l'analyse des SNP du gène OCA2 ont également indiqué qu'ils avaient probablement des cheveux clairs et un type de peau clair ou mate. La majorité des sibériens anciens analysés présentaient donc un phénotype typiquement européen.

L'ensemble des résultats obtenus à partir de l'analyse des SNP autosomaux suggère donc que les sujets sibériens anciens étudiés sont plus proches, à la fois génétiquement et phénotypiquement, des populations européennes actuelles que des populations est-asiatiques actuelles. Ces résultats confortent l'étude de Moiseyev qui a montré la présence de caractères crâniens discrets européides à partir d'échantillons sibériens anciens provenant de la même région géographique et datant des Ages du Bronze et du Fer (Moiseyev, 2006).

L'apparence physique probablement européenne des sujets sibériens anciens étudiés fait penser aux momies mises au jour dans le bassin du Tarim en Chine (Figure 11) et ce d'autant plus que l'un des sujets sibériens anciens analysé présentait le même haplotype mitochondrial que l'une de ces momies chinoises (Gao et al., 2008). Ces momies naturelles datées de l'Age du Bronze (II^{ème} et I^{er} millénaire avant J.-C.) ont été retrouvées parfaitement conservées (peau, cheveux et organes internes intacts) et vêtues d'étoffes de couleur et d'accessoires en très bon état. L'apparence physique typiquement européenne de ces momies (*e.g.* cheveux clairs, long nez et grande taille) a soulevé de nombreuses questions quant à leur origine et leur présence dans cette région de la Chine. Les deux hypothèses qui ont été émises quant à leur origine sont « l'hypothèse des steppes » et « l'hypothèse des oasis de la Bactriane » (Hemphill et

Mallory, 2004). Selon « l'hypothèse des steppes », le bassin du Tarim aurait été colonisé par des cavaliers venant de la région des steppes du Nord du bassin du Tarim (Figure 11). Deux vagues de migration se seraient succédées : la première au début de l'Age du Bronze par des populations nomades porteuses de la culture d'Afanasievo (culture ayant précédé la culture d'Andronovo) puis la seconde à l'Age du Bronze Tardif par des nomades porteurs de la culture d'Andronovo. Selon « l'hypothèse des oasis de la Bactriane » (la Bactriane correspond aux actuels Turkménistan, Ouzbékistan et Afghanistan), des colons agriculteurs sédentaires provenant des oasis de l'Ouest du bassin du Tarim auraient envahi cette région (Figure 11). Les résultats obtenus dans le cadre de cette étude apportent des arguments en faveur de « l'hypothèse des steppes ». En effet, les sujets sibériens anciens étudiés proviennent du Nord du bassin du Tarim, les plus anciens ont été rattachés à la culture d'Andronovo et ils ressemblaient aux momies du bassin du Tarim sur le plan phénotypique (même aspect physique) et génétique (même haplotype mitochondrial). Ainsi, les populations sibériennes anciennes de la région de Krasnoyarsk pourraient avoir joué un rôle dans la colonisation du bassin du Tarim.

V.2.4 Hypothèses émises sur l'origine des sujets sibériens anciens étudiés

Au vu de l'ensemble des données génétiques obtenues sur les échantillons sibériens anciens, il a été possible de formuler des hypothèses concernant leur origine et leur histoire (cf. article Keyser et al., 2009 reproduit dans la partie Publications).

Le fait que la majorité des échantillons sibériens de l'Age du Bronze aient été affiliés à des haplogroupes Y et mitochondriaux ouest-eurasiatiques suggèrerait une expansion ancienne de populations porteuses de ces haplogroupes vers l'Asie, au moins jusqu'au Sud de la Sibérie occidentale (actuelle région de Krasnoyarsk). La présence d'haplogroupes mitochondriaux est-eurasiatiques retrouvée à une plus forte fréquence chez les spécimens anciens de l'Age du Fer pourrait signifier que les migrants aient rencontré des populations asiatiques et qu'ils se soient mélangés à ces populations.

Plusieurs observations nous laissent penser que ces migrants étaient probablement originaires d'Europe de l'Est.

En effet, il existe des données historiques et des traces archéologiques bien attestées qui corroborent l'existence de mouvements de populations nomades à travers l'Eurasie (d'Ouest en Est) depuis le Nord de la Mer Noire jusqu'à la région péri-Baïkal en passant par l'Asie Centrale (Hemphill et Mallory, 2004).

De plus, l'haplogroupe R1a1 auquel ont été affiliés la majorité des sujets masculins anciens est largement répandu en Eurasie, et est d'autant plus fréquent que l'on se rapproche de l'Ukraine. Il est communément admis que le foyer d'origine de cet haplogroupe pourrait se situer dans la steppe pontique (actuelle Ukraine) (Semino et al., 2000 ; Passarino et al., 2001). A cela s'ajoute le fait que selon une étude récente, la mutation rs12913832 qui caractérise 60% des sujets sibériens anciens étudiés, serait apparue dans une région autour de la Mer Noire, à proximité de l'actuelle Ukraine ou en Turquie, il y a environ 6 000 à 10 000 ans, avant de diffuser vers le reste de l'Europe (Eiberg et al., 2008).

Au vu de données archéologiques et linguistiques, ces migrants venant d'Europe de l'Est pourraient être liés aux peuples locuteurs de la langue indo-européenne originelle (langue appelée « Indo-Européen commun » ou « Proto-Indo-Européen » dont sont issues toutes les langues indo-européennes qui correspondent à la plupart des langues parlées de l'Inde à l'Europe).

En effet, selon la théorie défendue par l'anthropologue Marija Gimbutas, les locuteurs de la langue indo-européenne originelle seraient les porteurs de la culture des Kourganes, culture unique dont les kourganes (tombes creusées dans le sol et surmontées d'un tumulus) seraient le mode de sépulture caractéristique (Gimbutas, 1970). Cette théorie propose une origine de cette culture située dans la steppe pontique à l'Age du Bronze (vers - 4000 ans avant J.-C.) et l'expansion de cette culture sur les 3 millénaires suivants vers l'Ouest en Europe centrale et occidentale (dispersion des langues indo-européennes) et vers l'Est en Sibérie (dispersion des langues indo-iraniennes).

Les sujets anciens que nous avons étudiés ayant été exhumés de kourganes et ayant été rattachés, pour les plus anciens, à la culture d'Andronovo, considérée comme une culture de langue indo-iranienne (Lamberg-Karlovsky, 2002), il nous a semblé qu'un lien pouvait être fait avec la théorie des Kourganes de Marija Gimbutas et ce d'autant plus que la majorité des sujets masculins étudiés ont été affiliés à l'haplogroupe R1a1 qui est considéré comme associé à la diffusion des langues indo-européennes (Rosser et al., 2000; Wells et al., 2001).

V.2.5 Intérêts de l'analyse des SNP de génomes humains anciens

Cette étude est originale car, pour la première fois, le fait de combiner l'analyse de SNP localisés sur les génomes nucléaires et mitochondriaux a permis d'apporter de nouveaux éléments quant à l'origine des sujets anciens étudiés. Elle atteste également de l'utilité de l'analyse des SNP en complément des marqueurs plus conventionnels des études d'ADNa

(i.e. polymorphismes de la région non codante de l'ADNmt, STR autosomaux et du chromosome Y) pour étudier l'origine de populations humaines anciennes.

En effet, s'il est possible de déduire un haplogroupe à partir d'un haplotype, au moins pour l'ADNmt, en comparant l'haplotype aux bases de données d'haplotypes et aux données de la littérature, l'analyse de SNP de la région codante de l'ADNmt et du chromosome Y permet de confirmer ou de déterminer un haplogroupe et représente donc un outil supplémentaire pour l'authentification des résultats. Comme illustré par nos résultats, les polymorphismes de la région non codante de l'ADNmt ne permettent parfois que de définir un macro-haplogroupe. Le séquençage complet de l'ADNmt présentant un investissement très important, l'analyse de SNP de la région codante de l'ADNmt semble être la stratégie de choix pour définir un sous-haplogroupe souvent beaucoup plus informatif qu'un macro-haplogroupe.

L'intérêt du typage de SNP autosomaux présentant des différences significatives de fréquence allélique entre les populations (AIM) a été largement démontré ces dernières années par les études de la diversité génétique des populations actuelles (Wolinsky, 2008). Alors que les marqueurs du chromosome Y et de l'ADNmt reflètent, respectivement, l'histoire paternelle et maternelle d'un individu, les SNP autosomaux présentent une transmission biparentale et se révéleraient être des marqueurs de choix pour éclaircir l'origine de populations ayant subies des mélanges. Dans notre étude, ces SNP autosomaux ont permis de conforter les résultats obtenus à partir des marqueurs uniparentaux.

L'analyse de SNP associés à des phénotypes pigmentaires s'est quant à elle révélée très utile pour obtenir des informations sur les traits physiques des spécimens sibériens anciens. L'utilisation de telles données génétiques permet d'obtenir une représentation plus précise de ce à quoi ils pouvaient ressembler. Par exemple pour l'Homme de Néandertal, l'étude de Lalueza-Fox et al. (2007) a donné plus de crédit aux représentations lui attribuant des cheveux roux ou blonds et un teint pâle.

V.3 Apport de l'analyse des SNP à l'histoire sanitaire de la Sibérie

Si la majorité des analyses génétiques réalisées à partir de squelettes humains anciens portent sur l'analyse d'ADN humain, il est également possible, à partir de ces mêmes échantillons, d'analyser de l'ADN de pathogènes ayant infectés les individus de leur vivant. Ces analyses génétiques présentent un intérêt pour mieux comprendre l'histoire des populations anciennes

puisque'elles permettent d'obtenir des informations quant à l'état sanitaire des sujets anciens étudiés mais aussi pour étudier l'évolution des génomes de pathogènes.

Ainsi, en plus de l'histoire du peuplement de la Sibérie, nous nous sommes également intéressés aux maladies qui ont affecté les populations sibériennes anciennes. Les échantillons anciens de la région de Krasnoyarsk n'ont pas suggéré un état sanitaire particulier. En revanche, il n'en est pas de même avec ceux collectés lors de plusieurs campagnes de fouilles de la MAFSO en Yakoutie. En effet, 7 des 73 squelettes exhumés ont présenté des lésions osseuses évocatrices d'atteintes tuberculeuses. La présence de tuberculose ancienne en Sibérie a récemment été rapportée par deux publications (Taylor et al., 2007; Murphy et al., 2009). Ces articles décrivent l'identification d'une infection tuberculeuse d'origine animale (*M. bovis*) à partir de vertèbres prélevées sur plusieurs sujets datés de l'Age du Fer et provenant du cimetière d'Amyrlyg (localisé au Sud de la Sibérie à proximité de la frontière avec la Mongolie, Figure 11). Il semblerait que ces sujets aient été contaminés suite à une exposition prolongée à des bovins infectés. Bien que les sujets yakoutes anciens étudiés dans ce travail soient beaucoup plus récents (XVII^{ème}-XVIII^{ème} siècles selon la datation par dendrochronologie), il est possible qu'ils aient été exposés à un risque infectieux similaire. En effet, les Yakoutes se distinguent de leurs voisins immédiats par leur mode de vie. Ils constituent une population semi-nomade composée d'éleveurs de bovidés et de chevaux alors que les populations voisines sont constituées de chasseurs-cueilleurs et d'éleveurs de rennes. Le mode de vie des Yakoutes se rapproche étrangement de celui des pasteurs nomades d'Asie Centrale et de Mongolie et en particulier de celui des sujets anciens atteints de tuberculose bovine exhumés au cimetière d'Amyrlyg.

Dans le but de confirmer ou d'infirmer le diagnostic de la tuberculose et de déterminer l'origine de l'infection (humaine ou animale), des SNP localisés au niveau du génome d'espèces bactériennes responsables de tuberculoses humaine et bovine ont été analysés à partir des échantillons yakoutes anciens présentant les altérations osseuses susceptibles d'avoir été provoquées par une telle infection.

V.3.1 Analyse de SNP du génome des agents pathogènes de la tuberculose

Les analyses moléculaires réalisées sur les 7 vertèbres présentant des lésions lytiques évocatrices d'une atteinte tuberculeuse ont permis de confirmer le diagnostic de cette maladie pour 3 des 7 sujets concernés.

Au vu de l'absence de la mutation caractéristique des espèces *M. bovis* au niveau du locus *gyrB*(1410), nous avons pu exclure une infection d'origine animale pour ces sujets yakoutes anciens. L'origine humaine de la tuberculose a pu être confirmée grâce aux génotypes obtenus pour les deux SNP au niveau des gènes *katG* (codon 463) et *gyrA* (codon 95). Ces SNP permettent de distinguer les trois groupes phylogénétiques, PGG1, PGG2 et PGG3, dans lesquels tous les agents pathogènes de la tuberculose peuvent être classés (Sreevatsan et al., 1997). Alors que les 7 espèces du CMTB peuvent être classées dans le groupe PGG1 (*i.e.* aussi bien des agents pathogènes des animaux que de l'Homme), seuls des isolats de *M. tuberculosis* constituent les groupes PGG2 et PGG3 (Figure 6). La présence de l'allèle muté pour le locus *katG*⁴⁶³ et de l'allèle ancestral pour le locus *gyrA*⁹⁵, pour les 3 échantillons yakoutes anciens, semble indiquer qu'une espèce *M. tuberculosis* du groupe phylogénétique PGG 2 aurait été responsable de l'infection de ces individus. Le schéma évolutif proposé par Sreevatsan et al suggère que l'apparition des espèces *M. tuberculosis* PGG2 et PGG3 dites « modernes » est récente et qu'elles auraient évolué à partir des espèces *M. tuberculosis* PGG1 dites « ancestrales » (Sreevatsan et al., 1997).

Par ailleurs, 6 lignées phylogénétiques de *M. tuberculosis* et *M. africanum* peuvent être distinguées dans les populations humaines (Gagneux et al., 2006; Gagneux et Small, 2007). Parmi ces 6 lignées, les isolats *M. tuberculosis* « modernes » (*i.e.* PGG2 et PGG3) forment la lignée 4. La répartition géographique actuelle de ces lignées est représentée sur la figure 28.

Si la plupart des lignées présente une répartition géographique spécifique et restreinte à un voire deux continents, ce n'est pas le cas de la lignée 4. Il s'agit de la lignée la plus fréquemment retrouvée en Europe et sur le continent américain mais elle est également présente en Afrique, en Asie Centrale et en Russie Centrale (Dubiley et al., 2009). Il est probable que cette lignée soit présente en Sibérie, cependant, nous n'avons pas eu accès aux données épidémiologiques concernant le territoire sibérien comme illustré sur la figure 28. Une étude d'ADNa a récemment attestée de sa présence en Hongrie aux XVIII^{ème} et XIX^{ème} siècles (Fletcher et al., 2003a). De plus, il semblerait que cette lignée soit à l'origine de l'épidémie de tuberculose qui a dévasté toute l'Europe pendant ces deux siècles (Greenblatt et Spigelman, 2003).

Nos résultats confirment l'existence de souches de *M. tuberculosis* « modernes » depuis plus de 300 ans et, de plus attestent, pour la première fois, de leur présence en Sibérie, et plus précisément en Yakoutie, aux XVII^{ème} et XVIII^{ème} siècles.

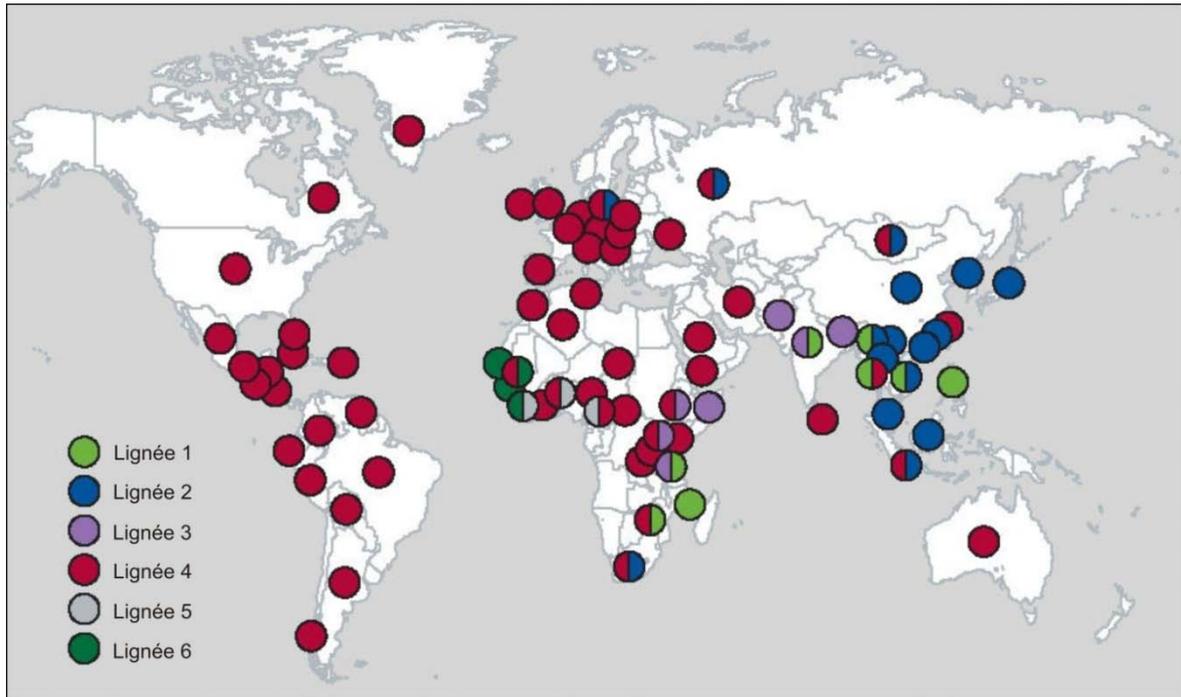


Figure 28 : Répartition géographique des 6 lignées phylogénétiques de *M. tuberculosis* et *M. africanum* décrites par Gagneux et al (2006).

(D'après Gagneux et Small, 2007).

V.3.2 Hypothèses émises sur la présence de la tuberculose en Yakoutie

L'apparition de la tuberculose humaine dans la population yakoute pourrait être liée à l'arrivée des Russes en Yakoutie au XVII^{ème} siècle. En effet, le contact entre les Russes et les Yakoutes à cette époque semble avoir été à l'origine d'importantes épidémies qui se sont succédées pendant plus de deux siècles (Crubézy et Alexeev, 2007). Ces épidémies ont fait beaucoup de ravages parmi les Yakoutes probablement en raison du terrain vierge qu'ils ont représenté. En effet, ils n'avaient jamais été en contact avec ces maladies infectieuses nouvelles et, de ce fait, n'étaient pas immunisés. Alors que les épidémies meurtrières telles que la variole ont marqué les esprits par leur brutalité et leur triste efficacité, il est possible que la tuberculose n'ait jamais été évoquée à cette époque en raison de son évolution beaucoup plus lente et donc nettement moins marquante.

V.3.3 Intérêts de l'analyse des SNP du génome d'agents pathogènes anciens

Cette étude a permis de confirmer le diagnostic de la tuberculose pour 3 des 7 échantillons analysés grâce à la détection d'une séquence d'ADN spécifique de l'élément répété IS6110, élément commun à tous les agents pathogènes de la tuberculose. Cependant, seule l'analyse des SNP a permis de trancher en faveur d'une tuberculose d'origine humaine et de déterminer qu'une souche de *M. tuberculosis* « moderne » était à l'origine de l'infection chez ces 3 sujets. Ces résultats illustrent clairement l'utilité des SNP dans les études de paléomicrobiologie. Ils permettent d'identifier directement une espèce pathogène lorsque les lésions observées ne sont pas spécifiques d'une maladie.

Par ailleurs, bien que notre étude ne soit pas la première à rapporter l'analyse de SNP pour étudier des infections tuberculeuses anciennes, elle démontre pour la première fois que le typage simultané de plusieurs SNP est possible à partir d'extraits d'ADNa de micro-organismes pathogènes. En effet, les articles parus dans la littérature ne font mention que de réactions individuelles pour l'analyse de chaque SNP.

VI. Conclusion et perspectives

Ce travail de thèse avait un objectif double : déterminer d'une part ce que pouvait apporter les SNP à l'étude des populations humaines du passé et établir d'autre part quelle méthode de typage de ces polymorphismes ponctuels se révélait la plus appropriée. Les résultats de nos travaux ont démontré que le miniséquençage à l'aide du kit SNaPshot™ (AB) est une méthode simple, fiable, robuste, sensible et particulièrement bien adaptée à l'analyse de SNP à partir de molécules d'ADN ancien puisqu'elle permet d'analyser plusieurs SNP simultanément. La méthode GenoSNIP en revanche n'a pas présenté une capacité de multiplexage suffisante pour l'étude de substrats anciens. Néanmoins, la détection d'analytes par spectrométrie de masse présentant des avantages indéniables au regard de l'électrophorèse capillaire, de nouveaux développements sont envisagés à partir d'une méthode récente de typage des SNP par spectrométrie de masse MALDI-TOF appelée iPLEX™ (Sequenom). Cette méthode est basée sur l'utilisation de la plateforme de spectrométrie de masse MassARRAY (Sequenom) qui a été développée spécifiquement pour l'analyse d'acides nucléiques. Cette méthode permettrait d'analyser jusqu'à 40 SNP simultanément et nécessiterait peu d'optimisations (<http://www.sequenom.com/>). Elle a déjà fait l'objet d'applications dans le domaine biomédical (Brion et al., 2008; Schaeffeler et al., 2008; Wright et al., 2008; Allegue et al., 2009; Korman et al., 2009; Young et al., 2009) et a été proposée pour l'analyse de SNP de la région codante de l'ADNmt dans les études de populations (Cerezo et al., 2009).

L'analyse de SNP du génome humain menée sur des échantillons anciens provenant du Sud de la Sibérie Centrale a permis de démontrer qu'aux Ages du Bronze et du Fer, cette région était peuplée de sujets à l'apparence physique européenne (yeux bleus ou verts, cheveux clairs et teint pâle) qui présentaient des lignées paternelles et maternelles majoritairement ouest-eurasiennes. Ces données génétiques semblent donc indiquer une origine européenne des sujets anciens étudiés, origine qui pourrait se situer en Europe de l'Est. Nos travaux ont donc permis de mettre en évidence l'existence de mouvements de populations d'Europe de l'Est jusqu'au Sud de la Sibérie avant 1800 avant J.-C. Il semblerait également qu'à l'Age du Fer ces migrants se soient unis à des populations autochtones. De plus, au vu de données linguistiques et archéologiques, il semblerait que le peuplement de la région de Krasnoyarsk puisse être lié à l'expansion vers l'Est des peuples de langue indo-européenne ou proto-indo-européenne. Enfin, les populations de la région de Krasnoyarsk pourraient avoir joué un rôle dans la colonisation du bassin du Tarim.

Bien entendu, il ne s'agit là que d'hypothèses qui méritent d'être approfondies. Cette étude va donc se poursuivre notamment par une analyse plus poussée des lignées R1a1 portées par les sujets masculins. En effet, deux foyers originels sont proposés pour cette lignée, l'un situé en Europe de l'Est, l'autre en Asie Centrale. Il va donc être intéressant de préciser l'origine des lignées paternelles observées chez les sujets sibériens anciens. Cela pourrait être réalisé grâce à l'analyse de deux SNP, M434 et M458, récemment décrits pour la différenciation des lignées R1a1 d'origine asiatique et européenne (Underhill et al., 2009).

Notre travail a également permis d'apporter de nouveaux éléments sur l'état sanitaire des populations yakoutes (Sibérie orientale) aux XVII^{ème} et XVIII^{ème} siècles. En effet, l'analyse de SNP du génome des agents pathogènes responsables de la tuberculose à partir de spécimens yakoutes anciens présentant des lésions osseuses évocatrices de cette maladie a permis de confirmer le diagnostic de la tuberculose pour 3 des 7 sujets anciens étudiés. Pour la première fois, cette étude a apporté la preuve que la tuberculose humaine sévissait en Yakoutie aux XVII^{ème} et XVIII^{ème} siècles. L'apparition de cette maladie chez les sujets anciens étudiés pourrait être liée à l'arrivée des colons russes en Yakoutie au XVII^{ème} siècle. De plus, nous avons pu déterminer que les souches de *M. tuberculosis* présentes en Yakoutie à cette époque appartiennent à la même lignée que celles qui sont probablement responsables de l'épidémie de tuberculose qui a dévasté toute l'Europe pendant les XVIII^{ème} et XIX^{ème} siècles (*M. tuberculosis* « moderne », plus précisément du groupe PGG2).

L'éventualité d'une épidémie de tuberculose aux XVII^{ème} et XVIII^{ème} siècles est également soulevée en Yakoutie. En effet, la proportion de sujets atteints présentant des lésions osseuses permet de se demander dans quelle mesure une grande partie de la population yakoute n'était pas atteinte à un moment donné. Cette hypothèse pourrait être testée puisque toutes les tombes yakoutes mises au jour lors des expéditions de la MAFSO ont fait l'objet de prélèvements biologiques (analysés au niveau génétique dans le cadre du travail de thèse de Sylvain Amory). Il serait donc envisageable de réaliser des analyses génétiques visant à détecter et identifier les agents pathogènes de la tuberculose à partir des échantillons pour lesquels des prélèvements, de la poudre d'os ou des extraits d'ADN sont encore disponibles. D'après la littérature, il est possible d'extraire et d'analyser l'ADN d'un agent pathogène de la tuberculose à partir d'échantillons osseux provenant d'un organisme infecté mais ne présentant aucune lésion osseuse apparente (Zink et al., 2005). Cela permettrait d'estimer la fréquence de la tuberculose dans la population yakoute ancienne considérée.

Par ailleurs, les analyses génétiques pourraient être étendues sur les 3 cas de tuberculose confirmés afin d'étudier l'évolution du génome de *M. tuberculosis*. Par exemple, il serait possible de déterminer quel sous-groupe ou SCG est à l'origine de ces cas infectieux. En effet, les isolats de *M. tuberculosis* « modernes » PGG2 peuvent être différenciés en SCG 3c, 3b, 4 ou 5 (Filliol et al., 2006). Selon une étude récente, le sous-groupe SCG 5 serait prévalent actuellement en Sibérie Centrale (Dubiley et al., 2009). Il serait donc intéressant de savoir s'il y a une continuité entre les souches qui circulaient il y a plus de 300 ans et actuellement en Sibérie. Enfin, d'autres SNP pourraient faire l'objet de nouvelles investigations à partir de ces échantillons anciens tels que ceux impliqués dans la résistance aux antituberculeux puisque nous disposons du génome d'au moins 3 isolats de *M. tuberculosis* n'ayant jamais été en contact avec ces médicaments.

La qualité des données qu'il a été possible d'obtenir ainsi que les hypothèses originales qui ont pu être proposées concernant l'origine et l'histoire du peuplement de la région de Krasnoyarsk et les nouveaux éléments qui ont pu être apportés quant à l'état sanitaire de la population yakoute confirment la pertinence de l'utilisation des SNP comme marqueurs génétiques pour étudier les populations du passé.

L'application de cette approche à des sujets anciens provenant de différentes régions permettra sans nul doute d'enrichir les connaissances sur l'histoire de la Sibérie et d'en préciser les modalités de peuplement. Les réactions multiplexes développées dans ce travail pour l'analyse des SNP du chromosome Y ont déjà été appliquées avec succès pour l'étude d'échantillons yakoutes anciens (cf. article Crubézy et al., 2010 reproduit dans la partie Publications).

La tuberculose n'est pas la seule maladie ancienne où les SNP peuvent être analysés. En effet, une étude récente de SNP à partir d'extraits d'ADNa de pathogènes de la lèpre a permis de démontrer que le génome de *M. leprae* est extrêmement stable au cours du temps (Watson et Lockwood, 2009). Il est donc possible d'envisager de développer une approche similaire pour d'autres pathogènes en fonction des lésions macroscopiques observées sur les squelettes anciens.

Outre les études de populations, ce travail ouvre de nouvelles perspectives dans le domaine médico-légal. En effet, le fait d'avoir démontré que des informations peuvent être obtenues sur les traits physiques d'un sujet ancien et son origine à partir d'échantillons osseux prouve qu'une telle application est possible à partir de prélèvements osseux pour des investigations

médico-légales. En France, la législation demeure très stricte et l'utilisation de tels tests génétiques prédictifs n'est pas envisageable dans un avenir proche mais nos travaux ont suscités l'intérêt de la communauté scientifique internationale et notamment des experts étrangers qui ont la possibilité d'inclure ces analyses dans leurs investigations génétiques. L'article publié en 2009 (cf. article Bouakaze et al., 2009 reproduit dans la partie Publications) a même fait l'objet d'une communication orale dans le cadre du 23^{ème} congrès de l'International Society of Forensic Genetics.

Le travail développé sur les échantillons bactériens contemporains ouvre également de nouvelles perspectives dans le domaine de la bactériologie clinique (cf. article Bouakaze et al., 2010 reproduit dans la partie Publications). En effet, si seuls 4 SNP ont été analysés sur les échantillons yakoutes anciens, deux séries de réactions multiplexes permettant d'analyser 8 SNP chacune ont été développées à partir de 56 échantillons d'ADN bactérien de différentes espèces du CMTB et 15 témoins négatifs. Nos résultats montrent que la première série de réactions permet d'identifier spécifiquement chacune des 7 espèces bactériennes du CMTB et de différencier les 3 groupes génétiques PGG. Cette série de réactions pourrait donc présenter un intérêt pour le diagnostic clinique de la tuberculose où l'identification de l'espèce impliquée dans l'infection est nécessaire. La seconde série de réactions multiplexes, qui permet d'identifier les 7 groupes phylogénétiquement distincts SCG de *M. tuberculosis* et *M. bovis*, pourrait servir à des fins phylogénétiques voire même épidémiologiques. En effet, il existe une concordance entre les familles de souches définies par les méthodes courantes d'épidémiologie moléculaire (*i.e.* IS6110-RFLP, typage MIRU-VNTR ou spoligotypage) et les SCG définis par notre méthode de typage de SNP (Gagneux et Small, 2007). Par exemple, les membres de la famille *Beijing* de *M. tuberculosis*, qui sont associés à une multirésistance à l'égard des médicaments antituberculeux et qui ont été identifiés comme étant responsables de plusieurs grandes épidémies, correspondent au groupe SCG 2. Ainsi, des développements à visée clinique sont également envisageables dans un avenir proche.

VII. Références bibliographiques

- Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, et Torroni A (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75:910-8.
- Akane A, Matsubara K, Nakamura H, Takahashi S, et Kimura K (1994) Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J Forensic Sci* 39:362-72.
- Alessandrini F, Turchi C, Onofri V, Buscemi L, Pesaresi M, et Tagliabracci A (2005) Multiplex PCR development of Y-chromosomal biallelic polymorphisms for forensic application. *J Forensic Sci* 50:519-25.
- Alland D, Lacher DW, Hazbon MH, Motiwala AS, Qi W, Fleischmann RD, et Whittam TS (2007) Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of *Mycobacterium tuberculosis* and the utility of LSPs in phylogenetic analysis. *J Clin Microbiol* 45:39-46.
- Allegue C, Sánchez-Diz P, Torres M, Abal R, Quintel I, Blanco-Verea A, Carracedo A, et Brion M (2009) Sequenom MassArray™ application in the long QT syndrome mutation detection. *FSI: Genet. Suppl. Ser.* 2:497-498.
- Alvarez-Iglesias V, Barros F, Carracedo A, et Salas A (2008) Minisequencing mitochondrial DNA pathogenic mutations. *BMC Med Genet* 9:26.
- Ammerman AJ, Pinhasi R, et Banffy E (2006) Comment on "Ancient DNA from the first European farmers in 7500-year-old Neolithic sites". *Science* 312:1875; author reply 1875.
- Amory S (2007) La diffusion des gènes de la période protohistorique à l'époque actuelle dans le complexe spatial Altaï-Baïkal. Thèse de doctorat : Archéologie, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, et Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-65.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Arnold C, Westland L, Mowat G, Underwood A, Magee J, et Gharbia S (2005) Single-nucleotide polymorphism-based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clin Microbiol Infect* 11:122-30.
- Austin JJ, Ross AJ, Smith AB, Fortey RA, et Thomas RH (1997) Problems of reproducibility--does geologically ancient DNA survive in amber-preserved insects? *Proc Biol Sci* 264:467-74.
- Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, et Bustamante CD (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19:795-803.
- Baker L, Brown T, Maiden MC, et Drobniowski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg Infect Dis* 10:1568-77.

- Bandelt HJ, Kong QP, Parson W, et Salas A (2005) More evidence for non-maternal inheritance of mitochondrial DNA? *J Med Genet* 42:957-60.
- Barbujani G, et Chikhi L (2006) Population genetics: DNAs from the European Neolithic. *Heredity* 97:84-5.
- Barnes PF, et Cave MD (2003) Molecular epidemiology of tuberculosis. *N. Engl. J. Med.* 349:1149-1156.
- Beauval C, Maureille B, Lacrampe-Cuyaubere F, Serre D, Peressinotto D, Bordes JG, Cochard D, Couchoud I, Dubrasquet D, Laroulandie V, Lenoble A, Mallye JB, Pasty S, Primault J, Rohland N, Paabo S, et Trinkaus E (2005) A late Neandertal femur from Les Rochers-de-Villeneuve, France. *Proc Natl Acad Sci U S A* 102:7085-90.
- Bennett DC, et Lamoreux ML (2003) The color loci of mice--a genetic century. *Pigment Cell Res* 16:333-44.
- Brandstatter A, Parsons TJ, et Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117:291-8.
- Brandstatter A, Salas A, Niederstatter H, Gassner C, Carracedo A, et Parson W (2006) Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27:2541-50.
- Branicki W, Brudnik U, Kupiec T, Wolanska-Nowak P, et Wojas-Pelc A (2007) Determination of phenotype associated SNPs in the MC1R gene. *J Forensic Sci* 52:349-54.
- Branicki W, Brudnik U, et Wojas-Pelc A (2009) Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann Hum Genet* 73:160-70.
- Braun M, Cook DC, et Pfeiffer S (1998) DNA from Mycobacterium tuberculosis complex identified in North American, pre-Columbian human skeletal remains. *J Archeol Sci* 25:271-277.
- Bray MS, Boerwinkle E, et Doris PA (2001) High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat* 17:296-304.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev VB, Golovanova LV, de la Rasilla M, Fortea J, Rosas A, et Paabo S (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318-21.
- Brinkmann B, Klintschar M, Neuhuber F, Huhne J, et Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62:1408-15.
- Brion M, Allegue C, Monserrat L, Hermida M, Castro-Beiras A, et Carracedo A (2008) Large scale analysis of HCM mutations in sudden cardiac death. *FSI: Genet. Suppl. Ser.* 1:549-550.
- Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, et Morling N (2005a) Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26:4411-20.
- Brion M, Sobrino B, Blanco-Verea A, Lareu MV, et Carracedo A (2005b) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int J Legal Med* 119:10-5.
- Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen

- D, et Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A* 99:3684-9.
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, et Cooper A (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717-28.
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Gutierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Martin C, Mokrousov I, Narvskaia O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofoa-Razanamparany V, Rasolonalana T, Rossetti ML, Rusch-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, et Sola C (2006) Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 6:23.
- Budowle B, et van Daal A (2008) Forensically relevant SNP classes. *Biotechniques* 44:603-8, 610.
- Burger J, Kirchner M, Bramanti B, Haak W, et Thomas MG (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A* 104:3736-41.
- Cano RJ, Poinar HN, Pieniazek NJ, Acra A, et Poinar GO, Jr. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* 363:536-8.
- Cerezo M, Cerny V, Carracedo A, et Salas A (2009) Applications of MALDI-TOF MS to large-scale human mtDNA population-based studies. *Electrophoresis* 30:3665-73.
- Chimara E, Ferrazoli L, et Leao SC (2004) Mycobacterium tuberculosis complex differentiation using gyrB-restriction fragment length polymorphism analysis. *Mem Inst Oswaldo Cruz* 99:745-8.
- Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstatter H, Berger C, Berger B, Falsetti AB, Gill P, Parson W, et Finelli LN (2009) Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* 4:e4838.
- Comas I, Homolka S, Niemann S, et Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in mycobacterium tuberculosis highlights the limitations of current methodologies. *PLoS One* 4:e7815.
- Cooper A, Mourer-Chauvire C, Chambers GK, von Haeseler A, Wilson AC, et Paabo S (1992) Independent origins of New Zealand moas and kiwis. *Proc Natl Acad Sci U S A* 89:8741-4.
- Cooper A, et Poinar HN (2000) Ancient DNA: do it right or not at all. *Science* 289:1139.
- Crubézy E, et Alexeev A (2007) Chamane - Kyys, jeune fille des glaces.
- Crubézy E, Ludes B, Poveda JD, Clayton J, Crouau-Roy B, et Montagnon D (1998) Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III* 321:941-51.
- Csanyi B, Bogacsi-Szabo E, Tomory G, Czibula A, Priskin K, Csosz A, Mende B, Lango P, Csete K, Zsolnai A, Conant EK, Downes CS, et Rasko I (2008) Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. *Ann Hum Genet* 72:519-34.

- Dai Z, Papp AC, Wang D, Hampel H, et Sadee W (2008) Genotyping panel for assessing response to cancer chemotherapy. *BMC Med Genomics* 1:24.
- Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, et Wallace DC (2002) Analysis of mitochondrial DNA diversity in the aleuts of the commander islands and its implications for the genetic history of beringia. *Am J Hum Genet* 71:415-21.
- Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I, Dorzhu C, Luzina F, Miscicka-Sliwka D, et Zakharov I (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet* 118:591-604.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, et Zakharov IA (2003) Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67:391-411.
- Djelouadji Z, Raoult D, Daffe M, et Drancourt M (2008) A Single-Step Sequencing Method for the Identification of Mycobacterium tuberculosis Complex Species. *PLoS Negl Trop Dis* 2:e253.
- Donoghue HD (2009) Human tuberculosis - an ancient disease, as elucidated by ancient microbial biomolecules. *Microbes Infect*.
- Donoghue HD, Hershkovitz I, Minnikin DE, Besrac GS, Leec OY-C, Galilid E, Greenblatt CL, Lemma E, Spigelman M, et Bar-Gal GK (2009a) Biomolecular archaeology of ancient tuberculosis: response to "Deficiencies and challenges in the study of ancient tuberculosis DNA" by Wilbur et al. (2009). *J Archeol Sci* 36:2797-2804
- Donoghue HD, Lee OY, Minnikin DE, Besra GS, Taylor JH, et Spigelman M (2009b) Tuberculosis in Dr Granville's mummy: a molecular re-examination of the earliest known Egyptian mummy to be scientifically examined and given a medical diagnosis. *Proc Biol Sci*.
- Donoghue HD, et Spigelman M (2006) Comment. Pathogenic microbial ancient DNA: a problem or an opportunity? *Proc Biol Sci* 273:641-2; discussion 643.
- Donoghue HD, Spigelman M, Greenblatt CL, Lev-Maor G, Bar-Gal GK, Matheson C, Vernon K, Nerlich AG, et Zink AR (2004) Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. *Lancet Infect Dis* 4:584-92.
- Drancourt M, Aboudharam G, Signoli M, Dutour O, et Raoult D (1998) Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proc Natl Acad Sci U S A* 95:12637-40.
- Drancourt M, et Raoult D (2005) Palaeomicrobiology: current issues and perspectives. *Nat Rev Microbiol* 3:23-35.
- Dubiley S, Ignatova A, Mukhina T, Nizova A, Blagodatskikh S, Stepanshina V, et Shemyakin I (2009) Molecular Epidemiology of Tuberculosis in Tula Area, Central Russia, before Dots. *Clin Microbiol Infect*.
- Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, et Sturm RA (2007) A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet* 80:241-52.
- Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, et Hansen L (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 123:177-87.
- Eisenach KD, Cave MD, Bates JH, et Crawford JT (1990) Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. *J Infect Dis* 161:977-81.

- Endicott P, Gilbert MT, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, et Cooper A (2003) The genetic origins of the Andaman Islanders. *Am J Hum Genet* 72:178-84.
- Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, et Sanchez JJ (2006) Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations. *PLoS One* 1:e81.
- Endicott P, Sanchez JJ, Pichler I, Brotherton P, Brooks J, Egarter-Vigl E, Cooper A, et Pramstaller P (2009) Genotyping human ancient mtDNA control and coding region polymorphisms with a multiplexed Single-Base-Extension assay: the singular maternal history of the Tyrolean Iceman. *BMC Genet* 10:29.
- Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, Soares P, Luciani S, Marota I, De Bellis G, Richards MB, et Rollo F (2008) Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18:1687-93.
- Faerman M, et Bar-Gal GK (1998) Determining the sex of infanticide victims from the late roman era through ancient DNA analysis. *J Arc Sci* 25:8621-8625.
- Fechner A, Quinque D, Rychkov S, Morozowa I, Naumova O, Schneider Y, Willuweit S, Zhukova O, Roewer L, Stoneking M, et Nasidze I (2008) Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia. *Am J Phys Anthropol* 137:41-7.
- Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, et Alland D (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* 188:759-72.
- Fletcher HA, Donoghue HD, Holton J, Pap I, et Spigelman M (2003a) Widespread occurrence of *Mycobacterium tuberculosis* DNA from 18th-19th century Hungarians. *Am J Phys Anthropol* 120:144-52.
- Fletcher HA, Donoghue HD, Taylor GM, van der Zanden AG, et Spigelman M (2003b) Molecular analysis of *Mycobacterium tuberculosis* DNA from a family of 18th century Hungarians. *Microbiology* 149:143-51.
- Frothingham R, Strickland PL, Bretzel G, Ramaswamy S, Musser JM, et Williams DL (1999) Phenotypic and genotypic characterization of *Mycobacterium africanum* isolates from West Africa. *J Clin Microbiol* 37:1921-6.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, et Small PM (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 103:2869-73.
- Gagneux S, et Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7:328-37.
- Gao S, Cui Y, Yang Y, Duan R, Abuduresule I, Mair VH, Zhu H, et Zhou H (2008) Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China. *Sci China C Life Sci* 51:205-13.
- Gibbons A (1994) Possible dino DNA find is greeted with skepticism. *Science* 266:1159.
- Gibbons A (2008) American Association of Physical Anthropologists meeting. Tuberculosis jumped from humans to cows, not vice versa. *Science* 320:608.
- Gilbert MT, Bandelt HJ, Hofreiter M, et Barnes I (2005) Assessing ancient DNA studies. *Trends Ecol Evol* 20:541-4.

- Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, et Cooper A (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet* 72:48-61.
- Gilbert MT, et Willerslev E (2007) Rescuing ancient DNA. *Nat Biotechnol* 25:872-4.
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, et Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130-5.
- Gimbutas M (1970) Proto-Indo-European culture: the Kourgane culture during the fifth, fourth and third millennia BC. In G Cardona, Hoenigswald, H.M., Seen, A.M. (ed.): *Indo-European and Indo-Europeans*. Philadelphia: University of Pennsylvania Press, pp. 155–195.
- Goh KS, Legrand E, Sola C, et Rastogi N (2001) Rapid differentiation of "Mycobacterium canettii" from other Mycobacterium tuberculosis complex organisms by PCR-restriction analysis of the hsp65 gene. *J Clin Microbiol* 39:3705-8.
- Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, et Zurawski G (1990) Chloroplast DNA sequence from a miocene Magnolia species. *Nature* 344:656-8.
- Graf J, Hodgson R, et van Daal A (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum Mutat* 25:278-84.
- Green RE, Malaspina AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, Prüfer K, Siebauer M, Burbano HA, Ronan M, Rothberg JM, Egholm M, Rudan P, Brajkovic D, Kucan Z, Gusic I, Wikstrom M, Laakkonen L, Kelso J, Slatkin M, et Paabo S (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134:416-26.
- Greenblatt C, et Spigelman M (2003) Emerging pathogens. *Archaeology, ecology and evolution of infectious disease*.
- Grignani P, Peloso G, Achilli A, Turchi C, Tagliabracci A, Alu M, Beduschi G, Ricci U, Giunti L, Robino C, Gino S, et Previdere C (2006) Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. *Int J Legal Med* 120:151-6.
- Groupe de travail du conseil supérieur d'hygiène publique de France (2004) Diagnostic clinique et bactériologique de la tuberculose. *Médecine et Maladies Infectieuses*. 34:364-370.
- Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, et Musser JM (2006) Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J Infect Dis* 193:121-8.
- Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, et Musser JM (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 162:1533-43.
- Gutierrez G, et Marin A (1998) The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Mol Biol Evol* 15:926-9.
- Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, et Vincent V (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog* 1:e5.
- Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AW, Meller H, et Alt KW (2008) Ancient DNA, Strontium isotopes, and osteological

- analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci U S A* 105:18226-31.
- Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, et Burger J (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310:1016-8.
- Haas CJ, Zink A, Palfi G, Szeimies U, et Nerlich AG (2000) Detection of leprosy in ancient human skeletal remains by molecular identification of *Mycobacterium leprae*. *Am J Clin Pathol* 114:428-36.
- Hagelberg E, Sykes B, et Hedges R (1989) Ancient bone DNA amplified. *Nature* 342:485.
- Halder I, Shriver M, Thomas M, Fernandez JR, et Frudakis T (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat* 29:648-58.
- Hanni C, Laudet V, Sakka M, Begue A, et Stehelin D (1990) [Amplification of mitochondrial DNA fragments from ancient human teeth and bones]. *C R Acad Sci III* 310:365-70.
- Hemphill BE, et Mallory JP (2004) Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am J Phys Anthropol* 124:199-222.
- Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, et Gagneux S (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 6:e311.
- Hershkovitz I, Donoghue HD, Minnikin DE, Besra GS, Lee OY, Gernaey AM, Galili E, Eshed V, Greenblatt CL, Lemma E, Bar-Gal GK, et Spigelman M (2008) Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* 3:e3426.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, et Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-4.
- Hirsh A, Tsolaki A, DeRiemer K, Feldman M, et Small P (2004) Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A*. 101.
- Ho SY, et Gilbert MT ((2009)) Ancient mitogenomics. *Mitochondrion* 10:1-11.
- Hodgson JA, et Disotell TR (2008) No evidence of a Neanderthal contribution to modern human diversity. *Genome Biol* 9:206.
- Hofreiter M, Serre D, Poinar HN, Kuch M, et Paabo S (2001) Ancient DNA. *Nat Rev Genet* 2:353-9.
- Huard RC, Fabre M, de Haas P, Lazzarini LC, van Soolingen D, Cousins D, et Ho JL (2006) Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol* 188:4271-87.
- Jehaes E, Pfeiffer H, Toprak K, Decorte R, Brinkmann B, et Cassiman JJ (2001) Mitochondrial DNA analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette. *Eur J Hum Genet* 9:185-90.
- Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci Int* 118:158-62.
- Jobling MA, et Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, et van Embden J (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35:907-14.

- Kappelman J, Alcicek MC, Kazanci N, Schultz M, Ozkul M, et Sen S (2008) First Homo erectus from Turkey and implications for migrations into temperate Eurasia. *Am J Phys Anthropol* 135:110-6.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, et Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830-8.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, et Hammer MF (2002) High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74:761-89.
- Kasai H, Ezaki T, et Harayama S (2000) Differentiation of phylogenetically related slowly growing mycobacteria by their gyrB sequences. *J Clin Microbiol* 38:301-8.
- Kayser M, et Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* 3:154-61.
- Keyser-Tracqui C, Crubezy E, et Ludes B (2003) Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *Am J Hum Genet* 73:247-60.
- Keyser-Tracqui C, et Ludes B (2005) Methods for the study of ancient DNA. *Methods Mol Biol* 297:253-64.
- Kharkov VN, Stepanov VA, Borinskaya SA, Kozhekbaeva ZM, Gusar VA, Grechanina EY, Puzyrev VP, Khusnutdinova EK, et Yankovsky NK (2004) Gene pool structure of eastern Ukrainians as inferred from the Y-chromosome haplogroups. *Russ J Genet* 40:415-421.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, et Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19:1737-51.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, et Zhang YP (2003) Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73:671-6.
- Konomi N, Lebowohl E, Mowbray K, Tattersall I, et Zhang D (2002) Detection of mycobacterial DNA in Andean mummies. *J Clin Microbiol* 40:4738-40.
- Korman BD, Seldin MF, Taylor KE, Le JM, Lee AT, Plenge RM, Amos CI, Criswell LA, Gregersen PK, Kastner DL, et Remmers EF (2009) The chromosome 7q region association with rheumatoid arthritis in females in a British population is not replicated in a North American case-control series. *Arthritis Rheum* 60:47-52.
- Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hanni C, Fortea J, de la Rasilla M, Bertranpetit J, Rosas A, et Paabo S (2007a) The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr Biol* 17:1908-12.
- Krause J, Orlando L, Serre D, Viola B, Prufer K, Richards MP, Hublin JJ, Hanni C, Derevianko AP, et Paabo S (2007b) Neanderthals in central Asia and Siberia. *Nature* 449:902-4.
- Kreader CA (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl Environ Microbiol* 62:1102-6.
- Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M, et Paabo S (2000) A view of Neandertal genetic diversity. *Nat Genet* 26:144-6.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, et Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30.

- Kristensen R, Berdal KG, et Holst-Jensen A (2007) Simultaneous detection and identification of trichothecene- and moniliformin-producing *Fusarium* species based on multiplex SNP analysis. *J Appl Microbiol* 102:1071-81.
- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, et Rosas A (2009) Bitter taste perception in Neanderthals through the analysis of the TAS2R38 gene. *Biol Lett* 5:809-11.
- Lalueza-Fox C, Krause J, Caramelli D, Catalano G, Milani L, Sampietro ML, Calafell F, Martinez-Maza C, Bastir M, Garcia-Taberner A, de la Rasilla M, Fortea J, Paabo S, Bertranpetit J, et Rosas A (2006) Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr Biol* 16:R629-30.
- Lalueza-Fox C, Rompler H, Caramelli D, Staubert C, Catalano G, Hughes D, Rohland N, Pilli E, Longo L, Condemi S, de la Rasilla M, Fortea J, Rosas A, Stoneking M, Schoneberg T, Bertranpetit J, et Hofreiter M (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453-5.
- Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martinez-Maza C, Bastir M, Fortea J, de la Rasilla M, Bertranpetit J, et Rosas A (2005) Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula. *Mol Biol Evol* 22:1077-81.
- Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, et Cheng KC (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310:1782-6.
- Lamberg-Karlovsky CC (2002) Archaeology and Language. The Indo-Iraniens. *Curr Anthropol* 43:63-75.
- Larcombe LA, Nickerson P, Hoppa RD, et Matheson C (2005) Detection of a single nucleotide polymorphism in the IL-6 promoter region of ancient nuclear DNA. *Infect Genet Evol* 5:117-22.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, et Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192-206.
- Lessig R, Zoledziewska M, Fahr K, Edelmann J, Kostrzewa M, Dobosz T, et Kleemann WJ (2005) Y-SNP-genotyping - a new approach in forensic analysis. *Forensic Sci Int* 154:128-36.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, et Venter JC (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
- Li H, Huang Y, Mustavich LF, Zhang F, Tan JZ, Wang LE, Qian J, Gao MH, et Jin L (2007) Y chromosomes of prehistoric people along the Yangtze River. *Hum Genet* 122:383-8.
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362:709-15.
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Lavolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, et Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24:381-6.
- Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L,

- Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventre A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Golge M, Mikerezi I, Usanga E, et Villems R (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21:2012-21.
- Lopez Herraes D, Bauchet M, Tang K, Theunert C, Pugach I, Li J, Nandineni MR, Gross A, Scholz M, et Stoneking M (2009) Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One* 4:e7888.
- Mathema B, Kurepina NE, Bifani PJ, et Kreiswirth BN (2006) Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 19:658-85.
- Mays S, et Taylor GM (2003) A first prehistoric case of tuberculosis from Britain. *Int. J. Osteoarchaeol.* 13:189-196.
- Mays S, Taylor GM, Legge AJ, Young DB, et Turner-Walker G (2001) Paleopathological and biomolecular study of tuberculosis in a medieval skeletal collection from England. *Am J Phys Anthropol* 114:298-311.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356-69.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, et Morling N (2004) MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome single-nucleotide polymorphisms. *Anal Chem* 76:6039-45.
- Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, et Herrera RJ (2009) Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur J Hum Genet* 17:1260-73.
- Moiseyev VG (2006) Nonmetric traits in Early Iron Age cranial series from Western and Southern Siberia. *Archaeol. Ethnol. Anthropol. Eurasia* 25:145-152.
- Mooder KP, Schurr TG, Bamforth FJ, Bazaliiski VI, et Savel'ev NA (2006) Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. *Am J Phys Anthropol* 129:349-61.
- Mosquera-Miguel A, Alvarez-Iglesias V, Cerezo M, Lareu MV, Carracedo A, et Salas A (2009) Testing the performance of mtSNP minisequencing in forensic samples. *Forensic Sci Int Genet* 3:261-4.
- Mostowy S, Cousins D, Brinkman J, Aranaz A, et Behr MA (2002) Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. *J Infect Dis* 186:74-80.
- Mullis KB, et Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-50.
- Murphy EM, Chistov YK, Hopkins R, Rutland P, et Taylor GM (2009) Tuberculosis among Iron Age individuals from Tyva, South Siberia: palaeopathological and biomolecular findings. *J. Archeol. Sci.* 36:2029-2038
- Myles S, Somel M, Tang K, Kelso J, et Stoneking M (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum Genet* 120:613-21.
- Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, et Seldin MF (2009) An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC Genet* 10:39.
- Niemann S, Harmsen D, Rusch-Gerdes S, et Richter E (2000) Differentiation of clinical Mycobacterium tuberculosis complex isolates by gyrB DNA sequence polymorphism analysis. *J Clin Microbiol* 38:3231-4.

- Noro M, Masuda R, Dubrovo IA, Yoshida MC, et Kato M (1998) Molecular phylogenetic inference of the woolly mammoth *Mammuthus primigenius*, based on complete sequences of mitochondrial cytochrome b and 12S ribosomal RNA genes. *J Mol Evol* 46:314-26.
- O'Rourke DH, Hayes MG, et Carlyle SW (2000) Ancient DNA studies in physical anthropology. *Annu Rev Anthropol* 29:217-242.
- Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, et Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int* 157:23-35.
- Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, et Hanni C (2006) Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr Biol* 16:R400-2.
- Ortner DJ (2003) Identification of Pathological Conditions in Human Skeletal Remains.: Academic Press.
- Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, et Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-3.
- Paabo S (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314:644-5.
- Paabo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A* 86:1939-43.
- Paabo S, Higuchi RG, et Wilson AC (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J Biol Chem* 264:9709-12.
- Paabo S, Kampe O, Severinsson L, Andersson M, Fernandez C, et Peterson PA (1985) The association between class-I transplantation antigens and an adenovirus membrane protein. *Prog Allergy* 36:114-34.
- Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, et Hofreiter M (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-79.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, et Stoneking M (2006) Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120:334-53.
- Pakendorf B, et Stoneking M (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165-83.
- Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, et Kidd KK (2009) SNPs for a universal individual identification panel. *Hum Genet*.
- Paracchini S, Arredi B, Chalk R, et Tyler-Smith C (2002) Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res* 30:e27.
- Parra EJ (2007) Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am J Phys Anthropol Suppl* 45:85-105.
- Parsons LM, Brosch R, Cole ST, Somoskovi A, Loder A, Bretzel G, Van Soolingen D, Hale YM, et Salfinger M (2002) Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. *J Clin Microbiol* 40:2339-45.
- Passarino G, Semino O, Magri C, Al-Zahery N, Benuzzi G, Quintana-Murci L, Andellnovic S, Bulc-Jakus F, Liu A, Arslan A, et Santachiara-Benerecetti AS (2001) The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. *Hum Immunol* 62:922-32.
- Petkovski E (2006) Polymorphismes ponctuels de séquence et identification génétique. Etude par spectrométrie de masse MALDI-TOF. Thèse de doctorat : Sciences du Vivant -

- Aspects moléculaires et cellulaires de la biologie, Université Louis Pasteur, Strasbourg.
- Petkovski E, Keyser-Tracqui C, Crubézy E, Hienne R, et Ludes B (2006) MALDI-TOF MS analysis of Y-SNPs in ancient samples. *International Congress Series* 1288:25-27
- Petkovski E, Keyser-Tracqui C, Hienne R, et Ludes B (2005) SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. *J Forensic Sci* 50:535-41.
- Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, et Paabo S (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281:402-6.
- Poinar HN, Hoss M, Bada JL, et Paabo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* 272:864-6.
- Pritchard JK, Stephens M, et Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-59.
- Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, et Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140:251-7.
- Ragoussis J (2009) Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet* 10:117-33.
- Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L, Menzies D, Schwartzman K, et Behr MA (2009) Major Mycobacterium tuberculosis lineages associate with patient country of origin. *J Clin Microbiol* 47:1119-28.
- Robert J, Morvan VL, Smith D, Pourquier P, et Bonnet J (2005) Predicting drug response and toxicity based on gene polymorphisms. *Crit Rev Oncol Hematol* 54:171-96.
- Roberts C, et Ingham S (2008) Using ancient DNA analysis in palaeopathology: a critical analysis of published papers, with recommendations for future work. *Int. J. Osteoarchaeol.* 18:600–613.
- Roberts CA, Pfister LA, et Mays S (2009) Letter to the editor: was tuberculosis present in *Homo erectus* in Turkey? *Am J Phys Anthropol* 139:442-4.
- Rogaev EI, Grigorenko AP, Moliaka YK, Faskhutdinova G, Goltsov A, Lahti A, Hildebrandt C, Kittler EL, et Morozova I (2009) Genomic identification in the historical case of the Nicholas II royal family. *Proc Natl Acad Sci U S A* 106:5258-63.
- Rompler H, Dear PH, Krause J, Meyer M, Rohland N, Schoneberg T, Spriggs H, Stiller M, et Hofreiter M (2006) Multiplex amplification of ancient DNA. *Nat Protoc* 1:720-8.
- Rosser ZH, Zerjal T, Hurler ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, et Jobling MA (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526-43.
- Rothschild BM, Martin LD, Lev G, Bercovier H, Bar-Gal GK, Greenblatt C, Donoghue H, Spigelman M, et Brittain D (2001) Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis* 33:305-11.

- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-4.
- Salo WL, Aufderheide AC, Buikstra J, et Holcomb TA (1994) Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A* 91:2091-4.
- Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Marti M, Bertranpetit J, et Lalueza-Fox C (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 274:2161-7.
- Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, et Morling N (2003) Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137:74-84.
- Sanchez JJ, Brion M, Parson W, Blanco-Verea AJ, Borsting C, Lareu M, Niederstatter H, Oberacher H, Morling N, et Carracedo A (2004) Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications. *Forensic Sci Int* 140:241-50.
- Sanchez JJ, et Endicott P (2006) Developing multiplexed SNP assays with special reference to degraded DNA templates. *Nat Protoc* 1:1370-8.
- Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, et Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713-24.
- Sauer S, Lehrach H, et Reinhardt R (2003) MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging. *Nucleic Acids Res* 31:e63.
- Schaeffeler E, Zanger UM, Eichelbaum M, Asante-Poku S, Shin JG, et Schwab M (2008) Highly multiplexed genotyping of thiopurine s-methyltransferase variants using MALD-TOF mass spectrometry: reliable genotyping in different ethnic groups. *Clin Chem* 54:1637-47.
- Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki H, Paabo S, et Smith FH (2002) The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc Natl Acad Sci U S A* 99:13342-7.
- Scholz M, Giddings I, et Pusch CM (1998) A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. *Anal Biochem* 259:283-6.
- Scorpio A, et Zhang Y (1996) Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* 2:662-7.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, et Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155-9.
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Menecier P, Hofreiter M, Possnert G, et Paabo S (2004) No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol* 2:E57.
- Sharma S, Rai E, Sharma P, Jena M, Singh S, Darvishi K, Bhat AK, Bhanwer AJ, Tiwari PK, et Bamezai RN (2009) The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. *J Hum Genet* 54:47-55.

- Smith NH, Gordon SV, de la Rúa-Domenech R, Clifton-Hadley RS, et Hewinson RG (2006) Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat Rev Microbiol* 4:670-81.
- Smith NH, Hewinson RG, Kremer K, Brosch R, et Gordon SV (2009) Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 7:537-44.
- Sobrino B, Brion M, et Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181-94.
- Soejima M, et Koda Y (2007) Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *Int J Legal Med* 121:36-9.
- Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, et Rastogi N (2003) Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol* 3:125-33.
- Soltis PS, Soltis DE, et Smiley CJ (1992) An *rbcL* sequence from a Miocene *Taxodium* (bald cypress). *Proc Natl Acad Sci U S A* 89:449-51.
- Sreevatsan S, Escalante P, Pan X, Gillies DA, 2nd, Siddiqui S, Khalaf CN, Kreiswirth BN, Bifani P, Adams LG, Ficht T, Perumaalla VS, Cave MD, van Embden JD, et Musser JM (1996) Identification of a polymorphic nucleotide in *oxyR* specific for *Mycobacterium bovis*. *J Clin Microbiol* 34:2007-10.
- Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et Musser JM (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 94:9869-74.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, et Wallace DC (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69:67-89.
- Sturm RA (2009) Molecular genetics of human pigmentation diversity. *Hum Mol Genet* 18:R9-17.
- Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, Martin NG, et Montgomery GW (2008) A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am J Hum Genet* 82:424-31.
- Sullivan KM, Mannucci A, Kimpton CP, et Gill P (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* 15:636-8, 640-1.
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, et van Soolingen D (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 44:4498-510.
- Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, et Fanning TG (2005) Characterization of the 1918 influenza virus polymerase genes. *Nature* 437:889-93.
- Taylor GM, Crossey M, Saldanha J, et Waldron T (1996) DNA from *Mycobacterium tuberculosis* identified in mediaeval human skeletal remains using polymerase chain reaction. *J. Archaeol. Sci.* 23:789-798.

- Taylor GM, Goyal M, Legge AJ, Shaw RJ, et Young D (1999) Genotypic analysis of *Mycobacterium tuberculosis* from medieval human remains. *Microbiology* 145 (Pt 4):899-904.
- Taylor GM, Mays SA, et Huggett JF (2009) Ancient DNA (aDNA) studies of man and microbes: general similarities, specific differences. *Int. J. Osteoarchaeol.* DOI: 10.1002/oa.966.
- Taylor GM, Murphy E, Hopkins R, Rutland P, et Chistov Y (2007) First report of *Mycobacterium bovis* DNA in human remains from the Iron Age. *Microbiology* 153:1243-9.
- Taylor GM, Young DB, et Mays SA (2005) Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. *J Clin Microbiol* 43:2236-40.
- The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299-320.
- The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-61.
- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-78.
- The Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339-48.
- Toothman MH, Kester KM, Champagne J, Cruz TD, Street WSt, et Brown BL (2008) Characterization of human DNA in environmental samples. *Forensic Sci Int* 178:7-15.
- Tost J, et Gut IG (2002) Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom Rev* 21:388-418.
- Tost J, et Gut IG (2005) Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin Biochem* 38:335-50.
- Tuross N (1994) The biochemistry of ancient DNA in bone. *Experientia* 50:530-5.
- Uchihi R, Tamaki K, Kojima T, Yamamoto T, et Katsumata Y (1992) Deoxyribonucleic acid (DNA) typing of human leukocyte antigen (HLA)-DQA1 from single hairs in Japanese. *J Forensic Sci* 37:853-9.
- Underhill PA, et Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41:539-64.
- Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CE, Semino O, Battaglia V, Kutuev I, Jarve M, Chaubey G, Ayub Q, Mohyuddin A, Mehdi SQ, Sengupta S, Rogaev EI, Khusnutdinova EK, Pshenichnov A, Balanovsky O, Balanovska E, Jeran N, Augustin DH, Baldovic M, Herrera RJ, Thangaraj K, Singh V, Singh L, Majumder P, Rudan P, Primorac D, VILLEMS R, et Kivisild T (2009) Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet.*
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, et Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43-62.
- Vallone PM, Just RS, Coble MD, Butler JM, et Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med* 118:147-57.
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, et et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31:406-9.

- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. *Science* 291:1304-51.
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, et Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-82.
- Watson CL, et Lockwood DN (2009) Single nucleotide polymorphism analysis of European archaeological *M. leprae* DNA. *PLoS One* 4:e7547.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, et Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* 98:10244-9.
- Wenzel T, Elssner T, Fahr K, Bimmler J, Richter S, Thomas I, et Kostrzewa M (2003) Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides. *Nucleosides Nucleotides Nucleic Acids* 22:1579-81.
- Wilbur AK, Bouwman AS, Stone AC, Roberts CA, Pfister L-A, Buikstra JE, et Brown TA (2009) Deficiencies and challenges in the study of ancient tuberculosis DNA. *J Archeol Sci* 36:1990-1997.
- Willerslev E, et Cooper A (2005) Ancient DNA. *Proc Biol Sci* 272:3-16.
- Willerslev E, Hansen AJ, et Poinar HN (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 19:141-7.
- Wirth T, Hildebrand F, Allix-Beguec C, Wolbeling F, Kubica T, Kremer K, van Soolingen D, Rusch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, et Niemann S (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog* 4:e1000160.
- Wolinsky H (2008) Our history, our genes. Population genetics lets researchers look back in time at human migrations. *EMBO Rep.* 9:127-129.
- Woodward SR, Weyand NJ, et Bunnell M (1994) DNA sequence from Cretaceous period bone fragments. *Science* 266:1229-32.
- Wright WT, Heggarty SV, Young IS, Nicholls DP, Whittall R, Humphries SE, et Graham CA (2008) Multiplex MassARRAY spectrometry (iPLEX) produces a fast and economical test for 56 familial hypercholesterolaemia-causing mutations. *Clin Genet* 74:463-8.

- Yao YG, Kong QP, Bandelt HJ, Kivisild T, et Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635-51.
- Young RP, Hopkins RJ, Hay BA, Epton MJ, Mills GD, Black PN, Gardner HD, Sullivan R, et Gamble GD (2009) Lung cancer susceptibility model based on age, family history and genetic variants. *PLoS One* 4:e5302.
- Yuasa I, Umetsu K, Harihara S, Kido A, Miyoshi A, Saitou N, Dashnyam B, Jin F, Lucotte G, Chattopadhyay PK, Henke L, et Henke J (2007) Distribution of two Asian-related coding SNPs in the MC1R and OCA2 genes. *Biochem Genet* 45:535-42.
- Zegura SL, Karafet TM, Zhivotovsky LA, et Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21:164-75.
- Zerjal T, Pandya A, et Santos FR (1999) The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. . In DR Papiha SS, Chakraborty R et al (ed.): *Genomic diversity: applications in human population genetics*. . New York: Plenum, pp. 91–101.
- Zink A, Haas CJ, Reischl U, Szeimies U, et Nerlich AG (2001) Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol* 50:355-66.
- Zink AR, Grabner W, et Nerlich AG (2005) Molecular identification of human tuberculosis in recent and historic bone tissue samples: The role of molecular techniques for the study of historic tuberculosis. *Am J Phys Anthropol* 126:32-47.
- Zink AR, Grabner W, Reischl U, Wolf H, et Nerlich AG (2003a) Molecular study on human tuberculosis in three geographically distinct and time delineated populations from ancient Egypt. *Epidemiol Infect* 130:239-49.
- Zink AR, Molnár E, Motamedi N, Pálffy G, Marcsik A, et Nerlich AG (2007) Molecular history of tuberculosis from ancient mummies and skeletons. *Int. J. Osteoarchaeol.* 17:380-391.
- Zink AR, Reischl U, Wolf H, et Nerlich AG (2002) Molecular analysis of ancient microbial infections. *FEMS Microbiol Lett* 213:141-7.
- Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, Wolf H, et Nerlich AG (2003b) Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* 41:359-67.

VIII. Annexes

Annexe 1: Tableau récapitulatif présentant la séquence des amorces PCR utilisées pour l'amplification des fragments d'ADN contenant les SNP d'intérêt, la taille des fragments amplifiés et la concentration finale de ces amorces utilisée dans les réactions PCR multiplexes.

SNP	Séquence de l'amorce PCR (5'→3')		Taille (pb)	Concentration (µM)
	Sens	Antisens		
PCR multiplexe 1 (7 SNP du chromosome Y)				
M173	TTTCTTACAATCAAGGGCATTAG	CTGAAAACAAAACACTGGCTTATCA	81	0,3
M175	AGTACCCAAATCAACTCAACTCCAGTG	CTGATACCTTTGTTTCTGTTCAITCTTGA	100	0,2
M217	GATTCTTTAACTTGTGAAGGAGAATGAA	CGTAAGCATTGATAAAGCTGCTGTG	115	0,3
M45	GGTGTGGACTTTACGAACCAACCTT	TATCTCTGGCCTGGACCTCAGAAG	119	0,4
RPS4Y ₇₁₁	TGGTGGGATGTTGTTTTCTCTCT	CAACAGTAAGTCGAATGCCCTTTCC	123	0,35
M17	ATTGGGGAATACCTGGTCATAACA	AATAGTTTGGCCACTTAACAAACCC ³	124	0,4
M216	TCCTCAACCAGTTTTTATGAAGCTAGA	GCAAAAAGATAATTGTTCCAGGGTAAGC	145	0,2
PCR multiplexe 2 (6 SNP du chromosome Y)				
M9	TCAGGACCTGAAATACAGAACTGC	TTGAACGTTTGAACATGCTAAATTAAG	93	0,3
92R7	TTTAAATCCCTCCTATTGTGCTAACCA	CACTTCTTTTCAGAAAAATGCATGAAC	107	0,35
M3	AATGTGGCCAAGTTTTATCTGCTG	GGCATCTTTCATTTTAGGTACCAGCTC	111	0,15
TAT	GACTCTGAGTGTAGACTTGTGA	GAAAGTGCCGTA AAAAGTGTGAA	112	0,2
M89	TGGATTCAGCTCTCTCCTAAGGTTAT	CTGCTCAGGTACACACAGAGTATCA	135	0,2
M242	GCATAGAAAGTTTGTGCAAAAAGGTGAC	GGGCTTTCAGCATAATACCTTACCTAGAA	155	0,25
Multiplexe 3 (9 SNP de l'ADNmt)				
13563	CCACATCATCGAAACCGCAAA	GCGCTTGTGAGGGAGGTAGC	84	0,4
663	CGGGCTACATCACCCATA	ACGGGATGCTTGCATGTGT	87	0,11
12358	TTTTGGTGCAACTCCAAATAAAA	GGTTAACGAGGGTGGTAAGGA	102	0,5
14766 et 14793	TTCAACTACAAGAACACCAATGACC	CATGCGGAGATGTTGGATGG	108	0,35
14318	ACAAAGCCCCCGACCAATA	GATGGGGTGGTGGTTGTGGT	109	0,5
7028	TGGCATTGTATTAGCAAACATCA	TGAAGCCTCCTATGATGGCAAA	119	0,14
10609	CGCTCACACCTCATATCCTCCCTA	GCAGGCGGCAAAGACTAGTATGG	141	0,14
5301	TGGGCCATTATCGAAGAATCACA	GGGGAGTAGTGTGATTGAGGTGGA	129	0,2
Multiplexe 4 (9 SNP de l'ADNmt)				
12633	TTCATCCCTGTAGCATTGTTCG	GAACTGATTAATGTTTGGGTCTGAG	87	0,16
8508	CTACCACCTACCTCCCTCACC	GGGGCAATGAATGAAGCGAAC	101	0,5
11299	CCCTTCCCCTACTCATCGCACT	TGGCTCAGGAGTTTGATAGTTCTTGG	105	0,18
6752 et 6776	TGGTCTGAGCTATGATATCAATTGG	GCGGAGGTGAAATATGCTCGTG	106	0,35
3720	CAAACCTCAAACATGACCCCTGATCG	GTTGGAGAGGTTAAAGGAGCCACT	125	0,5
4917	GACATCCGGCCTGCTTCTTCTC	TCCACCTCAAACCTGCTATG	136	0,1
14620	CACCCGACCACACCCGCTAAC	TGATGTATGCTTTGTTTCTGTTGAG	118	0,35
6371	ACCCTGGAGCCTCCGTAGACCT	GAAAGAGGGGCGTTTGGTATTGG	140	0,35
Multiplexe 5 (7 SNP de l'ADNmt)				
4336	AGGAGCTTAAACCCCTTATTTT	GGTGTGATAGGTGGCACGGAGA	86	0,42
13101	CCACCCAGTCTCAGCCCTACT	TTTTCTGCTAGGGGGTGAAGC	88	0,34
16482	GCCATAACACTTGGGGGTAGC	AAGGGGAACGTGTGGGCTATTT	95	0,5
3010	ACGACCTCGATGTTGGATCAGG	CGGTCTGAACTCAGATCACGTAGGA	99	0,2
951	AACCCAGGGTGGTCAATTTT	AGCTTTATTGGGGAGGGGGTGA	110	0,34
4769 et 4793	TCCGGACAATGAACCATAACCA	AGGGGTGCCTTGGGTAACCTCT	129	0,35
Multiplexe 6 (2 SNP de l'ADNmt)				
11467	CCATGTCGAAGCCCCCATC	TGTGTTTTGTCAGGGGGTTGAGA	113	0,2
8901	CCTTATGAGCGGGCACAGTGAT	TGGGGATAAGGGGTGTAGGTGTG	99	0,2
Multiplexe 7 (10 SNP autosomaux)				
rs12913832	TTGTCTTCATGGCTCTCTGTGTCTG	AGAGAAGCCTCGGCCCTGA	108	0,4
rs1805007	CCGCTACATCTCCATCTTCTACGC	GCGTGCTGAAGACGACACTGG	110	0,55
rs1805008	ATCTTCTACGCACTGCGCTACCAC	GAAGAGCGTGCTGAAGACGACAC	102	0,55
rs7495174	AGGCCAGGCGGACTCAG	AGGCAGGGAGGGTTTACACAGC	128	0,5
rs6497268	TACAGCCACTCTGGAAAGCAGTTTG	TTGCGTGTAGGGTTTGTGTGAATG	133	0,4
rs11855019	CCAGGATTCAAAAAGAAAGTCTCAAGG	TTTCTCCATCACTGATTTAGCTG	115	0,55

SNP	Séquence de l'amorce PCR (5'→3')		Taille (pb)	Concentration (µM)
	Sens	Antisens		
rs16891982	CCCTATAGTGCACACAACCTCCACAGAG	TGAGGAAAACACGGAGTTGATGC	96	0,3
rs1426654	CCCTTGGATTGTCTCAGGATG	TGAGTAAGCAAGAAGTATAAGGAGCAA	118	0,1
rs2031526	CCTTGAATTGCTCTTGAAAACTAA	CAGCCCAATGATACACTTTCATTTAAC	149	0,55
rs1545397	TGGAATTGGATACTGACAAATGGTTG	CATGGGGGAGAGAGAATGACTCAG	144	0,55
Multiplexe 8 (8 SNP pour la différentiation des espèces du CMTB)				
katG ⁴⁶³	CGGTCGAAACTAGCTGTGAGACAGT	AAGCCGAGATTGCCAGCCTTA	72	0,2
hsp65 ⁶³¹	GTCTCAAACGCGGCATCGAAAAG	AATCTGCTCCTTGGTCTCGACCTC	89	0,3
katG ²⁰³	GTTCACGTAGATCAGCCCCATCTG	ATGAGGTCTATTGGGGCAAGGAAG	113	0,2
16SrRNA ¹⁴²⁹	TGTCCAGGGCTTCACACATGC	TTCACGGGGTCGAGTTGCAG	120	0,2
gyrA ⁹⁵	CGAGACCATGGGCAACTACCA	ATTGCCTGGCGAGCCGAAGT	130	0,5
gyrB(1410)	TCAAATCGTTTGTGCAGAAGGTCTGT	CACCAACTCTCGTGCCTTACGTG	143	0,3
gyrB(675+756)	AGATCAAGCGCGACGGGTA	CGAAGTCGTATTCCGTGGTTTCG	150	0,2
Multiplexe 9 (8 SNP pour la caractérisation des SCG)				
232574	CCACGGCGGGGACAAGAT	AGAAAGGCGCCGCTGTAGG	81	0,3
74092	GACGGTCCGAATTGCCTTGG	GACCAGGAGAAGGCCATCAAAGAG	88	0,5
105139	ATAACGTCGGGCACTGACAAAGAG	TCCCGTATCAACTCGTAGGATCTGG	94	0,5
3352929	TCGACGTCCGGGTAGCATT	GCGTCGCAAGCATCTGACATT	108	0,2
311613	CACCACTGTTGCCACGATGTTCTT	GGCGACTTGCTACGCGTCCTAC	114	0,3
2460626	GGGCTCGCAGCCAGACTTC	ATGATCACGGCGACCCAGAC	120	0,5
1977	TGAGAAGCTCTACGGTTGTTGTTTCG	TTCACCTCACGATGAGTTTCGATCC	131	0,3
913274	GCAATCGCCGTGCAACC	CTGCATGTTATGGGTGACGATGAC	141	0,4

Annexe 2 : Tableau récapitulatif présentant la séquence des amorces PEX utilisées pour le typage par spectrométrie de masse MALDI-TOF de 8 SNP autosomaux et 8 SNP permettant la différenciation des espèces du CMTB et la séquence ainsi que la masse des produits d'extension attendus.

SNP	Séquence des amorces (5'→3') ^a	Séquence des produits d'extension clivés (5'-3')	(*)	Masse de l'amorce non étendue (Da)	Masse attendue pour l'allèle 1 (Da)	Masse attendue pour l'allèle 2 (Da)
Amorces PEX non biotinylées (8 SNP autosomaux)						
rs12913832	GAT AGC GTG CAG AAC TLGACA	GACAddT/GACACddT	R	1264,82	A : 1553,01	G : 1842,20
rs7495174	AAG GCA AGT TCC CCT ALA GGT	AGGTddT/AGGTCAgddT	R	1293,83	A : 1582,02	G : 2513,62
rs6497268	TGT TGG CTG GTA GTT GLA ^{OMe} A ^{OMe} TT	A ^{OMe} A ^{OMe} TT ddT/A ^{OMe} A ^{OMe} TTGddT	R	1315,82	A : 1604,01	C : 1933,22
rs11855019	CAT CAG TGA TTT AGC TGT GLT CTG	T CTGddT/T CTGCCAAAddT	R	1246,79	A : 1534,99	G : 2739,76
rs16891982	GGT TGG ATG TTG GLGC TT	GCTTCddT/GCTTGddT	F	1245,79	C : 1823,17	G : 1863,19
rs1426654	GTC TCA GGA TGT TGCL G GC	G GCACAACddT/GGCGCAACddT	F	967,61	A : 2773,79	G : 2789,79
rs2031526	ACA CTT TCA TTT AAC TTT TGC LTT TG	TTTGddT/TTTGcddT	R	1260,80	A : 1549,00	G : 1838,18
rs1545397	TTT ATC TTG CAA AAT TAT ATC ALT CAG	TCAGddT/TCAGAGddT	R	1255,80	A : 1544,00	T : 2186,41
Amorces PEX biotinylées (8 SNP autosomaux)						
rs12913832	biotine-GAT AGC GTG CAG AAC TLG ^{OMe} ACA	G ^{OMe} ACAAddT/G ^{OMe} ACACddT	R	1294,82	A : 1583,01	G : 1872,20
rs7495174	biotine-AAG GCA AGT TCC CCT ALA ^{OMe} GGT	A ^{OMe} GGTddT/A ^{OMe} GGTCAgddT	R	1325,83	A : 1614,02	G : 2545,62
rs6497268	biotine-TGT TGG CTG GTA GTT GLA ^{OMe} A ^{OMe} TT	A ^{OMe} A ^{OMe} TT ddT/A ^{OMe} A ^{OMe} TTGddT	R	1315,82	A : 1604,01	C : 1933,22
rs11855019	biotine-CAT CAG TGA TTT AGC TGT GLT C ^{OMe} TG	T C ^{OMe} TGddT/T C ^{OMe} TGCCAAAddT	R	1275,79	A : 1564,99	G : 2769,76
rs16891982	biotine-GGT TGG ATG TTG GLGC TT	GCTTCddT/GCTTGddT	F	1245,79	C : 1823,17	G : 1863,19
rs1426654	biotine-GTC TCA GGA TGT TGCL G ^{OMe} GC	G ^{OMe} GCACAACddT/G ^{OMe} GCGCAACddT	F	997,61	A : 2803,79	G : 2819,79
rs2031526	biotine-ACA CTT TCA TTT AAC TTT TGC LTT TG	TTTGddT/TTTGcddT	R	1260,80	A : 1549,00	G : 1838,18
rs1545397	biotine-TTT ATC TTG CAA AAT TAT ATC ALT C ^{OMe} AG	TC ^{OMe} AGddT/TC ^{OMe} AGAGddT	R	1283,80	A : 1572,00	T : 2216,41
Amorces PEX non biotinylées (8 SNP pour la différenciation des espèces du CMTB)						
hsp65 ⁶³¹	GGT CTC GAC CTC CLT GGC	TGGCddG/TGGCACCCCTddG	R	1223,79	C : 1584,01	T : 3349,13
katG ⁴⁶³	CAG CCT TAA GAG CCA LAT CC	ATCCddG/ATCCTddG	R	1214,78	C : 1527,99	A : 1832,18
gyrA ⁹⁵	ACG CGT CGA TCT ALGACA	GACAddG/GACACCCCTddG	F	1263,82	G : 1577,03	C : 2748,77
katG ²⁰³	TAT TGG GGC AAG GAA LCC AC	CCACTTddG/CCACTCddG	R	1199,77	A : 2121,36	G : 2106,35
ARNr 16S ¹²⁴⁹	TTA ACC TCG CGG CLT CGC	T CGCAddG/T CGCddG	R	1230,78	T : 1857,19	C : 1543,99
gyrB(675)	CIT CTC ATA AAC CTG AGA CLA CTC	ACTCddG/ACTCATACCCddG	R	1214,78	C : 1527,99	T : 3326,14
gyrB(756)	AAG ACG GGG TCA LCGGT	CG GTCddG/CG GTTddG	F	1255,80	G : 1544,00	A : 2186,41
gyrB(1410)	CGT TTG TGC AGA AGG TCL GTA A	GTAACddG/GTAATddG	F	1278,83	C : 1881,22	T : 1896,23
Amorces PEX biotinylées (3 SNP pour la différenciation des espèces du CMTB)						
katG ⁴⁶³	biotine-CAG CCT TAA GAG CCA LA ^{OMe} T C ^{OMe} C	A ^{OMe} TC ^{OMe} CddG/A ^{OMe} TC ^{OMe} CTddG	R	1276,78	C : 1589,99	A : 1894,18
gyrA ⁹⁵	biotine-ACG CGT CGA TCT ALG ^{OMe} ACA	G ^{OMe} ACAAddG/G ^{OMe} ACACCCCTddG	F	1294,82	G : 1608,03	C : 2779,77
gyrB(1410)	biotine-CGT TTG TGC AGA AGG TCL GTA A	GTAACddG/GTAATddG	F	1278,83	C : 1881,22	T : 1896,23

^a La molécule photoclivable est représentée par un « L » dans la séquence des amorces. (*)Orientation des amorces selon la séquence du génome de *M. tuberculosis* H37Rv (NC_000962).

Annexe 3 : Tableau récapitulatif présentant la séquence des amorces SBE utilisées pour le typage par miniséquençage des différents SNP d'intérêt, la taille de ces amorces et leur concentration finale dans les réactions de miniséquençage multiplexes.

SNP	Poly (C)	Séquence neutre (5'→3')	Amorce spécifique (5'→3')	(*)	Taille (nt)	Concentration (µM)
SBE multiplexe 1 (13 SNP du chromosome Y)						
M3	-		- GGTACCAGCTCTCCTAATT	R	20	0,08
M217	-		- TTATGTATTTTTCCTTCTGAAGAGTT	R	26	0,1
M242	-	GAAAGTCTGACAA	AAAAGGTGACCAAGGTGCT	F	32	0,2
M45	-	CGTCGTGAAAGTCTGACAA	CTCAGAAGGAGCTTTTTGC	R	38	0,58
M89	-	GCCACGTCGTGAAAGTCTGACAA	AACTCAGGCAAAGTGAGAGAT	R	44	0,2
RPS4Y ₇₁₁	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGCAATAAACCTTGGATTTC	F	50	0,5
92R7	-	TAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATGAACACAAAAGACGTAGAAG	R	56	0,57
M17	-	TGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CCAAAATTCACTTAAAAAACCC	R	60	0,18
M175	2	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CACATGCCTTCTCACTTCTC	F	62	0,2
TAT	2	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GCTCTGAAATATTAATTAACAAC	R	68	0,13
M9	12	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATGTCATAATTAAGAAAAATAAGAG	R	80	0,27
M216	21	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TGCTAGTTATGTATACCTGTGAAT	R	86	0,2
M173	31	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CAATTC AAGGGCATTTAGAAC	F	92	0,2
SBE multiplexe 2 (9 SNP de l'ADNmt)						
13563	-		CAA GAGGTAGCGATGAGAGTAATAGA	R	26	0,5
663	-		GTCTGACAA CATAACAAATAGGTTTGGTCCT	F	32	0,3
12358	-	TCGTGAAAGTCTGACAA	ATAACCATGCACACTACTATA	F	38	0,5
14766	-	GGTGCCACGTCGTGAAAGTCTGACAA	TGACCCCAATACGCAAAA	F	44	0,45
14793	-	TAGGTGCCACGTCGTGAAAGTCTGACAA	CCCTAATAAAATTAATTAACC	F	50	0,5
14318	-	GACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGTGGTTGTGGTAAACTTAA	R	56	0,5
7028	1	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CGACACGTAACGTTGTAGC	F	62	0,2
10609	5	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GCTGTTCATTATAGCTACTCTCA	F	68	0,1
5301	16	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TGGCTATGATGGTGGGA	R	74	0,1
SBE multiplexe 3 (9 SNP de l'ADNmt)						
12633	-		AA ATATCACAGTGAGAATTCTATGAT	R	26	0,2
8508	-	AAAGTCTGACAA	ATTTTGGTTCACAGGTTTG	R	32	0,4
11299	-	GAAAGTCTGACAA	GCTCACTAACATTCTACTACTCAC	F	38	0,15
6752	-	GGTGCCACGTCGTGAAAGTCTGACAA	TGCTCACACGATAAACCC	R	44	0,1
6776	-	GTGCCACGTCGTGAAAGTCTGACAA	CTACGTCTATTCTACTGTAATAT	R	50	0,4
3720	-	TAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	AGGGTGACTTCATATGAGATTGT	R	56	0,5
4917	-	TGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATATACCAAATCTCTCCCTACTA	F	62	0,15
14620	8	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTAGTAATGGGGTTTGTGGG	R	68	0,4
6371	12	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	AATTGATGGCCCTAAGATAGA	R	74	0,3
SBE multiplexe 4 (7 SNP de l'ADNmt)						
4336	-	CTGACAA	GGATGGGTTTCGATTCTCAT	R	26	0,2
13101	-	AAGTCTGACAA	GCGGATGAGTAAGAAGATTCC	R	32	0,1
16482	-	TCGTGAAAGTCTGACAA	CAGATGTCGGATACAGTTAC	R	38	0,2
3010	-	GGTGCCACGTCGTGAAAGTCTGACAA	GTTGGATCAGGACATCCC	F	44	0,2
951	-	GACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CCGGCGTAAAGAGTGTTTTA	F	56	0,2
4769	-	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTCCTAGTTTATTGCTATAGC	R	62	0,2
4793	9	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CTCAGAAGTGAAAGGGGGC	R	68	0,1
SBE multiplexe 5 (2 SNP de l'ADNmt)						
11467	-	AGTCTGACAA	GCTCTAAGATTAATAATGCCCT	F	32	0,4
8901	-	AGGTGCCACGTCGTGAAAGTCTGACAA	TAGTACTTGGCCGACTACTCTTA	F	50	0,4
SBE multiplexe 6 (10 SNP autosomaux)						

SNP	Poly (C)	Séquence neutre (5'→3')	Amorce spécifique (5'→3')	(*)	Taille (nt)	Concentration (µM)
rs12913832	-	AGTCTGACAA	TAGCGTGCAGA AACTTGACA	R	29	0,3
rs1805007	-	AAAGTCTGACAA	CATCTCCATCTTCTACGCACTG	F	34	0,58
rs1805008	-	GCCACGTCGTGAAAGTCTGACAA	CATCGTGACCCTGCCG	F	39	0,58
rs7495174	-	GCCACGTCGTGAAAGTCTGACAA	AAGCAAGTCCCTAAAGGT	R	44	0,58
rs6497268	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTGGCTGGTAGTTGCAATT	R	49	0,5
rs11855019	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATCACTGATTAGCTGTGTTCTG	R	54	0,58
rs16891982	1	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGTTGGATGTTGGGGCTT	F	59	0,4
rs1426654	5	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GTCTCAGGATGTTGCAGGC	F	64	0,1
rs2031526	4	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CAC TTTCATTTAACTTTTGTCTTTG	R	69	0,5
rs1545397	10	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	ATCTTGCAAAATTATATCATTACAG	R	74	0,58
SBE multiplexe 7 (8 SNP pour la différenciation des espèces du CMTB)						
hsp65 ⁶³¹	-	AAGTCTGACAA	GTCTCGACCTCCTTGGC	R	28	0,35
katG ⁴⁶³	-	CGTGAAAGTCTGACAA	GCCTTAAGAGCCAGATCC	R	34	0,15
gyrA ⁹⁵	-	CCACGTCGTGAAAGTCTGACAA	ACGCGTCGATCTACGACA	F	40	0,5
katG ²⁰³	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	GTCATCGCCGAGCCA	F	46	0,14
16SrRNA ¹⁴²⁹	-	ACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGCCGGTACAAAGGGC	F	55	0,25
gyrB(675)	2	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TTCATAAACCTGAGACCACCTC	R	64	0,14
gyrB(756)	13	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	AAGACGGGGTCAACGGT	F	70	0,4
gyrB(1410)	15	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GTTTGTGCAGAAGGTCTGTAA	F	76	0,13
SBE multiplexe 8 (8 SNP pour la caractérisation des SCG)						
1977	-	GTGAAAGTCTGACAA	GACTGCCAACGACGAA	R	31	0,3
3352929	-	TCGTGAAAGTCTGACAA	CTGACATTGGTGCACAAAAC	R	37	0,2
74092	-	GGTGCCACGTCGTGAAAGTCTGACAA	GCGGTGAACCCGGTACT	R	43	0,58
105139	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	AGTTTCACGGTTATCAGCG	R	49	0,58
2460626	-	ACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGACCGAGGGTCTGGC	F	55	0,58
232574	4	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	ACAGGGCAATCACCTCG	F	61	0,2
311613	11	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CCGTTTCGTAGCCGAG	F	67	0,2
913274	16	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TCGCTGGACCCAATCTC	F	73	0,5

(*) Orientation des amorces selon la séquence du génome humain ou du génome de *M. tuberculosis* H37Rv.

Annexe 5 : Tableau présentant les informations indiquées par les 36 individus dont les prélèvements ont été utilisés comme échantillons contemporains, les génotypes obtenus pour les 10 SNP autosomaux à partir de l'analyse de ces échantillons et l'origine biogéographique de ces individus qui a été estimée à partir des génotypes obtenus pour les 4 SNP « AIM » grâce au logiciel STRUCTURE.

Echantillon ^a	Informations indiquées par les individus			rs12913832 HERC2	rs1805007 MC1R	rs1805008 MC1R	OCA2 diplotype ^b	rs16891982 SLC24A2	rs1426654 SLC24A5	rs2031526 DCT	rs1545397 OCA2	Origine biogéographique déduite ^c		
	Couleur des yeux	Couleur des cheveux	Couleur de peau									E	As	Af
E1	bleu	roux	claire	<u>G/G</u>	C/C	C/T	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.963	0.012	0.024
E2	vert	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E3	bleu	blond	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.954	0.024	0.022
E4	bleu	blond	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.960	0.020	0.020
E5	bleu/gris	auburn	claire	<u>G/G</u>	C/T	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.961	0.013	0.026
E6	vert/gris	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	C/G	A/A	G/G	A/A	0.787	0.038	0.175
E7	vert/noisette	brun clair	claire	A/G	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.955	0.022	0.024
E8	vert/noisette	brun foncé	claire	A/A	C/C	C/C	TGT/CTC	G/G	A/A	G/G	A/A	0.961	0.013	0.027
E9	vert/noisette	brun foncé	claire	A/A	C/C	C/C	TTT/CTC	G/G	A/A	G/G	A/A	0.963	0.013	0.024
E10	bleu	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	C/G	A/A	G/G	A/A	0.789	0.049	0.163
E11	vert	auburn	claire	<u>G/G</u>	C/T	C/C	TGT/TGC	G/G	A/A	G/G	A/A	0.958	0.014	0.028
E12	bleu/noisette	brun clair	claire	A/G	C/C	C/C	TGT/TTT	G/G	A/A	G/G	A/A	0.962	0.012	0.026
E13	bleu/noisette	brun clair	claire	A/G	C/C	C/C	TGT/TTT	G/G	A/A	G/G	A/A	0.965	0.013	0.022
E14	vert	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	C/G	A/A	G/G	A/T	0.763	0.165	0.073
E15	marron	brun foncé	claire	A/G	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.957	0.022	0.021
E16	marron	brun foncé	claire	A/A	C/C	C/C	TGT/CTC	C/G	A/A	A/G	A/T	0.669	0.283	0.048
E17	vert/noisette	brun foncé	mate	A/G	C/C	C/C	TGT/TTT	C/G	A/A	G/G	A/T	0.755	0.170	0.076
E18	bleu	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/T	0.935	0.045	0.021
E19	marron	roux	claire	A/G	C/T	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.964	0.013	0.022
E20	vert	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	C/G	A/A	G/G	A/A	0.792	0.047	0.161
E21	vert/gris	blond	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.957	0.022	0.021
E22	bleu	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.959	0.014	0.026
E23	vert/noisette	brun clair	claire	A/G	C/C	C/C	TGT/TTT	G/G	A/A	A/G	A/A	0.957	0.020	0.022
E24	vert	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	C/G	A/A	G/G	A/A	0.786	0.049	0.166
E25	marron	roux	claire	A/G	C/C	T/T	TGT/TGC	G/G	A/A	G/G	A/A	0.963	0.014	0.023
E26	bleu	brun clair	claire	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E27	bleu	roux	claire	<u>G/G</u>	C/C	C/T	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.958	0.014	0.028
Af1	marron	noir	foncé	A/A	C/C	C/C	TGC/TTC	C/C	G/G	A/G	A/A	0.028	0.094	0.878
Af2	marron	noir	foncé	A/A	C/C	C/C	TGC/TTC	C/C	G/G	G/G	A/A	0.023	0.031	0.946
Af3	marron	noir	foncé	A/A	C/C	C/C	TGC/TTC	C/C	A/G	G/G	A/A	0.164	0.041	0.795
As1	-	-	-	A/A	C/C	C/C	TTT/CTC	C/C	G/G	A/G	A/T	0.042	0.649	0.308
As2	-	-	-	A/A	C/C	C/C	CTC/CTC	C/C	G/G	A/G	T/T	0.020	0.921	0.060
As3	-	-	-	A/A	C/C	C/C	CTC/CTC	C/C	G/G	A/A	T/T	0.013	0.964	0.023
As4	-	-	-	A/G	C/C	C/C	TTT/CGC	C/C	A/G	A/A	A/T	0.212	0.708	0.080
As5	-	-	-	A/A	C/C	C/C	TTC/CGC	C/C	G/G	A/G	T/T	0.019	0.922	0.059
As6	-	-	-	A/A	C/C	C/C	CTC/CTC	C/G	G/G	A/A	T/T	0.119	0.858	0.023

^aE, échantillons issus d'individus d'origine européenne ; Af, échantillons issus d'individus d'origine africaine ; As, échantillons issus d'individus d'origine asiatique.

^bLe diplotype OCA2 correspond aux marqueurs rs7495174/rs6497268/rs11855019. Le diplotype OCA2 et le génotype rs12913832 qui sont associés au phénotype « couleur bleue des yeux » sont soulignés.

^cLa probabilité pour chaque échantillon d'être d'origine européenne, asiatique et africaine a été déterminée en utilisant le programme STRUCTURE. La probabilité la plus élevée, qui indique l'estimation de l'ascendance la plus probable, est indiquée en gras. E, origine européenne; Af, origine africaine; As, origine asiatique.

Annexe 6 : Tableau présentant les génotypes obtenus pour les 10 SNP autosomaux à partir de l'analyse des 25 échantillons anciens provenant de la région de Krasnoyarsk et l'origine biogéographique de ces sujets qui a été estimée à partir des génotypes obtenus pour les 4 SNP « AIM » grâce au logiciel STRUCTURE.

Echantillon	Culture	rs12913832 HERC2	rs1805007 MC1R	rs1805008 MC1R	OCA2 diplotype ^a	rs16891982 SLC24A2	rs1426654 SLC24A5	rs2031526 DCT	rs1545397 OCA2	Origine déduite par STRUCTURE ^b		
										E	As	Af
S07	Andronovo	A/A	C/C	C/C	TTT/CTC	C/C	G/A	A/G	T/T	0.143	0.811	0.046
S08	Andronovo	A/A	-	-	-	C/C	G/G	A/A	A/T	0.030	0.870	0.100
S09	Andronovo	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.967	0.011	0.022
S10	Andronovo	-	C/C	C/C	TGT/TTT	G/G	A/A	A/G	A/A	0.966	0.017	0.017
S11	Andronovo	<u>G/G</u>	-	-	-	C/G	A/A	G/G	-	0.837	0.054	0.109
S13	Andronovo	-	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.967	0.016	0.017
S14	Andronovo	A/G	C/C	C/C	TGT/TGC	C/G	A/A	G/G	A/A	0.861	0.033	0.107
S15	Andronovo	<u>G/G</u>	C/C	C/C	-GT/-GT	G/G	A/A	A/A	A/A	0.948	0.033	0.018
S16	Andronovo	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.964	0.016	0.020
S18	Karasuk	<u>G/G</u>	C/C	-	-	G/G	A/G	G/G	-	0.744	0.075	0.181
S19	Karasuk	<u>G/G</u>	-	-	-	-	A/A	G/G	A/A	0.942	0.014	0.044
S21	Tagar	<u>G/G</u>	C/C	-	TGT/TGC	G/G	A/A	G/G	A/A	0.967	0.011	0.023
S22	Tagar	A/G	C/C	C/C	-GT/-GT	G/G	A/A	G/G	-	0.967	0.013	0.020
S23	Tagar	<u>G/G</u>	-	-	-	G/G	A/A	G/G	-	0.967	0.014	0.020
S24	Tagar	<u>G/G</u>	-	-	-	G/G	A/A	G/G	A/A	0.971	0.011	0.018
S25	Tagar	<u>G/G</u>	C/C	-	-GT/-GT	G/G	A/A	G/G	A/T	0.946	0.038	0.016
S26	Tagar	A/G	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/A	A/A	0.953	0.033	0.014
S28	Tagar	A/A	-	-	<u>TGT/TGT</u>	G/G	A/A	G/G	A/T	0.946	0.037	0.016
S29	Tagar	<u>G/G</u>	-	-	-	G/G	A/A	-	-	0.963	0.019	0.018
S32	Tagar	A/G	C/C	-	-	C/C	A/A	A/G	A/T	0.491	0.419	0.090
S33	Tashtyk	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/A	A/A	0.948	0.034	0.018
S34	Tashtyk	<u>G/G</u>	C/C	-	-GT/-GT	G/G	A/A	A/A	-	0.928	0.056	0.016
S35	Tashtyk	A/G	-	-	-	G/G	A/A	A/G	A/T	0.920	0.066	0.015
S36	Tashtyk	<u>G/G</u>	C/C	C/C	<u>TGT/TGT</u>	G/G	A/A	A/G	A/A	0.962	0.018	0.020
S37	Tashtyk	<u>G/G</u>	C/T	C/C	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.969	0.010	0.021

^a Le diplotype OCA2 correspond aux marqueurs rs7495174/rs6497268/rs11855019. Le diplotype OCA2 et le génotype rs12913832, associés au phénotype « couleur bleue des yeux » sont soulignés.

^b La probabilité pour chaque échantillon d'être d'origine européenne, asiatique, et africaine a été déterminée en utilisant le programme STRUCTURE. La probabilité la plus élevée, qui indique l'estimation de l'ascendance la plus probable, est indiquée en gras. E, origine européenne; Af, origine africaine; As, origine asiatique.

IX. Publications

First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA.

Bouakaze C., Keyser C., Amory S., Crubézy E, Ludes B.

International Journal of Legal Medicine, 2007, 121:493-499.

Use of SNPs for the study of ancient *Mycobacterium tuberculosis* genome:

Validation of the strategy on modern DNA.

Bouakaze C., Keyser C., Amory S., Piemont Y., Dabernat H., Crubezy E., Ludes B.

Forensic Science International: Genetics Supplement Series, 2008, 1: 547-548. Progress in

Forensic Genetics 12 - Proceedings of the 22nd International ISFG Congress.

Tracing back ancient south Siberian population history using mitochondrial and Y-chromosome SNPs.

Keyser C., Romac S., Bouakaze C., Amory S., Crubézy E., Ludes B.

Forensic Science International: Genetics Supplement Series, 2008, 1: 343-345. Progress in Forensic Genetics 12 - Proceedings of the 22nd International ISFG Congress.

**Pigment phenotype and biogeographical ancestry from ancient skeletal remains:
inferences from multiplexed autosomal SNP analysis.**

Bouakaze C., Keyser C., Crubézy E., Montagnon D., Ludes B.

International Journal of Legal Medicine, 2009, 123:315-325.

Ancient DNA provides new insights into the history of south Siberian Kurgan people.

Keyser C., Bouakaze C., Crubézy E., Nikolaev V.G., Montagnon D., Reis T., Ludes B.

Human Genetics, 2009, 126:395-410.

Human evolution in Siberia: from frozen bodies to ancient DNA.

Crubézy E., Amory S., Keyser C., Bouakaze C., Bodner M., Gibert M., Röck A., Parson W.,
Alexeev A., Ludes B.

BMC Evolutionary Biology, 2010, 10:25

**Identification and genotyping of *M. tuberculosis* complex species using a SNaPshot
minisequencing-based assay.**

Bouakaze C., Keyser C., de Martino S. J., Sougakoff W., Veziris N., Dabernat H., Ludes B.

Journal of Clinical Microbiology, 2010, 48:1758-1766.

Analyse de polymorphismes ponctuels d'ADN extraits de tissus humains anciens : application à l'étude de populations sibériennes anciennes

Depuis deux décennies, les études d'ADN ancien ne cessent d'enrichir nos connaissances sur l'histoire et l'évolution de l'Homme. Cependant, la plupart de ces études sont basées sur l'analyse de polymorphismes ponctuels (SNP) de la région non codante de l'ADN mitochondrial en raison du plus grand nombre de copies de cet ADN dans les cellules humaines alors que cette région ne représente qu'une infime partie du génome humain et de ce fait restreint le type d'information qu'il est possible d'obtenir à partir de spécimens anciens.

Dans ce travail de thèse, une approche méthodologique originale basée sur l'analyse de différents SNP a été utilisée sur des échantillons anciens dans le but d'étudier l'histoire de populations sibériennes du passé.

Des SNP du génome humain, localisés au niveau de la région codante de l'ADN mitochondrial mais aussi de l'ADN nucléaire, ont été analysés à partir d'échantillons anciens provenant du Sud de la Sibérie Centrale afin d'apporter de nouveaux éléments quant au peuplement de cette région. Les données génétiques obtenues indiquent qu'aux Ages du Bronze et du Fer, cette région était peuplée de sujets au teint pâle, aux cheveux clairs et aux yeux bleus ou verts qui étaient originaires d'Europe de l'Est. Nos résultats ont donc permis de mettre en évidence l'existence de mouvements de populations d'Europe de l'Est jusqu'au Sud de la Sibérie avant 1800 avant J.-C. De plus, au vu de données archéologiques et linguistiques, il semblerait que ces mouvements puissent être liés à l'expansion vers l'Est de peuples ayant parlé le proto indo-européen.

L'analyse de SNP du génome d'agents pathogènes responsables de la tuberculose menée à partir de vertèbres yakoutes présentant des lésions osseuses évocatrices d'une atteinte tuberculeuse a permis d'appréhender l'état sanitaire de la population yakoute (Sibérie Orientale). En effet, le fait d'avoir pu diagnostiquer la tuberculose chez 3 de ces sujets et d'identifier l'espèce *M. tuberculosis* comme agent responsable de l'infection a apporté, pour la première fois, la preuve de la présence de la tuberculose humaine en Yakoutie aux XVII^{ème} et XVIII^{ème} siècles. L'apparition de cette maladie chez les Yakoutes pourrait être liée à l'arrivée des colons russes en Yakoutie au XVII^{ème} siècle.

La qualité des résultats obtenus ainsi que les hypothèses originales qui ont pu être proposées suite à ce travail de recherche attestent de l'intérêt que présente l'analyse de SNP pour les études d'ADN ancien. Il démontre en outre que le typage des SNP basé sur la technique de miniséquençage à l'aide du kit SNaPshotTM est une méthode particulièrement bien adaptée à l'étude des molécules d'ADN dégradées.

Genotyping of single nucleotide polymorphisms in DNA extracted from ancient human tissues: application to the study of ancient Siberian populations

Ancient DNA studies have allowed a better understanding of the history and evolution of humankind since their beginning only 20 years ago. Most of them have typically focused on the analysis of single nucleotide polymorphisms (SNPs) in the non-coding region of the mitochondrial DNA, which is largely due to the high copy number of this DNA in human cells. However, since this region represents only a small part of the human genome, it restricts the type of information that can be obtained from the ancient DNA molecules.

In the present work, an original methodological approach based on the study of various SNP was used on ancient DNA samples to study the history of ancient Siberian populations.

On the one hand, SNPs of the human genome, located in the coding region of the mitochondrial DNA and also in the nuclear DNA, were analyzed on ancient South Siberian samples in order to shed light on the peopling of this region. The resulting data indicates that, at the Bronze and Iron Ages, this region was inhabited by fair-skinned, light-haired and blue- or green-eyed people that originated from Eastern Europe. Our results thus show that population movements from Eastern Europe to South Siberia occurred before 1800 Before Christ. Moreover, according to archeological and linguistic data, it seems that these movements were linked to the eastward expansion of people who spoke the ancestral Indo-European or Proto-Indo-European language.

On the other hand, SNPs enabling the distinction between common tuberculosis pathogens were analyzed on ancient Yakut vertebrae with lytic lesions evocative of the diagnosis of this disease in order to investigate the sanitary state of the Yakut population (Oriental Siberia). The fact that tuberculosis was diagnosed in 3 cases and that *M. tuberculosis* was found to be responsible for these infections provide for the first time the proof that human tuberculosis was present in Yakutia in the 17th and 18th centuries. The appearance of this disease could be linked to the Russian colonization that started in the 17th century.

The quality of the ancient data obtained during this work as well as the possibility to propose new hypotheses give evidence that SNPs are valuable markers for ancient DNA studies that aim at retracing the origin of past human populations and studying their health status. Moreover, it demonstrates that the SNaPshotTM-based minisequencing is a method of choice for SNP typing on degraded DNA templates.