

TABLE DES MATIÈRES

I LA YAKOUTIE : CONTEXTE GENERAL ET PEUPLEMENT	13
I.1 LA GEOGRAPHIE : CONTEXTE GENERAL	14
I.1.1 <i>Présentation du terrain d'étude</i>	14
I.1.2 <i>L' alas : une particularité du paysage de la Yakoutie Centrale</i>	15
I.1.3 <i>Le permafrost</i>	17
I.2 DES PREMIERS PEUPLEMENTS A LA COLONISATION RUSSE	20
I.2.1 <i>Premiers peuplements de la Yakoutie</i>	20
I.2.2 <i>Pré et protohistoire</i>	21
I.2.3 <i>Le Moyen-Âge</i>	22
I.2.4 <i>La colonisation russe</i>	25
I.3 LE PEUPLEMENT ACTUEL : CARACTERISTIQUES ET REPARTITION	27
I.3.1 <i>Tchouktches et Youkaguir</i>	27
I.3.2 <i>Les populations toungousses : Evenks et Evènes</i>	28
I.3.3 <i>Les Yakoutes</i>	31
I.3.4 <i>Répartition actuelle des populations</i>	33
I.4 PROBLEMATIQUE ET HYPOTHESES SUR L'ETHNOGENESE YAKOUTE	35
I.4.1 <i>L'ethnogenèse yakoute : une problématique complexe</i>	35
I.5 INTERETS METHODOLOGIQUES ET MOYENS MIS EN OEUVRE	37
II L'ADN ANCIEN	39
II.1 LES CHAMPS D'ÉTUDE DE L'ADN ANCIEN	40
II.1.1 <i>Historique de la discipline</i>	40
II.1.1.1 Les premières études	40
II.1.1.2 Les premières désillusions.....	41
II.1.2 <i>La réaction d'amplification en chaîne : Polymérase Chain Reaction (PCR)</i>	42
II.1.3 <i>Les champs d'application</i>	44
II.1.3.1 Etudes de restes humains	44
II.1.3.1.1 Détermination du sexe des individus.....	44
II.1.3.1.2 Etude des relations de parenté	45
II.1.3.1.3 Etude des mouvements de populations.....	46
II.1.3.1.4 La lignée néandertalienne.....	48
II.2 LES SOURCES POTENTIELLES D'ADN ANCIEN	51
II.2.1 <i>Les os</i>	51
II.2.1.1 Structure et propriétés.....	51
II.2.1.2 Diagenèse des tissus osseux et préservation de l'ADN	52
II.2.2 <i>Les dents</i>	54
II.2.2.1 Structure	54
II.2.2.2 Utilisation des dents en paléogénétique	55
II.2.3 <i>Les cheveux</i>	56
II.2.3.1 Structure du cheveu	57
II.2.3.2 Propriétés structurales et l'analyse moléculaire des cheveux	58
II.2.4 <i>Les autres substrats</i>	59
II.3 LES SPÉCIFICITÉS DES ÉTUDES SUR L'ADN ANCIEN	61
II.3.1 <i>Les dégradations de la molécule d'ADN</i>	61
II.3.1.1 Nature des dégradations	61
II.3.1.1.1 Fragmentation du brin d'ADN	61
II.3.1.1.2 Hydrolyse	63
II.3.1.1.3 Oxydation	64
II.3.1.1.4 Liaisons croisées	64
II.3.1.2 Sources de dégradations et facteurs environnementaux.....	66
II.3.1.2.1 La température	66
II.3.1.2.2 Le pH	67
II.3.1.2.3 L'humidité	67
II.3.1.2.4 La quantité d'oxygène	67
II.3.1.2.5 Le rayonnement ultraviolet	68
II.3.1.3 Conditions environnementales et taphonomiques dans notre étude.....	68
II.3.1.4 Influence des dégradations post-mortem sur l'analyse et l'interprétation des résultats.....	71
II.3.1.5 Les méthodes de "réparation" de l'ADN	72
II.3.2 <i>Les inhibitions</i>	75

II.3.2.1	Les différents types d'inhibiteurs	75
II.3.2.2	Mode d'action des inhibiteurs	76
II.3.2.2.1	Action sur la molécule d'ADN.....	76
II.3.2.2.2	Action sur la Taq polymérase.....	76
II.3.2.3	Les méthodes de détection des inhibiteurs.....	77
II.3.2.4	Les méthodes de purification des inhibiteurs.....	78
II.3.2.4.1	L'extraction.....	78
II.3.2.4.2	L'amplification.....	79
II.3.3	<i>Les contaminations</i>	81
II.3.3.1	Les sources de contaminations	81
II.3.3.2	Précautions nécessaires à l'étude de l'ADN ancien.....	83
II.3.3.2.1	La fouille et le prélèvement des échantillons	83
II.3.3.2.2	Typage des personnes impliquées	83
II.3.3.2.3	Stockage	84
II.3.3.2.4	Bonnes pratiques de laboratoire	84
II.3.3.3	Méthodes de décontamination	85
II.3.3.3.1	Décontamination physique	85
II.3.3.3.2	Décontamination chimique.....	86
II.3.3.3.3	Décontamination des locaux et du matériel de laboratoire	86
II.4	CRITÈRES D'AUTHENTIFICATION DES RÉSULTATS	88

III APPROCHE ADOPTEE ET MARQUEURS ANALYSES94

III.1	METHODES DE DECONTAMINATION ET DE PREPARATION DES SUBSTRATS	95
III.1.1	<i>Prélèvement des échantillons archéologiques</i>	95
III.1.2	<i>Les échantillons anciens</i>	96
III.1.2.1	Méthodes de fouilles.....	96
III.1.2.2	Répartition géographique	99
III.1.2.3	Datations.....	101
III.1.3	<i>Les substrats osseux</i>	105
III.1.4	<i>Les dents</i>	105
III.1.5	<i>Les cheveux</i>	106
III.1.6	<i>Trépanation et cryobroyage des tissus durs</i>	107
III.2	EXTRACTION ET PURIFICATION DE L'ADN	109
III.2.1	<i>L'extraction</i>	109
III.2.2	<i>Lyse et décalcification</i>	109
III.2.3	<i>Extraction organique</i>	110
III.2.4	<i>Purification et concentration</i>	110
III.3	LES MARQUEURS MOLECULAIRES	112
III.3.1	<i>Quantification de l'ADN par PCR en temps réel</i>	112
III.3.1.1	Intérêt de la quantification.....	112
III.3.1.2	Principe de la quantification	112
III.3.2	<i>Les marqueurs à transmission biparentale</i>	114
III.3.2.1	Détermination du sexe des individus : le gène de l'amélogénine	114
III.3.2.2	Les STR autosomaux.....	115
III.3.2.2.1	Généralités et propriétés.....	115
III.3.2.2.2	Artéfacts d'amplification.....	115
III.3.2.2.3	Amplification et analyse	117
III.3.2.2.4	Détermination des relations de parenté.....	120
III.3.3	<i>Les marqueurs à transmission uniparentale</i>	121
III.3.3.1	Haplotypes et haplogroupes.....	121
III.3.3.1.1	La notion d'haplotype	121
III.3.3.1.2	La notion d'haplogroupe	122
III.3.3.2	Les marqueurs du chromosome Y	124
III.3.3.2.1	Les STR	125
III.3.3.2.2	Les SNP	127
III.3.3.3	L'ADN mitochondrial	129
III.3.3.3.1	Généralités et propriétés.....	129
III.3.3.3.2	Les contraintes	132
III.3.3.3.2.1	Les hétéroplasmes	132
III.3.3.3.2.2	Les hot spots de mutation <i>in vivo</i> et <i>postmortem</i>	134
III.3.3.3.2.3	Les mutations fantômes	135
III.3.3.3.2.4	Les pseudogénomes	136
III.3.3.3.3	Amplification, séquençage et analyse	136
III.3.3.3.4	Procédure de clonage	138
III.3.3.4	Analyse des données des marqueurs à transmission uniparentale	139

III.3.3.4.1	Constitution et mise en place des bases de données	139
III.3.3.4.2	Haplotypes et haplogroupes partagés	140
III.3.3.4.3	Analyses statistiques	142
IV	RESULTATS	144
IV.1	QUANTIFICATION DE L'ADN NUCLEAIRE	145
IV.1.1	<i>Les substrats osseux</i>	145
IV.1.2	<i>Utilisation du PTB</i>	148
IV.1.3	<i>Les dents</i>	151
IV.1.4	<i>Les cheveux</i>	152
IV.2	AMPLIFICATION TOTALE DU GENOME: KIT GENOMIPHITM	154
IV.3	LES STR AUTOSOMAUX	156
IV.3.1	<i>Résultats généraux</i>	156
IV.3.2	<i>Efficacité de typage</i>	159
IV.3.2.1	Comparaison des kits utilisés sur les substrats osseux	159
IV.3.2.2	Comparaison des substrats analysés	161
IV.3.2.3	Influence du PTB	163
IV.3.3	<i>Equilibre de Hardy-Weinberg</i>	165
IV.3.4	<i>Analyse Moléculaire de Variance</i>	167
IV.3.5	<i>Relations de parenté au sein des différents ensembles funéraires</i>	168
IV.3.5.1	Le site d'Arbre Chamanique (AC)	168
IV.3.5.2	Le site de Jarama	171
IV.3.5.3	Le site de Bekh Alas	171
IV.3.5.4	Le site de Ken Ebe	172
IV.3.5.5	Le site de Bouogaryma	173
IV.3.5.6	Les tombes n°9 et n°10	174
IV.3.5.7	Site de la rivière Tandy	174
IV.4	LES MARQUEURS DU CHROMOSOME Y	175
IV.4.1	<i>Les STR</i>	175
IV.4.1.1	Efficacité de typage	175
IV.4.1.2	Analyse des données	179
IV.4.2	<i>Les SNP</i>	184
IV.4.3	<i>Analyse Network</i>	186
IV.4.4	<i>Distances génétiques et analyse multidimensionnelle</i>	187
IV.5	L'ADN MITOCHONDRIAL	191
IV.5.1	<i>Résultats généraux et paramètres de diversité</i>	191
IV.5.2	<i>Vérification des séquences : construction d'un Reduced Median network</i>	193
IV.5.3	<i>Haplotypes partagés</i>	196
IV.5.4	<i>Détermination des haplogroupes</i>	198
IV.5.5	<i>Analyse multidimensionnelle</i>	202
V	DISCUSSION	205
V.1	L'ADN ANCIEN : APPROCHE FONDAMENTALE	206
V.1.1	<i>Spécificités des échantillons anciens de Yakoutie Centrale</i>	206
V.1.1.1	Authentification des résultats	206
V.1.1.2	Importance des conditions taphonomiques	208
V.1.2	<i>Types de prélèvements et qualité des résultats</i>	210
V.1.2.1	Les dents : "coffre fort" à ADN	210
V.1.2.2	Les os	210
V.1.2.3	Les cheveux	212
V.1.2.4	Influence du bromure de N-Pénacyl Thiazolium lors de l'extraction	212
V.1.2.5	Amplification totale du génome	213
V.1.2.6	Efficacité des kits d'amplification STR	214
V.1.3	<i>Perspectives méthodologiques</i>	216
V.1.3.1	Protocole d'extraction d'ADN	216
V.1.3.2	Quantification de l'ADN mitochondrial	216
V.1.3.3	Les STR de faible poids moléculaire	217
V.1.3.4	Détermination des haplogroupes mitochondriaux par la technique SNaPshot	218
V.2	LE PEUPLEMENT DE LA YAKOUTIE CENTRALE	219
V.2.1	Le recrutement des ensembles funéraires	219
V.2.1.2	La période ancienne	220
V.2.1.3	La période turco-mongole	222
V.2.1.4	La période mongole	225

V.2.1.5	La période chrétienne	228
V.2.1.6	Approche globale des lignées paternelles anciennes	230
V.2.1.7	Approche globale des lignées maternelles anciennes	231
V.2.1.8	Des différences régionales au sein de la Yakoutie Centrale	232
V.2.1.9	Conclusions générales	235
VI REFERENCES BIBLIOGRAPHIQUES	238	
VII ANNEXES	266	
VII.1	ANNEXE 1	267
VII.2	ANNEXE 2	269
VII.3	ANNEXE 3	270
VII.4	ANNEXE 3	281
VII.5	ANNEXE 5	305
VII.6	ANNEXE 6	314
VII.7	ANNEXE 7	315
VII.8	ANNEXE 8	316
VII.9	ANNEXE 9	319
VII.10	ANNEXE 10	320
VII.11	ANNEXE 11	321
VII.12	ANNEXE 12	323
VII.13	ANNEXE 13	324
VII.14	ANNEXE 14	325
VII.15	ANNEXE 15	326
VII.16	ANNEXE 16	327
VII.17	ANNEXE 17	328
VIII PUBLICATIONS	329	

Liste des figures

Figure 1 : Vue aérienne de la succession de différents alas dans la région de Churapcha.	p16
Figure 2 : Distribution du permafrost continu et discontinu en Sibérie.	p18
Figure 3 : Tombe de Jarama 1 découverte en sous bois lors de l'expédition MAFSO 2004.	p19
Figure 4 : Localisation des sites Paléolithiques mis au jour en Yakoutie.	p21
Figure 5 : Répartition des populations autochtones en Sibérie Orientale au XVII ^{ème} siècle.	p28
Figure 6 : Répartition actuelle des populations de Sibérie.	p30
Figure 7 : Peintures rupestres des falaises du site de Shishkino.	p32
Figure 8 : Schéma des différentes étapes de l'amplification par PCR.	p43
Figure 9 : Répartition géographique des sites ayant livré des restes de Néandertaliens analysés au niveau moléculaire.	p51
Figure 10 : Structure hiérarchique de l'os compact humain.	p52
Figure 11 : Modèle schématique de fixation de l'ADN à la surface d'un cristal d'hydroxyapatite.	p53
Figure 12 : Schéma de coupe longitudinale d'une dent.	p55
Figure 13 : Structure du bulbe capillaire.	p57
Figure 14 : Schéma d'une tige de cheveu.	p58
Figure 15 : Sites de lésions sur la molécule d'ADN par les différent type de dégradation.	p62
Figure 16a : Rupture de brin par hydrolyse de liaison.	p63
Figure 16b : Rupture de brin liée à la perte de base azotée.	p63
Figure 17 : Différents types de réactions chimiques conduisant à la formation de liaisons croisées.	p65
Figure 18 : Différentes couvertures en écorce de bouleau sur la partie supérieure externe du coffre, partie intérieure recouvrant le corps et sur le fond du coffre. Tombe de <i>Munur Urek</i>	p70
Figure 19 : Localisation des trois <i>ulus</i> de Yakoutie Centrale au sein desquelles les fouilles ont été réalisées. ...	p99
Figure 20 : Tombe de <i>Oulakhan Alas</i> , correspondant à la première période médiévale.	p102
Figure 21 : Tombe de <i>Kyys Ounouoga</i> dont le sujet féminin porte un manteau décoré par les perles caractéristiques des tombes de cette période.	p103
Figure 22 : Tombe de <i>Seden</i> , caractéristique de la période post-colonisation par la présence d'une croix ainsi que d'un cierge.	p104
Figure 23 : Hémi-diaphyse de fémur abrasée et trépanée.	p107
Figure 24 : Principe de fonctionnement des sondes Taqman.	p113

Figure 25 : Modes de formation des <i>stutters</i> : a : réPLICATION normale, b : apparition d'un stutter en raison d'un glissement en arrière, c : mode le plus fréquent de formation des stutters par glissement en avant de la polymérase.	p117
Figure 26 : Fluorophores et échelle de taille des produits de PCR suivant les loci inclus dans le kit Ampf/STR® Profiler Plus™.	p118
Figure 27 : Fluorophores et échelle de taille des produits de PCR suivant les loci inclus dans le kit Ampf/STR® Identifiler™.	p118
Figure 28 : Carte de la distribution globale des haplogroupes du chromosome Y.	p123
Figure 29 : Exemple de réticulation d'un haplogroupe en sous-haplogroupes.	p123
Figure 30a : Fluorophores et échelle de taille des produits de PCR des loci inclus dans le kit PowerPlex® Y.	p126
Figure 30b : Fluorophores et échelle de taille des produits de PCR des loci inclus dans le kit Ampf/STR® Yfiler™.	p126
Figure 31 : Carte schématique de l'ADN mitochondrial humain.	p129
Figure 32 : Carte schématique de la région contrôle de l'ADN mitochondrial.	p130
Figure 33 : Exemple d'hétéroplasmie en position 16093 observée sur des séquences obtenues à partir d'extraits d'ADN provenant de différents substrats (sujet Munur Urekh 1).	p133
Figure 34 : Proportions pour chaque ordre de grandeur de concentration mesuré avec le kit Quantifiler™ observées pour les différents types d'échantillons osseux.	p147
Figure 35 : Comparaison des valeurs moyennes de quantifications obtenues pour les différents types de substrats.	p153
Figure 36 : Contrôle des produits d'amplification Genomiphi par dépôt sur gel d'électrophorèse à 1% d'agarose.	p155
Figure 37a : Efficacité de typage du kit Profiler Plus™ évaluée sur 64 amplifications réalisées à partir d'échantillons osseux.	p160
Figure 37b : Efficacité de typage du kit Identifiler™ évaluée sur 168 amplifications réalisées à partir d'échantillons osseux.	p161
Figure 38a : Efficacité de typage du kit Identifiler™ évaluée sur 168 amplifications réalisées à partir d'échantillons osseux.	p162
Figure 38b : Efficacité de typage du kit Identifiler™ évaluée sur 31 amplifications réalisées à partir d'échantillons dentaires.	p162
Figure 39 : Efficacité de typage des STR autosomaux pour les trois groupes d'échantillons.	p163
Figure 40 : Efficacité de typage comparative pour l'ensemble des échantillons testés avec le tampon de lyse contenant du PTB.	p165
Figure 41 : Corps des 4 sujets inhumés dans le coffre de la tombe d'Arbre Chamanique 1.	p168
Figure 42 : Relations de parenté au sein de la tombe Arbre Chamanique 1.	p169
Figure 43 : Ensemble des tombes du site principal de Ken Ebe.	p172
Figure 44a : Efficacité de typage du kit PowerPlex® Y évaluée à partir de 31 amplifications.	p178

Figure 44b : Efficacité de typage du kit AMPf/STR® Y-Filer™ évaluée à partir de 65 amplifications.	p179
Figure 45 : Comparaison des indices de diversité génétique pour les STR du chromosome Y entre différentes populations.	p180
Figure 46 : <i>Median joining</i> network des haplotypes STR N3 de différentes populations.	p187
Figure 47 : Analyse multidimensionnelle basée sur les paires de distances F_{st} calculées à partir des haplotypes de 7 STR du chromosome Y.	p188
Figure 48 : Analyse multidimensionnelle chronologique basée sur les paires de distances F_{st} calculées à partir des haplotypes de 7 STR du chromosome Y.	p189
Figure 49 : Analyse multidimensionnelle géographique basée sur les paires de distances F_{st} calculées à partir des haplotypes de 7 STR du chromosome Y.	p190
Figure 50 : <i>Reduced median</i> network construit à partir des haplotypes mitochondriaux de la population Yakoute ancienne.	p195
Figure 51 : Analyse multidimensionnelle basée sur les paires de distance F_{st} calculées à partir des haplotypes de la région HV1 des sujets anciens répartis en sous groupes chronologiques.	p202
Figure 52 : Analyse multidimensionnelle basée sur les paires de distance F_{st} calculées à partir des haplotypes de la région HV1 des sujets anciens répartis en sous groupes géographiques.	p204
Figure 53 : Contacts entre populations et schémas de migration proposés.	p237

Rapport: Génétique

Liste des tableaux

Tableau 1 : Liste des échantillons anciens prélevés.	p97-98
Tableau 2 : Répartition géographique des sujets anciens étudiés suivant la région de leur mise au jour.	p100
Tableau 3 : Récapitulatif des données de la quantification par PCR en temps réel.	p145
Tableau 4 : Résultats de la quantification obtenus pour les sujets testés avec le protocole d'extraction contenant du PTB dans le tampon de lyse. Les échantillons sont répartis en 3 groupes suivant les résultats obtenus avec le PTB : quantité d'ADN mesurée plus faible (tableau 4a), quantité équivalente (tableau 4b) et augmentation du rendement (tableau 4c)	p150
Tableau 5 : Comparaison des quantifications des extraits obtenus à partir de dents et d'os.	p150
Tableau 6 : Concentrations respectives des différents échantillons anciens utilisés pour le test Genomiph TM (Amersham).	p154
Tableau 7 : Profils consensus des STR autosomaux obtenus à partir de l'ensemble des substrats.	p157-158
Tableau 8 : Résultats obtenus pour le test exact de Hardy Weinberg basé sur les STR autosomaux.	p166
Tableau 9 : Résultats de l'analyse AMOVA réalisée à partir des données des STR autosomaux et basée sur les répartitions chronologiques et géographiques de l'ensemble des sujets anciens.	p167
Tableau 10 : Profils consensus établis à partir des différentes amplifications réalisées pour les STR du chromosome Y.	p177
Tableau 11 : Diversités haplotypiques basées sur les STR du chromosome Y de différentes populations.	p181
Tableau 12 : Haplotypes minimum des STR du chromosome Y observés dans la population Yakoute ancienne.	p182
Tableau 13 : Comparaisons des haplotypes du chromosome Y dans notre base de données.	p183
Tableau 14 : Affiliation des sujets anciens aux différents haplogroupes du chromosome Y.	p184
Tableau 15 : Fréquences de l'haplogroupe N3 dans les populations eurasiennes.	p185
Tableau 16 : Comparaison des diversités haplotypiques et génétiques basées sur la région HV1 de l'ADN mitochondrial de différentes populations d'Asie Centrale et de Sibérie.	p193
Tableau 17 : Haplotypes mitochondriaux partagés au sein de l'ensemble des sujets anciens.	p196
Tableau 18 : Correspondances obtenus lors de la comparaison des haplotypes mitochondriaux des sujets anciens avec les bases de données.	p197
Tableau 19 : Distribution des haplotypes mitochondriaux chez les sujets anciens et la population moderne. .	p200

Liste des abréviations

A	adénine
AD	<i>Anno Domini</i> (après l'an 0)
ADN	acide désoxyribonucléique
ADNmt	ADN mitochondrial
BET	bromure d'éthidium
BP	<i>before present</i> (avant l'époque présente)
BSA	sérum albumine bovine
C	cytosine
ddNTP	didéoxyribonucléotides triphosphates
dNTP	déoxyribonucléotides triphosphates
DTT	Dithiothréitol
EDTA	acide éthylène diamine tétra acétique
G	guanine
JC	Jésus Christ
Kb	kilo base
KCL	chlorure de potassium
HV1	première région hyper variable de la région de contrôle de l'ADN mitochondrial
HV2	deuxième région hyper variable de la région de contrôle de l'ADN mitochondrial
LR	<i>Likelihood Ratio</i> (rapport de vraisemblance)
M	molaire (mole/litre)
mM	millimolaire
MgCl ²	chlorure de magnésium
min	minute
ml	millilitre
NaOCl	hypochlorite de sodium
ng	nanogramme
ng.µl ⁻¹	nanogramme par microlitre
np	<i>nucleotide position</i> (position nucléotidique)
NR	non retrouvé
NRY	partie non recombinante du chromosome Y
nt	nucléotide
pb	paire de base
pg	picogramme
PCR	amplification élective in vitro (Polymerase Chain Reaction)
pH	potentiel hydrogène
PTB	n-phenacyl thiazolium de bromure
RFLP	<i>Restriction Fragment Length Polymorphism</i> (polymorphisme de longueur des fragments de restriction)
rpm	rotation par minute
SDS	sodium dodécyl sulfate
SNP	<i>Single Nucleotide Polymorphism</i> (polymorphisme ponctuelle de séquence)
sec	seconde
STR	<i>Short Tandem Repeat</i> (microsatellite)
STR-Y	microsatellite présent sur le chromosome Y
T	thymine
TBE	tris-borate-EDTA
TE	Tris-EDTA
UV	Ultra Violets
µl	microlitre
°C	degré Celsius

INTRODUCTION GENERALE

L'évolution des techniques de biologie moléculaire associée à une meilleure connaissance des phénomènes régissant les modifications subies par l'ADN au cours du temps sont à l'origine de la progression rapide des recherches en paléogénétique. Depuis les découvertes des premières séquences anciennes il y a une vingtaine d'années, les progrès réalisés permettent aujourd'hui l'étude d'un grand nombre de sujets ou encore l'analyse de génomes entiers de sujets exceptionnellement bien conservés. Si les possibilités techniques rendent actuellement possible la compréhension des phénomènes de dégradation et les interactions permettant la pérennité des acides nucléiques, nul doute que les découvertes des années à venir contribueront encore à élargir les possibilités d'études sur l'ADN ancien. Ces techniques profitent à de nombreux domaines tels la génétique des populations humaines, animales ou végétales, les études sur la domestication ou encore la paléopathologie. La paléoanthropologie en particulier a grandement bénéficié des recherches sur l'ADN ancien. En effet, bien que toutes les problématiques ne soient pas accessibles par le biais de la paléogénétique, il a néanmoins été possible d'apporter des éléments intéressants concernant, entre autres, les relations entre les Hommes anatomiquement modernes et les Néandertaliens ou encore entre les populations de fermiers néolithiques et les chasseurs-cueilleurs qui vivaient en Europe.

Toutefois, la dégradation au cours du temps est inéluctable et le savoir actuel ne permet pas l'analyse de molécules d'ADN trop fortement fragmentées en raison de conditions environnementales trop défavorables. Il est, par conséquent, très avantageux de pouvoir étudier des spécimens ayant été inhumés dans des conditions permettant une bonne conservation des acides nucléiques. Le permafrost qui recouvre la quasi-totalité du sol de la Sibérie Orientale représente un milieu très favorable à la préservation de l'ADN rendant cette région particulièrement intéressante pour les recherches en anthropologie moléculaire.

En outre, le contexte général du peuplement de la Sibérie Orientale est passionnant du fait de la position géographique de cette région à l'interface des mondes européens, asiatiques et américains, mais aussi en raison du grand nombre de populations autochtones vivant sur cet immense territoire. Ces groupes ethniques sont rattachés à des familles linguistiques différentes et possèdent des modes de vie variés allant des éleveurs semi-nomades aux

chasseurs et éleveurs de rennes. Au sein de la Sibérie Orientale, les Yakoutes représentent une énigme. En effet, seuls éleveurs de bétail et de chevaux entourés par des chasseurs-cueilleurs, ils contrastent également par leurs spécificités linguistiques et culturelles. L'origine des Yakoutes faisant l'objet de différentes hypothèses, une approche pluridisciplinaire combinant l'anthropologie classique et moléculaire, la génétique des populations actuelles et l'archéologie semblait particulièrement adaptée pour répondre à cette problématique. L'accès possible à un nombre important d'échantillons anciens, couvrant une période de temps s'étendant du III^{ième} siècle avant JC au XIX^{ième} siècle AD, représentait une opportunité unique d'étudier l'ethnogenèse et l'évolution de cette population. C'est dans ce contexte que le Professeur Eric Crubézy a initié en 2003, avec le soutien du Ministère des Affaires Etrangères, le programme de recherche des « Missions Archéologiques Françaises en Sibérie Orientale » et l'ACI « Le complexe spatial Altaï Baïkal, Plaque tournante des flux géniques en Haute Asie, de la période protohistorique à l'époque actuelle » avec l'appui du Ministère de la Recherche qui a soutenu le travail de recherche présenté dans ce mémoire.

Le premier chapitre de ce mémoire sera consacré à la description de la Yakoutie dans le contexte général de la Sibérie Orientale ainsi qu'à la présentation des différentes populations autochtones vivant dans cette région géographique permettant d'appréhender la problématique générale de ce travail de thèse. Après un bref historique de la discipline puis la description des spécificités et des contraintes propres aux études en anthropologie moléculaire, nous détaillerons la méthodologie ainsi que les marqueurs génétiques employés lors de ce travail. Les résultats obtenus seront ensuite exposés puis discutés au regard des problématiques anthropologique et moléculaire. Nous terminerons ce mémoire en apportant les éléments concernant la formation des Yakoutes et leurs relations avec les autres populations de Sibérie Orientale.

I LA YAKOUTIE : CONTEXTE GENERAL ET PEUPLEMENT

I.1 La géographie : contexte général

I.1.1 Présentation du terrain d'étude

La Sibérie, vaste territoire recouvrant la totalité de l'Asie septentrionale, s'étend d'ouest en est depuis les montagnes de l'Oural à Vladivostok et de la Mongolie au sud jusqu'à l'océan glacial arctique au nord. Cette immense aire géographique est traditionnellement divisée en trois grandes régions du fait de la diversité des milieux naturels rencontrés. Aux frontières de l'Europe, se trouve la Sibérie Occidentale dont les plaines s'étendent des Montagnes de l'Oural au Ienisseï ; faisant suite à cette partie les plateaux de Sibérie Orientale couvrent la zone allant du Ienisseï jusqu'à la Lena qui marque la limite avec les territoires d'extrême Orient. Ceux-ci comprennent la vallée de l'Amour, la façade Pacifique, la péninsule du Kamchatka ainsi que le détroit de Béring. La Yakoutie couvre environ un quart de cet ensemble et représente la quasi-totalité de la Sibérie Orientale.

La République Sakha, ou Yakoutie, est la plus grande des 21 Républiques autonomes qui constituent aujourd'hui la Fédération de Russie avec une superficie de 3 103 200 km². La Yakoutie s'étend sur 2000 km d'est en ouest, des Plateaux de Sibérie Centrale aux Montagnes de Verkhoïansk, couvrant ainsi les bassins de cinq grands fleuves, l'Olenek, la Léna, la Yana, l'Indighirka et la Kolyma. La République Sakha est bordée au sud par la Mongolie et la Chine et 2500 km au nord par la Mer de Laptev et la Mer de Sibérie Orientale, 60% du territoire est situé dans le grand nord et 40 % se trouve au delà du cercle polaire. La masse continentale engendre un climat froid et sec ainsi qu'une amplitude thermique extrême puisque les températures sont fréquemment de l'ordre de -50°C en hiver en Yakoutie Centrale, et peuvent atteindre -70°C dans la région de Verkhoïansk, alors qu'en été il est possible de rencontrer des températures avoisinant les 40°C. Ces conditions climatiques particulièrement difficiles entraînent la présence d'une taïga dense, composée essentiellement de mélèzes et de bouleaux sur la plus grande partie du territoire excepté les régions arctiques où la taïga fait place à la toundra. Malgré ce climat extrêmement, rigoureux des groupes humains, dont l'origine et les dates de peuplement diffèrent, sont présents dans l'ensemble des régions de Sibérie.

Il est possible de scinder la Yakoutie en trois grandes régions en fonction de critères linguistiques et par conséquent des populations qui occupent le territoire, ou en tenant compte de la géographie physique. Selon Vorokin (1999), il existe trois groupes dialectaux distincts apparus au gré des contacts entre populations de langue différente et dont la répartition se

calque sur les séparations géographiques. Ces particularités linguistiques seraient par ailleurs retrouvées au niveau génétique (Boeva, 1988). (i) La partie comprenant la Vilyuy et la région d'Olékmansk dont une fraction importante de la population et représentée par les Evenks. (ii) Au delà du cercle polaire, le nord et le nord-est de la Yakoutie où les populations Evènes, mais aussi certains groupes minoritaires tels les Youkaguirs ou les Tchouktches, partagent la toundra avec les Yakoutes. (iii) Enfin, la Yakoutie Centrale qui fera l'objet de cette étude. Cette région est représentée par le triangle des fleuves Lena-Aldan-Amga et les régions mitoyennes. La Yakoutie Centrale présente la plus forte proportion de Yakoutes et se caractérise par des zones favorables à l'élevage du bétail.

De manière administrative, la République Sakha est divisée en 33 régions, ou *ulus*. Ce terme qui fut largement adopté par les Yakoutes n'était pas utilisé avant l'arrivée des russes (Gogolev, 1993). Ce découpage administratif date du XIX^{ième} siècle et fait suite à de nombreux remaniements qui se sont succédés depuis le XVII^{ième} siècle. Ces découpages administratifs étaient motivés par la volonté des colons russes de faciliter les recensements en vue de la collecte du *yassak* (impôt prélevé par les Russes sous forme de fourrures). Le premier fut effectué en 1642, il concernait les Hommes mais également le bétail afin d'estimer les ressources des paysans yakoutes.

I.1.2 L'*alas* : une particularité du paysage de la Yakoutie Centrale

La région comprise entre la rive est de la Lena et le fleuve Amga est principalement formée par des plateaux de basse altitude correspondant à d'anciennes terrasses alluviales de la Lena. Le réchauffement ayant suivi le Dernier Maximum Glaciaire a entraîné la formation d'un paysage très particulier, résultant de l'abaissement du niveau du permafrost. Ces bouleversements climatiques ont favorisé la formation d'effondrements qui, en se remplissant d'eau, vont finir par se rejoindre pour former des ensembles atteignant parfois plusieurs kilomètres carrés. L'absence de nappe phréatique sous jacente combinée à un abaissement du permafrost peut conduire à l'assèchement de la zone créant ainsi de vastes aires déboisées et l'apparition de pingos, ou *bulgunnjakh*, qui sont des collines de faible hauteur possédant un noyau de glace et sur lesquelles les inhumations pouvaient être pratiquées. Ce processus se déroulant en trois phases successives, s'étale généralement sur une période de 6000 ans (Harada *et al.*, 2006).



Figure 1 : Vue aérienne de la succession de différents *alas* dans la région de Churapcha (photo Annie Géraut, 2006).

Ces grandes dépressions, appelées *alas* (Figure 1), possèdent des conditions propices au développement d'une végétation herbacée qui furent mises à profit par les Yakoutes pour le pâturage des troupeaux ainsi que la fenaison. Ce phénomène thermokarstique particulièrement fréquent en Yakoutie Centrale (Brouchkov *et al.*, 2004) aurait conduit à la formation d'un terroir permettant le développement de l'élevage et pourrait, par conséquent, avoir favorisé la fixation dans cette zone des populations d'éleveurs ayant migré depuis les régions au sud. De plus, la densité ainsi que l'ampleur des *alas* était responsable de la richesse en ressources pastorales des propriétaires terriens de la région ce qui aurait pu avoir une incidence sur leur statut social.

L'*alas* constituait traditionnellement le territoire d'une famille, le plus souvent petite et monogame (parfois polygame), composée des parents et de leurs enfants. La famille patriarcale était la cellule de base de la société traditionnelle chez les Sakhas. La grande famille patriarcale, composée des parents, des fils et de leurs familles, et parfois de serfs voire d'esclaves, était plus rare. Celle-ci était généralement le fait d'un *tojon* (chef de clan) et comprenait ses frères, ses fils, ses neveux ainsi que leurs femmes et leurs enfants. Dans le cas des grandes familles polygames, le domaine familial pouvait s'étendre sur plusieurs *alas* (entre deux et quatre), chaque épouse devenant alors responsable d'une partie du cheptel de la famille (Le Berre-Semenov, 2000). Il est possible d'imaginer que les terroirs dont la

concentration en *alas* était particulièrement importante auraient pu favoriser l'expansion de certains clans et, par conséquent, de certaines lignées paternelles. L'*ulus* de Churapcha, dans lequel nous avons réalisé un nombre de fouilles et de prélèvements génétiques important, est unique du point de vue de la densité des *alas*.

I.1.3 Le permafrost

Le permafrost, ou pergélisol, représente la fraction du sol qui ne dégèle pas durant l'année. L'épaisseur de cette couche peut varier suivant les zones de 0 à plus de 1500 m et la température varie entre 0°C, dans les zones les plus au sud, à -15°C, dans les régions les plus septentrionales (Fotiev, 1997). Malgré le réchauffement climatique actuel qui a tendance à modifier la répartition, la majorité du territoire de la République Sakha est occupée par une zone de permafrost continu qui ne dégèle pas durant l'année (Kitabata *et al.*, 2006) (cf figure 2). Lors des fouilles réalisées durant ces trois années, nous avons pu observer que la profondeur à laquelle le sol était gelé variait de 50 cm à un mètre en fonction par exemple de l'orientation ou de l'exposition au soleil du site.

Le permafrost représente de manière évidente un obstacle pour les inhumations puisqu'il paraît difficilement concevable que des populations ne disposant que d'outils rudimentaires aient pu réaliser des fosses alors qu'il est impossible d'attaquer la couche gelée à l'aide de pelles en métal. Les populations devaient par conséquent soit stocker les corps durant l'hiver en attendant que la surface du sol dégèle, soit recourir à des bûchers afin de faire fondre une épaisseur suffisante de sol. Lors des fouilles que nous avons menés, les coffres sont par conséquent fréquemment retrouvés posés sur le permafrost. Cette particularité présente de nombreux avantages quant à la conservation des corps et du matériel archéologiques. Les spécificités du permafrost sont également très favorables à la conservation de l'ADN au cours du temps (Willerslev *et al.*, 2004b).

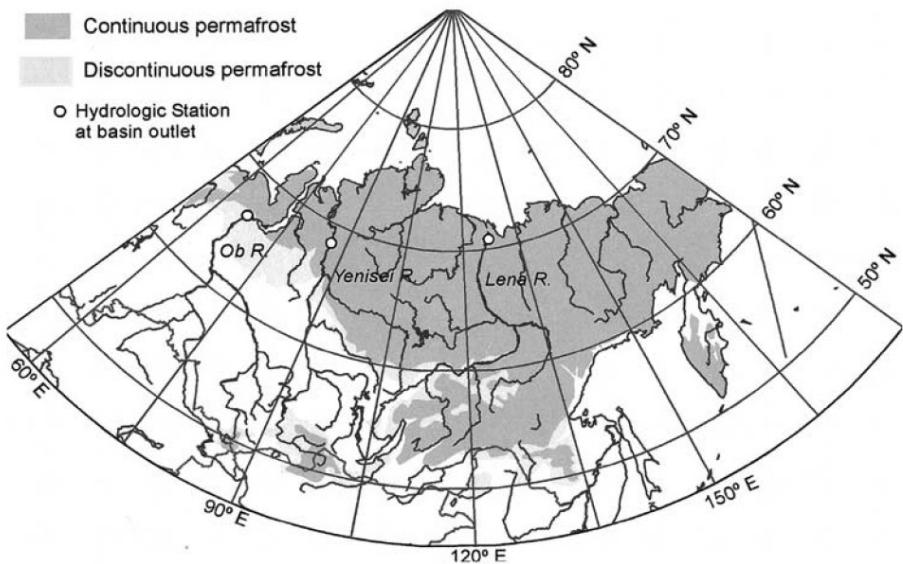


Figure 2 : Distribution du permafrost continu et discontinu en Sibérie (d'après Yang *et al.*, 2002).

En effet, la température et en particulier les variations au cours du temps sont, comme nous le verrons dans le chapitre sur les dégradations de l'ADN, l'un des principaux facteurs délétère pour l'ADN. Or le permafrost va maintenir le corps à une température quasi-constante, proche de 0°C. Lors des fouilles, nous avons observé à de nombreuses reprises la présence de glace à l'intérieur des coffres indiquant que, même durant la période estivale, le corps est maintenu à une température constante (cf figure 3). En outre, le permafrost est un environnement relativement sec puisque la plupart de l'eau est présente sous forme solide (92/97%) (Gilichinsky, 2002). Les dommages hydrolytiques vont par conséquent être diminués. De plus, une des particularités du permafrost rencontré en Yakoutie est la concentration en méthane ainsi que son potentiel d'oxydoréduction qui suggèrent des conditions anaérobies. Les dommages oxydatifs seraient par conséquent réduits. Enfin, son pH est neutre ce qui permettrait de diminuer les risques d'alkylation qui sont plus fréquemment rencontrés dans le permafrost présent en Antarctique en raison du pH légèrement alcalin (Gilichinsky, 2002).

L'ensemble des ces propriétés font du permafrost un environnement particulièrement favorable à la préservation de l'ADN au cours du temps. La Yakoutie apparaît par conséquent comme une zone géographique particulièrement adaptée au développement d'une approche paléogénétique nécessitant l'analyse d'un nombre important d'échantillons anciens.



Figure 3 : Tombe de Jarama 1 découverte en sous bois lors de l'expédition MAFSO 2004. On note la présence de glace sur les deux tiers inférieurs du corps.

I.2 Des premiers peuplements à la colonisation russe

Comme nombre de peuples de Sibérie, les Yakoutes ont une tradition essentiellement orale et il n'existe par conséquent aucune trace écrite de leur histoire. Les premiers rapports historiques n'apparaissent que tardivement avec l'arrivée des colons russes au XVII^{ème} siècle et les informations concernant les périodes antérieures sont bien souvent fragmentaires. De plus, les données des études archéologiques sont difficilement exploitables du fait des datations imprécises voire fantaisistes et/ou de la méthodologie employée. Cependant, les recherches effectuées ces dernières années par des archéologues locaux mais également par des équipes internationales d'archéologues et de généticiens permettent d'éclaircir quelque peu la période antérieure à la colonisation.

I.2.1 Premiers peuplements de la Yakoutie

Les premières occupations humaines sur le territoire de ce qui est aujourd'hui la Yakoutie sont certainement à mettre en relation avec les populations de chasseurs du Paléolithique. Du fait de l'étendue du territoire et des conditions climatiques, les recherches archéologiques restent difficiles, aussi peu de sites très anciens ont été décrits et la chronologie demeure très fragmentée.

Le site le plus ancien qui atteste d'une occupation humaine a été décrit en 2004 par Pitulko *et al.* et se situe dans le nord de la Yakoutie, au dessus du cercle polaire (cf figure 4). Ce site, nommé Yana RHS en raison de sa localisation sur les rives de la rivière Yana, est datée de 27 000 ans. L'industrie lithique mise au jour sur le site regroupe de nombreux outils dont la typologie diffère des industries ultérieures par l'absence de fabrication de lames et lamelles. Le site de Yana RHS eut un impact important dans le domaine archéologique puisque, pour la première fois, la présence humaine dans l'Extrême nord était attestée avant le Dernier Maximum Glaciaire. Les sites de Berelekh (Mochanov, 1977) dans la région de Verkhoïansk et Dyuktaï (Mochanov, 1977) dans la région de l'Aldan, sont datés aux environs de 11 000 à 13 000 ans (cf figure 4). Ces deux sites sont caractéristiques de la culture Dyuktaï qui est spécifique du nord de la Sibérie et se caractérise, en plus des outils bifaciaux, par la présence d'outils lamellaires.

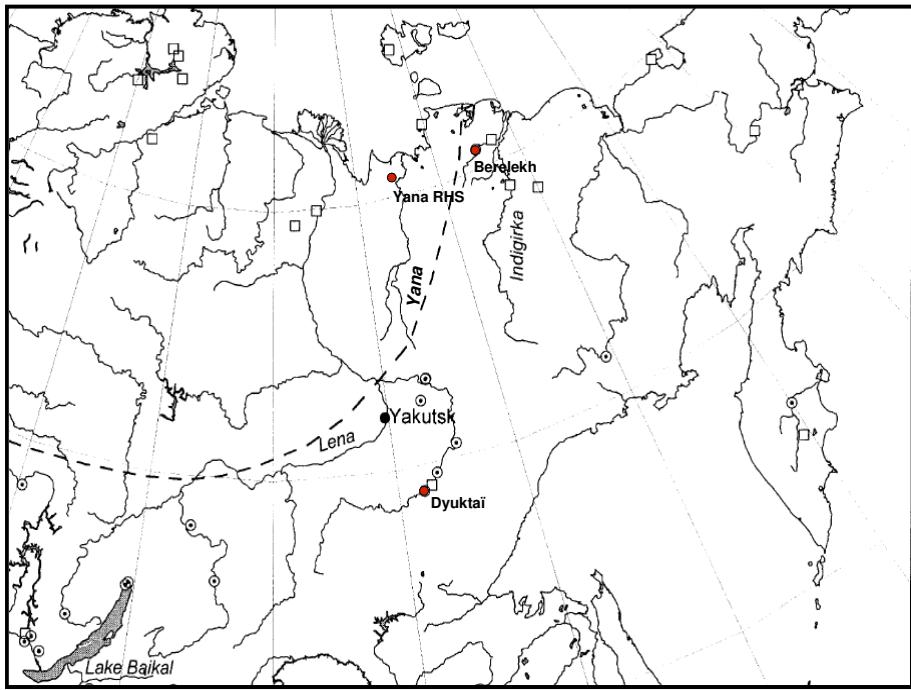


Figure 4 : Localisation des sites Paléolithiques mis au jour en Yakoutie (modifié d'après Kuzmin *et al.*, 2001).

I.2.2 Pré et protohistoire

Du fait de sa position géographique, la Sibérie Occidentale semble avoir été occupée depuis le Paléolithique aussi bien par les populations ouest et est eurasiatiques (Okladnikov, 1964). La région du lac Baïkal, souvent proposée comme berceau de la population yakoute actuelle, a été largement étudiée du fait du nombre de sites archéologiques présents et de sa position charnière entre la Sibérie Occidentale et Orientale.

Dans la région péri-Baïkale, correspondant au nord et à l'ouest du lac Baïkal, la culture matérielle durant le Paléolithique Supérieur présente des similitudes avec les populations d'Europe de l'est (Okladnikov, 1959, 1964). De plus, les rares données ostéologiques (Turner, 1987 ; Ishida et Dodo, 1996) ainsi que les représentations artistiques suggèrent que les habitants de l'est de la Sibérie était d'ascendance est-eurasiatique (hypothèse à considérer toutefois avec circonspection).

Les cultures Kitoi (transition Mésolithique/Néolithique : 7000-5000 ans BP) ainsi que Serovo-Glazkovo (Néolithique : 4200-3200 BP) de la région du lac Baïkal sont en revanche

un peu mieux documentées du point de vue archéologique et génétique. Les Kitois étaient caractérisés par un mode de subsistance quasi-exclusivement lié à la pêche et par des pratiques funéraires comprenant l'utilisation d'ocre lors des inhumations (Bazaliiskiy et Savelyev, 2003 *in* Moodyer *et al.*, 2006). La culture Serovo-Glaskovo, qui est séparée de la précédente par un hiatus d'environ 800 ans, diffère par le mobilier associé au corps ainsi que par les pratiques funéraires qui incluent l'utilisation de feu et une utilisation d'ocre plus occasionnelle (Tiutrin et Bazaliiskii, 1996 ; Weber *et al.*, 2002 *in* Moodyer *et al.*, 2006).

La séparation entre ces deux périodes serait à mettre en relation avec une dégradation climatique durant le milieu du Néolithique qui aurait entraîné une raréfaction des ressources. Les populations de la culture Kitoi auraient été contraintes de quitter la région du lac Baïkal pour rejoindre le cours moyen de l'Ob et du Ienisseï conduisant, respectivement, à l'apparition des Shorians et des Kets (Moodyer *et al.*, 2006). La population de la culture Serovo-Glaskovo se serait en revanche maintenue dans la région du Baïkal et il semble que les lignées maternelles présentes dans cette population aient perdurées au moins jusqu'à la période Xiongnu. En effet, des similitudes importantes ont été retrouvées entre les sujets étudiés par Moodyer *et al.* (2006) et les Xiongnu de la nécropole d'Egyin Gol (Keyser-Tracqui *et al.*, 2003) dans le nord de la Mongolie.

Durant le Néolithique, il semble que d'importants mouvements de populations soient survenus, en particulier concernant les anciennes populations eurasiatiques qui se seraient déplacées vers le sud et le centre de la Sibérie (Levin-Potapov, 1964). Par la suite, durant les deux derniers millénaires avant notre ère, les expansions turcs, mongoles et toungousses semblent avoir provoqué le déplacement des groupes autochtones moins nombreux vers les zones périphériques de leurs aires de répartition initiale ou avoir causé leur assimilation au sein des cultures dominantes (Levin-Potapov, 1964). Ces changements auraient pu conduire à l'apparition de nouveaux groupes ethniques de manière relativement récente et dont les influences culturelles seraient variées (Schurr, 2004).

I.2.3 Le Moyen-Âge

La langue et le mode de vie des Yakoutes rapprochent indéniablement cette population des peuples des steppes du sud de la Yakoutie. Cependant, la localisation précise de cette origine

et les modalités concernant les dates, le nombre de vagues de migrations ainsi que l'itinéraire suivi restent incertains.

Deux hypothèses principales peuvent être considérées pour retracer l'ethnogenèse des Sakhas.

Selon l'hypothèse la plus ancienne, proposée à la fin du XIX^{ième} siècle par WL Siéroszewski (1973, *in* : Le Berre-Semenov 2002), un exilé politique d'origine polonaise, les ancêtres des Yakoutes seraient arrivés dans la zone correspondant à la Yakoutie Centrale, au niveau du cours moyen de la Lena, et auraient été contraints de se fixer sur ce terroir en raison de l'occupation toungouse alentour. L'expansion des Yakoutes ayant conduit à leur domaine de répartition actuel serait à mettre en relation avec l'arrivée des Russes au XVII^{ième} siècle. Les épidémies, notamment de variole, qui suivirent l'arrivée des colons affectèrent grandement les populations autochtones entraînant un dépeuplement quasi-total de certaines régions. Ces déprises démographiques pourraient être à l'origine de l'expansion des Sakhas (Forsyth *et al.*, 1992). La dichotomie actuelle entre les Sakhas du centre, de la Vilyuy et du nord aurait eu lieu à cette période (*in* Le Berre-Semenov, 2002).

Selon une autre hypothèse formulée par G Ksenofontov, puis reprise plus tard par AP Okladnikov, les premières populations venant du sud seraient arrivées au I^{er} siècle de notre ère depuis la région du péri-Baïkal (partie ouest du lac Baïkal correspondant à la région d'Irkoutsk). Selon cet auteur, ces tribus d'éleveurs de rennes, appelées *Uraangkhaj*, seraient d'origine toungousse et auraient subit l'influence des populations turks et mongoles. Les Xiongnus, qui occupaient à partir du IV^{ième} siècle av JC un vaste territoire incluant la majorité de la région du lac Baïkal y compris la région du péri-Baïkal, pourraient correspondre aux populations turks mentionnées par Ksenofontov. Cette première vague de nomades aurait par la suite été partiellement assimilée et/ou partiellement repoussée vers le nord et le fleuve Olenek par une vague d'éleveurs de bétail nommés *Kurykans*, une tribu turk apparentée aux Ouïgours. Ces derniers se seraient ensuite déplacés puis fixés en Yakoutie Centrale qui présente, comme nous l'avons décrit plus haut, certaines particularités géographiques favorables à une économie d'élevage. Pour AP Okladnikov, la migration des *Kurykans* depuis la région du péri-Baïkal constitue le point de départ de la formation du peuple sakha. L'arrivée des ces populations d'éleveurs des steppes du sud en Yakoutie Centrale entre le X^{ième} et le XVI^{ième} siècle, leur adaptation aux conditions climatiques et les contacts avec les populations autochtones d'origine toungousse auraient conduit à l'émergence des Yakoutes.

D'après des fouilles réalisées par l'historien A Gogolev (2000), l'apparition de la société yakoute telle qu'elle fut décrite au XVII^{ième} siècle serait à mettre en relation avec la culture de

Kulun-Atakh du XIV^{ième} siècle et localisée sur le cours moyen de la Lena. Cette culture se diviserait en deux phases : une précoce allant du XIV^{ième} au XV^{ième} siècle et une plus tardive allant jusqu’au XVI^{ième} siècle. L’habitat retrouvé dans ces sites était présent sous forme de locus regroupant des ossements de chevaux et de bovins. Les dessins ainsi que les formes présentes sur les céramiques évoquent, d’une part, le style rencontré entre le VI^{ième} et le X^{ième} siècle chez les *Kurykans* de l’ouest du Baïkal et préfigurent d’autre part les motifs rencontrés dans la culture yakoute. La population à l’origine de la culture Kulun-Atakh aurait adapté l’élevage des chevaux et des bovins aux conditions climatiques de la région de la Lena et leurs contacts avec les populations autochtones auraient conduit à la formation de la culture yakoute traditionnelle.

Au nord de la culture de Kulun-Atakh, dans la région de contact entre la Vilyuy et la Lena, l’archéologue russe AP Okladnikov mit en évidence dans les années 1940 une autre culture pouvant être en relation avec l’émergence des Sakhas. Cette culture dite des petites maisons ou *Malyški Doma*, dont la datation mal définie serait comprise entre le XIII^{ième} et le XVI^{ième} siècle, pourrait également marquer une assimilation des tribus locales par des populations arrivant du sud.

Les travaux de la Mission Archéologique Française en Sibérie Orientale (MAFSO) ont permis d’apporter de nouveaux éléments concernant le peuple sakha durant la période médiévale. En effet, au cours des fouilles réalisées durant ces dernières années l’examen des pratiques funéraires a révélé deux traditions distinctes dont les particularités seront détaillées ultérieurement (*III.1.2.2 Les datations*). De manière succincte nous avons pu distinguer : une première tradition dite turco-mongole qui semble ancienne (XIV-XV^{ième} siècles) et pour laquelle l’essentiel des sujets mis au jour sont masculins. Des éléments de harnachement (étriers, selle, mors) sont généralement associés au corps et aucun mobilier d’importation (perles ou vêtements en soie) n’est observé. La seconde période se distingue par l’absence de selle et la présence de nombreuses perles ainsi que d’autres éléments d’importation chinoise notamment. Ces deux périodes soulèvent la question des contacts entre la population yakoute, dite traditionnelle, et les populations de Mongolie, d’Asie Centrale et de Chine après leur arrivée en Yakoutie.

I.2.4 La colonisation russe

Les prémisses de la colonisation des étendues sibériennes par les Russes s'amorcèrent dans le milieu du XVI^{ème} siècle sous le règne d'Ivan IV qui s'était proclamé Empereur de Sibérie suite à l'annexion de la région de la rivière Kama, dans les Montagnes de l'Oural. L'expansion au-delà de cette région commença vraiment en 1582 et notamment avec les incursions de cosaques menés par Yermak, figure emblématique des nationalistes russes. En 1620, l'annexion de la Sibérie Occidentale était complète et les Russes contrôlaient un territoire s'étendant des Montagnes de l'Altai jusqu'au Ienisseï. La même année, l'expédition dite de Penda partit de Touroukhansk sur l'Ienisseï et remonta le cours de la Toungouska inférieure jusqu'à la Lena. Penda descendit le cours de la Lena jusqu'à la région de l'actuelle Yakoutsk et arrêta sa progression vers le nord pour rebrousser chemin et remonter la Lena jusqu'à la région du Baïkal. Le retour vers Touroukhansk s'effectua à travers les steppes de Bouriatie jusqu'à la Toungouska supérieure puis le Ienisseï. La colonisation de la région de la Lena ne fut officielle que dix ans plus tard avec la construction du premier fort de Yakoutsk en 1632 sous le règne de Michel III.

Les motivations de cette expansion vers les régions sibériennes étaient principalement d'ordre économique et notamment en raison de l'engouement pour les fourrures. Ces territoires riches en animaux sauvages représentaient par conséquent une source importante de peaux. Dès le XVI^{ème} siècle les russes mirent en place le *yassak*, un impôt à payer en fourrures, auquel étaient soumises toutes les populations autochtones. Pour une même famille, l'ensemble des sujets masculins âgés de plus de quinze ans devaient fournir une fois par an un nombre fixé de fourrures de zibeline. Chez les Yakoutes, il semble que le nombre de fourrures était fonction du nombre de têtes de bétail possédées par la famille (Forsyth, 1992).

Il est attesté que, durant les quarante premières années qui suivirent le début de la colonisation de la Yakoutie, les révoltes menées par les *tojons* furent écrasées d'une main de fer par les représentants de l'autorité tsaristes. Les nombreuses exactions combinées à la pression du paiement du *yassak* entraînèrent des mouvements de migrations des Sakhas depuis la Lena vers les régions de la Vilyuy et vers le cercle polaire, notamment vers les fleuves Olenek et Anabar dans le nord-ouest. Lors de ces expansions, les contacts avec les populations toungousses conduisirent à la formation de populations fortement métissées telles les Dolganes qui parlent un dialecte yakoute et dont le mode de vie se rapproche des toungousses.

Les données concernant les rapports entre les Russes et les Sakhas sont néanmoins difficiles à cerner. Il semble, en effet, que la colonisation ait pu profiter à une certaine partie de la population locale. Les décrets mis en place en 1677-78 concernant l'autorité des chefs de clans et leur responsabilité lors de la collecte des fourrures pour le *yassak* renforçaient grandement l'influence de cette noblesse locale. Ces premiers décrets furent suivis en 1733, après l'abolition de l'esclavage par l'Impératrice Anna, par la confirmation officielle du rôle d'administrateurs locaux et de juges des *tojons*. De plus, l'attribution des terres se faisait en fonction de la quantité de fourrures remises en paiement du *yassak*. Les chefs de clans purent donc étendre leurs terres au détriment des plus pauvres car ils fournissaient alors le plus de peaux de zibeline. Ce renforcement de l'autorité des *tojons* aurait favorisé l'émergence d'une aristocratie locale qui a sans doute pu étendre son influence à cette époque. De plus, au cours du XVIII^{ième} siècle la langue yakoute a connu une expansion importante et au début du XIX^{ième} elle avait été adoptée dans la majeure partie de la Sibérie Orientale, y compris par les officiels russes habitant à Yakoutsk (Forsyth, 1992).

Une autre conséquence de la colonisation russe fut la christianisation des populations de tradition chamanique et animiste qui occupaient les territoires conquis. Au XVIII^{ième} siècle la quasi-totalité de la population Yakoute avait été baptisée mais les signes effectifs de ces conversions ne sont visibles qu'à partir du XIX^{ième} siècle, excepté en ce qui concerne l'adoption de patronymes d'origines russes. En effet, la conversion à la foi orthodoxe était principalement motivée par l'exemption de paiement du *yassak* pendant les trois ans suivant le baptême. La plupart des Yakoutes conservèrent ainsi leurs traditions chamaniques et la pratique des rites en rapport avec leur cosmologie sur les sites sacrés.

Il paraît assuré que les Sakhas existaient déjà en tant qu'entité avant l'arrivée des Russes, toutefois leur aire de répartition et de nombreux changements se produisirent après la colonisation. Il est donc important de garder à l'esprit qu'une partie de l'ethnogenèse yakoute aurait pu avoir lieu de manière concomitante à l'expansion tsariste du XVII^{ième} et XVIII^{ième} siècles.

I.3 Le peuplement actuel : caractéristiques et répartition

Selon le dernier recensement effectué en 2002, la République Sakha compte actuellement 949 280 d'habitants dont 41,15% de russes et 59,85% de populations autochtones. Celles-ci comprennent 45,54% de Yakoutes qui forment l'ethnie majoritaire et 14,31% de groupes minoritaires représentés par les populations toungousses (Evenks, Evènes, et Dolganes), les Youkaguirs et les Tchouktches. Ces différents groupes ethniques ne sont pas répartis de manière homogène sur le territoire et il est vraisemblable que différents mouvements de populations se sont produits en raison des expansions de certains groupes, notamment les Yakoutes, et l'arrivée des colons russes.

I.3.1 Tchouktches et Youkaguirs

Sur le territoire de la République Sakha vivent actuellement environ 600 Tchouktches et un peu plus de 500 Youkaguirs. Les Tchouktches et les Youkaguirs peuvent être considérés comme des populations reliques dans la mesure où leurs effectifs sont extrêmement restreints et compte tenu du métissage important des populations vivant en Yakoutie avec les ethnies toungousses.

Les Tchouktches sont l'un des plus anciens peuples de la Sibérie Orientale et leurs origines semblent provenir de populations de chasseurs cueilleurs venus de Sibérie Orientale à la fin du Néolithique. Leurs déplacements vers la péninsule du Tchoukotka ainsi que l'assimilation des populations d'esquimaux auraient entraîné la formation d'un groupe sédentaire. Le passage à l'élevage de rennes, dont la datation est mal assurée (entre le VIII^{ième} et le XVI^{ième} siècle), aurait pu être responsable d'une migration en retour vers l'ouest en raison de la raréfaction des pâturages liée à un augmentation de la taille des troupeaux durant le XIX^{ième} siècle (Le Berre-Semenov, 2002).

Il semble que les Youkaguirs aient occupé un territoire assez vaste au cours du XVII^{ième} siècle dans tout le nord de la Yakoutie (cf figure 5) : de la Lena jusqu'à l'Anadyr dans la Péninsule du Tchoukotka. Cependant, les expansions sakhas puis russes auraient conduit à l'assimilation et parfois l'élimination de cette population qui fut alors repoussée dans des zones refuges sur les rives de l'océan glacial arctique et au delà des monts de Verkhoïansk, vers les sources de

l'Indigirka et de la Kolyma. L'ethnogenèse de ces populations reste ambiguë et il est difficile de déterminer si ces populations sont le fait d'une classification réalisée par les russes ou si les Youkaguirs constituaient une population clairement individualisée. Dans ce cas les relations avec les Yakoutes seraient à rechercher considérant leur aire de répartition qui concorde largement avec celles des Sakhas (cf figure 5).

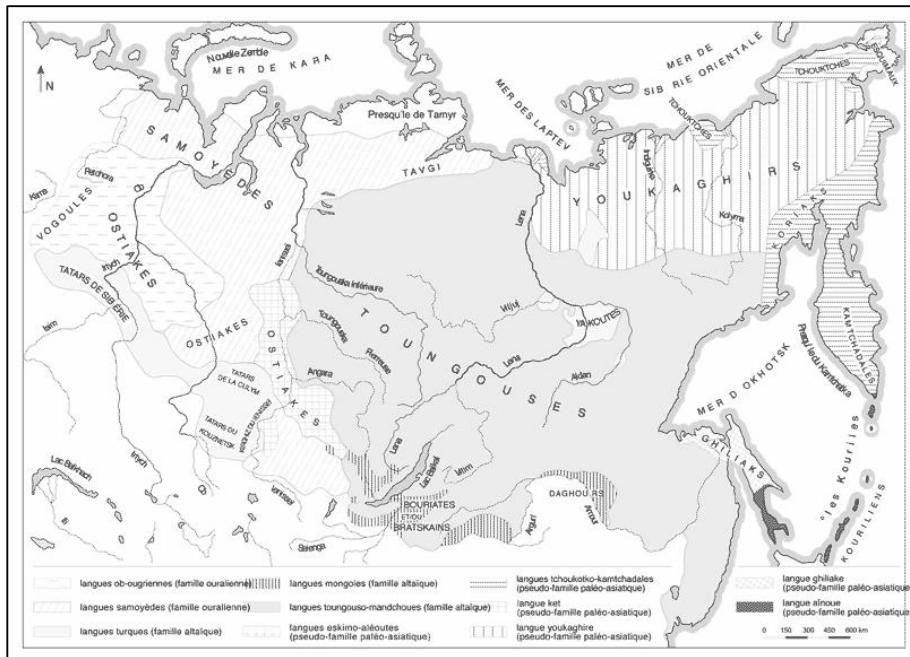


Figure 5 : Répartition des populations autochtones en Sibérie Orientale au XVII^e siècle.

I.3.2 Les populations toungousses : Evenks et Evènes

Les Evenks représentent aujourd’hui le peuple le plus dispersé en Sibérie puisqu’on rencontre des foyers d’occupation depuis le Ienisseï à l’ouest jusqu’aux côtes de la mer d’Okhotsk et des steppes du sud à l’océan glacial arctique (cf Figure 6). L’Arrondissement Autonome des Evenks se situe au nord de l’Angara et à l’ouest du Ienisseï. Cependant en plus de cette localisation administrative et de la population vivant en Yakoutie, il existe une communauté importante dans le nord-est de la Chine, en Mongolie intérieure ainsi qu’en Mandchourie. Les Evenks appartiennent à la famille linguistique toungouss-mandchou. La langue evenk est très diversifiée et peut être séparée en trois dialectes majeurs : septentrional, méridional et oriental, eux même divisés en sous dialectes. L’origine de cette population fait l’objet d’une

polémique quant à sa localisation et deux positions s'affrontent. De manière synthétique, les travaux d'Okladnikov mettent en avant deux vagues de migration depuis la région du Baïkal lors du premier siècle AD suite à la pression des populations Xiongnu, puis vers le X^{ième} siècle en raison des expansions des tribus mongoles conduisant à la formation des groupes actuels. La deuxième théorie, qui apparaît chez Shirokogoroff (1923), fait venir les proto-toungousses du nord de la Chine actuelle (région de la Mandchourie et de la région de l'Amour). Cependant, la répartition et les différences observées de nos jours sont, selon A. Lavrillier, à mettre en relation avec leur histoire et les migrations constantes de ce peuple. Lors de ces déplacements, les populations toungousses se seraient trouvées en contact avec les peuples autochtones tels les Youkaguir, les Koryaks ou les Tchouktches qui auraient alors été en partie déplacés vers le nord. Il semblerait qu'au cours de ce processus, la langue toungouse ait prévalu, tandis que leurs descendants adoptaient la culture des chasseurs-pêcheurs de la taïga et héritaient du type anthropologique des populations dites "Paléoasiatiques". Finalement, au cours des siècles, ces mélanges de populations dans l'est de la Yakoutie auraient donné lieu à une différenciation des Toungouses en deux groupes ethnographiques distincts : les Evenks et les Evènes (Vasilievich 1969 *in* Le Berre-Semenov 2002).

Les Evènes possèdent un mode de vie nomade principalement basé sur l'élevage de rennes, la chasse et la pêche. Toutefois, les tribus vivant sur les côtes de la mer d'Okhotsk sont devenus sédentaires et pratiquent la pêche ainsi que la chasse des mammifères marins révélant ainsi l'influence des populations Tchouktches et Youkaguir.

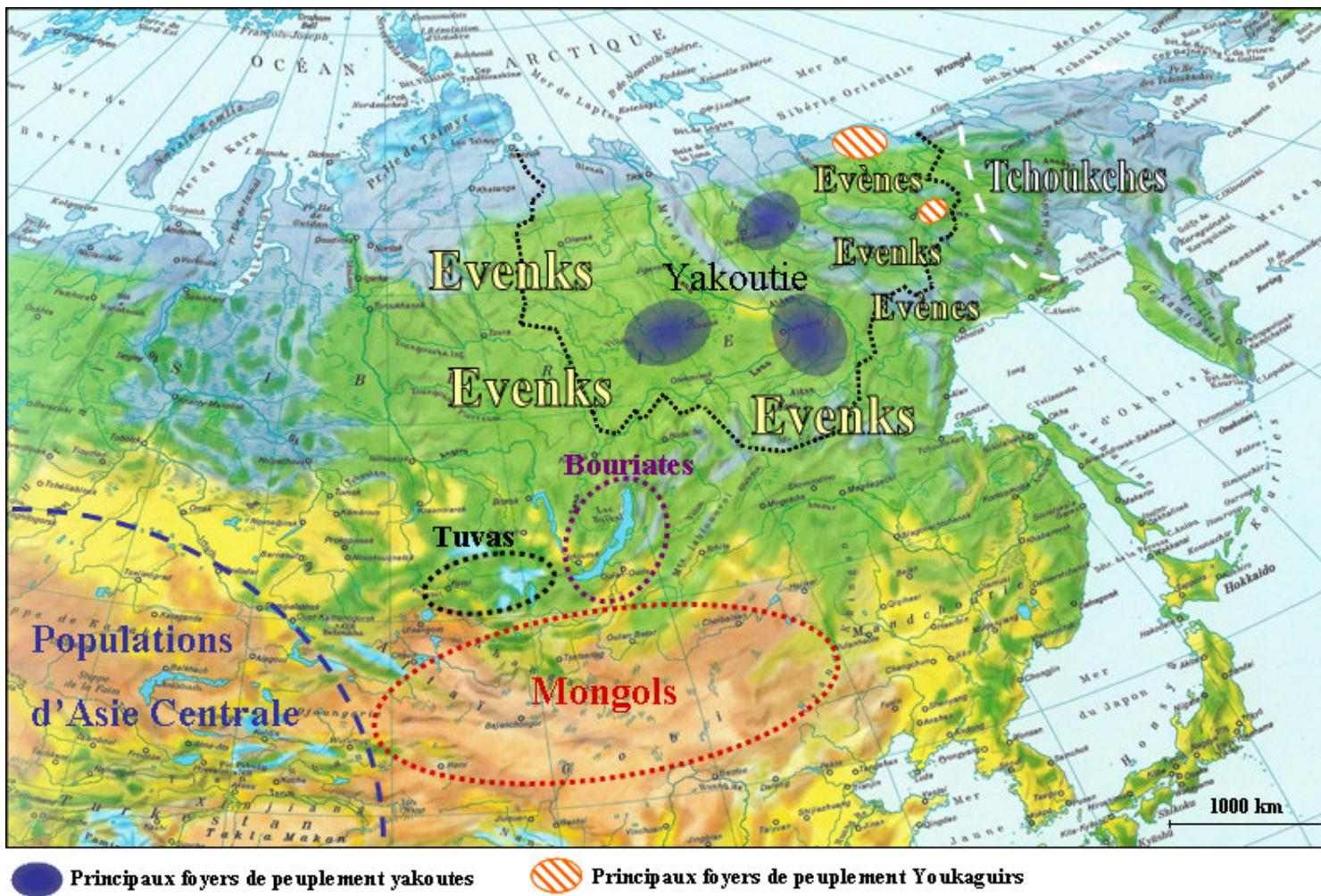


Figure 6 : Aire de répartition schématique des principales populations de Sibérie Orientale.

I.3.3 Les Yakoutes

Le nom "yakoute" correspond à la dénomination donnée par les russes à ce peuple et provient certainement de la déformation d'un mot d'origine toungousse *yeket* (Forsyth, 1992). Le terme yakoute désignait, pour les russes, l'ensemble des éleveurs de bétail et de chevaux. Les Yakoutes quant à eux se nomment traditionnellement *Sakha*.

Les Yakoutes apparaissent comme un peuple atypique qui contraste par sa culture et son mode de vie avec les populations environnantes. Selon Okladnikov (1972) "*les Yakoutes ont fasciné les érudits pendant 300 ans en raison de leur langue türk, leur économie d'élevage de bétail et surtout leur culture qui étaient totalement déplacées dans cet environnement subarctique et étrangères aux populations environnantes*". Les Sakhas vivent dans un environnement extrêmement froid qui ne semble que peu propice à l'élevage des bovidés et des chevaux. Ils représentent d'ailleurs le seul exemple d'adaptation de ce type d'économie sous des latitudes aussi septentrionales. La chasse et la pêche peuvent néanmoins occuper, suivant les lieux ou les époques, une part non négligeable de leurs modes de subsistance.

Les Sakhas parlent une langue appartenant à la famille des langues türks (comme les populations de l'Altaï, les Kazakhs, Kirghiz, Tuvas et Ouigours) (Balzer 1994) de part sa grammaire ainsi que son vocabulaire mais présente également l'influence des langues mongoles et même toungousses (Forsyth, 1992). Environ 60% de mots sont originaires des langues türks (avec une affinité marquée pour la langue des Tuvas) et 40% d'origine Mongole/Bouriate (langue mongole) (Schönig 1990).

Les traditions et croyances des Yakoutes reflètent également cette influence des populations des steppes du sud. La fête de l'*Ysyakh*, célébrée durant neuf jours et devenue aujourd'hui fête nationale de la République Sakha, n'est pas sans rappeler le culte tengrique des Huns (adoration du ciel et des feux célestes). Au cours de l'*Yhyakh*, des prières incantatoires (*algys*) sont adressées par des "chamanes blancs" aux divinités célestes (*Ajyy*), puis aux Esprits de la terre et aux Esprits Protecteurs (Le Berre-Semenov, 2002). Durant cette cérémonie les *olonkhos*, des poèmes épiques de plusieurs milliers de vers, étaient récités retracant l'Elleïade : le mythe de la formation des Sakhas. Cette Elleïade retrace l'arrivée de 3 tribus venues du sud menées par trois chefs de clans : Omogoj, Ellej et Uluu-Khoro. Omogoj, dont l'origine et la langue sont inconnues, fut le premier arrivé et ses descendants occuperaient aujourd'hui la rive ouest de la Lena dans les *ulus* de Churapchinsky, Tattinsky et Aldansky. Le second fut Ellej, d'ascendance et de langue türk dont la tribu se serait fixée dans les *ulus*

de Khangalassky, Megino-Khangalassky et Namsky qui se trouve de part et d'autre de la Lena, dans la région de Yakoutsk. Enfin, Uluu-Khoro arriva de l'est mais son peuple étant considéré comme étranger en raison d'une langue différente (non précisée), il se serait dispersé par la suite dans toute la Yakoutie. Pour AP Okladnikov cette Elleïade se reflétait dans les peintures rupestres découvertes sur le site de Shishkino situé sur les rives de la Lena, plusieurs centaines de kilomètres au sud de Yakoutsk. Ces peintures représentent des cavaliers chevauchant debout et brandissant des étendards mais également des vaches ainsi que des chevaux (cf figure 7). Malheureusement ces fresques n'ont pas été datées avec précision et les premières estimations les situant dans le Paléolithique Supérieur ont bien entendu été rejetées (Bednarik et Devlet, 1992).



Figure 7 : Peintures rupestres des falaises du site de Shishkino.

Compte tenu des données archéologiques et ethnologiques, il est aujourd'hui admis par la plupart des spécialistes que les ancêtres des Yakoutes étaient originaires du sud de la Sibérie et qu'une migration vers le nord le long des rives de la Lena les aurait conduit jusqu'en Yakoutie Centrale. Les modalités de l'ethnogenèse telles que la zone de départ précise, le nombre de vagues de migrations ainsi que leurs datations font cependant l'objet de nombreuses divergences entre les chercheurs.

La population yakoute actuelle a fait l'objet d'études sur les marqueurs génétiques dits classiques telles les protéines du sang (Pakendorf *et al.*, 1999), les groupes sanguins (Tarskaïa *et al.*, 2002) ou le système HLA (Gouriev, 2004) mais également sur les marqueurs tels que l'ADN mitochondrial (Derenko *et al.*, 2000 ; Pakendorf *et al.*, 2003, 2006) ou le chromosome Y (Santos *et al.*, 1999 ; Pakendorf *et al.*, 2002, 2006). Ces études sont assez homogènes et confirment des affinités marquées avec les populations à la fois du sud de la Sibérie, Mongolie, Bouriatie et Asie Centrale, mais également avec les populations Evenks (Pakendorf

et al., 2003, 2006). Cependant, il semble exister des différences importantes entre les origines des lignées maternelles, en relation avec les populations d'Asie Centrale et du sud de la Sibérie, et les lignées paternelles dont les particularités sont uniques à la population yakoute et dont l'origine reste énigmatique (Pakendorf *et al.*, 2006).

Les travaux de FX Ricaut sur 11 sujets yakoutes anciens datés du XV^{ième} au XVII^{ième} siècle (Ricaut, 2003) ont permis d'apporter des éléments inédits sur l'origine du peuple sakha. Considérant les lignées maternelles, il semble qu'une certaine continuité ait prévalu puisque la distribution des haplogroupes dans cet échantillon ancien reflète la composition actuelle de la population yakoute moderne. Certains éléments, dont la présence d'un sujet daté du XV^{ième} siècle présentant un haplotype dit ouest eurasiatique, ont permis de nuancer les hypothèses qui avaient été avancées concernant la provenance russe des haplotypes européens retrouvés dans la population actuelle. La nécessité d'une étude sur un échantillon plus large de la population ancienne paraissait, par conséquent, nécessaire afin de préciser les modalités anciennes de la formation des Yakoutes mais également l'influence d'événements plus récents comme la colonisation russe.

I.3.4 Répartition actuelle des populations

Les deux principaux centres de peuplement yakoute sont la Yakoutie Centrale, qui correspond à la zone formée par la Lena, l'Amga et l'Aldan ainsi que la région de la Vilyuy (cf Figure 6). La Yakoutie Centrale totalise presque un tiers de la population yakoute, Yakutsk à elle seule regroupant 15% de la population. La région de la Vilyuy ne compte que 12% des Sakhas, le reste de la population yakoute se répartissant sur l'ensemble du territoire : elle est présente essentiellement dans les régions du nord, dans les Monts de Verkhoïansk, dans la Moyenne-Kolyma au nord-est, dans l'Anabar au nord-ouest et, plus marginalement, au sud de Yakutsk dans la région d'Olekmansk.

La répartition des ethnies minoritaires est également hétérogène sur le territoire. Les Evenks représentent la deuxième ethnie autochtone la plus importante après les Yakoutes suivie par les Evènes. Ils occupent principalement la région occidentale de la Yakoutie, de la Lena dans la région d'Olekmansk jusqu'à la région de l'Anabar en passant par la Vilyuy. Les Evènes se concentrent essentiellement dans le nord-est de la République Sakha dans la région comprenant les montagnes de Verkhoïansk ainsi que la Kolyma. Le peuplement dolgane n'est

qu'anecdotique en Yakoutie et localisé essentiellement dans le nord-ouest de la Yakoutie puisque leur zone de peuplement principal correspond à la péninsule du Taymyr. La population Tchouktche est relativement faible en Yakoutie et se concentre principalement à l'embouchure de la rivière Indigirka ainsi que sur la rive ouest de la Kolyma. Enfin, les Youkaguirs représentent une des plus petites minorités de Sibérie et la plupart vivent sur le territoire yakoute. Ils se divisent en deux groupes les Youkaguirs de la toundra, ou *Vadul* et les Youkaguirs de la taïga, ou *Odul*.

I.4 Problématique et hypothèses sur l'ethnogenèse Yakoute

I.4.1 L'ethnogenèse yakoute : une problématique complexe

Comme nous l'avons évoqué dans les chapitres précédents, la présence de la population yakoute dans l'environnement inhospitalier que constitue la Sibérie Orientale représente une énigme. En effet, les Sakhas dénotent totalement des populations qui les entourent de part leur mode de vie semi-nomade et leur économie basée sur l'élevage des bovins et des chevaux. Leur langue d'origine turco-mongole représente l'un des exemples les plus septentrionaux des langues turciques et contraste avec les dialectes toungousso-mandchous des populations environnantes, Evenks et Evènes. Leurs traditions et leur culture, notamment leur mythologie représentée par l'Elleïade, les séparent également des groupes ethniques alentours les rapprochant des croyances présentes, par exemple, chez les Huns.

L'ensemble de ces spécificités a conduit, de longue date, à les affilier aux populations d'éleveurs nomades des steppes de Mongolie et d'Asie Centrale qui se seraient adaptés par la suite aux rigueurs du climat. Cette adaptation ethnologique hors norme associée à la localisation géographique de la Sibérie Orientale, au carrefour des mondes européen, asiatique et américain font de la Yakoutie un sujet d'étude passionnant.

Les hypothèses archéologiques sur les origines méridionales des Sakhas, présentées dans le chapitre précédent, ont été confrontées aux données analysées sur les populations actuelles dans différentes études (Tarskaia *et al.*, 2002 ; Gouriev *et al.*, 2003 ; Pakendorf *et al.*, 2002, 2003, 2006 ; etc.). Cependant, (i) les différents mouvements de population qui ont pu survenir par le passé, accompagnés ou non de métissage avec les populations autochtones ; (ii) l'impact de la colonisation russe concernant les modifications démographiques et les mouvements de populations après leur arrivée ; (iii) les différents épisodes de réduction de la population en raison des famines, des épidémies ou des conflits obscurcissent considérablement l'interprétation des données génétiques. De plus, la structure et la répartition particulière des populations sibériennes en petits groupes isolés et fortement dispersés, les rendent particulièrement sensibles aux effets de la dérive génétique (Santos *et al.*, 1999) compliquant encore l'étude des phénomènes ayant eu lieu par le passé.

L'accès aux données génétiques des populations du passé constitue par conséquent une opportunité unique d'étudier la formation du peuple sakha et son évolution. Les premiers travaux réalisés sur une douzaine de sujets inhumés entre le XV^{ième} et le XVIII^{ième} siècles avaient permis de mieux préciser les pratiques funéraires des Yakoutes et également dégager certaines hypothèses concernant l'origine du peuplement (Ricaut, 2003).

Le présent travail de thèse s'inscrit dans la continuité des études de la MAFSO menées depuis 2000. Les expéditions de la MAFSO ayant permis la fouille de nombreuses tombes durant ces trois dernières années, nous étions en mesure de proposer une étude portant sur un nombre important de sujets inhumés en Yakoutie Centrale et dont la répartition chronologique s'étend du Moyen-Âge à la période postérieure à la colonisation russe.

Cet échantillon unique nous a permis, dans un premier temps, d'approfondir la connaissance des pratiques funéraires chez les Sakhas, tant sur le plan des modes de recrutement que sur les schémas d'organisation. Dans un second temps, nous avons pu envisager l'évolution de la population Yakoute *in situ* par la comparaison des données obtenues sur des sujets anciens à celles disponibles pour la population yakoute actuelle mais également pour les populations environnantes et en particulier les Evenks. La comparaison de nos résultats avec les données génétiques anciennes disponibles sur les régions du sud de la Sibérie telles que : le nord de la Mongolie (Keyser-Tracqui *et al.*, 2003), la région du lac Baïkal (Mooder *et al.*, 2006) ; l'Altaï (Ricaut, 2003) et l'ouest de la Sibérie (Keyser, données non publiées) nous permettra de confronter les hypothèses archéologiques et d'essayer d'éclaircir les modalités de l'ethnogenèse des Sakhas. Enfin, ces données génétiques devraient également permettre de répondre à plusieurs des interrogations survenues à la suite des découvertes archéologiques de la MAFSO telles que : l'existence de contacts avec les populations du sud de la Sibérie durant la période Xiongnu ; l'incidence au niveau génétique des changements d'influences culturelles ; l'importance des métissages avec les populations autochtones puis ultérieurement avec les colons russes ; l'influence de l'émergence d'une aristocratie locale au sein de la population yakoute au XVII^{ième} siècle ; etc...

I.5 Intérêts méthodologiques et moyens mis en oeuvre

Le climat extrêmement froid et sec de la République Sakha ainsi que la présence d'un permafrost continu sur l'ensemble de la Yakoutie Centrale combiné à des pratiques funéraires particulières engendrent des conditions taphonomiques propices à la conservation des restes humains et du matériel archéologique. Lors des fouilles réalisées ces trois dernières années, de nombreuses tombes ont livré des sujets dans un état de préservation exceptionnel. Ces conditions sont également particulièrement favorables à la conservation de l'ADN ce qui permet notamment l'analyse de l'ensemble des marqueurs moléculaires étudiés couramment sur les populations modernes (*Short Tandem Repeats* (STR) autosomaux, STR du chromosome Y et régions hypervariables de l'ADN mitochondrial) de façon quasi-systématique (Keyser-Tracqui *et al.*, 2003). La collaboration avec les anthropologues et archéologues yakoutes nous permet également de procéder à des prélèvements dans de bonnes conditions eut égard aux mesures contre les contaminations mais également de manière plus pratique, à la quantité de matériel disponible. Ainsi, pour la plupart des échantillons, nous avons pu disposer de suffisamment de matériel afin de mener au mieux les analyses.

L'étude des différents marqueurs moléculaires, dont les propriétés et l'utilisation seront détaillées plus amplement dans un chapitre ultérieur, nous permettent d'aborder différentes problématiques.

Tout d'abord les STR autosomaux sont utilisés, en plus de l'évaluation des éventuelles contaminations par des membres de l'équipe, pour la détermination moléculaire du sexe des individus. Cette détermination du sexe permet de confronter les diagnoses pratiquées au niveau morphologique et de proposer un résultat pour les individus immatures. Il est en outre possible d'aborder les relations de parenté pouvant exister au sein des ensembles funéraires grâce à ces marqueurs. L'analyse des STR rend donc possible une meilleure compréhension des modes de recrutement au sein des lieux d'inhumation et par conséquent des pratiques funéraires de la population yakoute suivant les époques ou les régions étudiées.

L'étude des marqueurs à transmission uniparentale tels que la région hypervariable de l'ADN mitochondrial ou les STR et *Single Nucleotide Polymorphisms* (SNP) du chromosome Y peuvent compléter les informations apportées par les marqueurs autosomaux dans la recherche des liens de parenté entre les sujets d'un même site. Mais leur intérêt principal réside dans la possibilité de suivre les lignées maternelles et paternelles des individus étudiés

et par la comparaison avec les données portant sur d'autres populations de proposer des hypothèses sur les schémas de migrations. De plus, les taux de mutations variables entre ces marqueurs permettent d'appréhender les mouvements de populations à des échelles de temps différentes, les SNP du chromosome Y signant des événements plus anciens alors que les STR renseignent plutôt sur des évolutions récentes.

En tenant compte de la répartition spatiale et temporelle de nos échantillons et grâce à l'analyse conjointe des différents marqueurs moléculaires avec les données génétiques modernes ainsi qu'avec les éléments archéologiques et anthropologiques il sera possible d'apporter des informations (i) sur les changements survenus au cours du temps au sein de la population sakha ; (ii) sur les contacts avec les populations environnantes et les éventuels évènements de dérive génétique ou les effets fondateurs ayant pu survenir

La comparaison de nos données génétiques anciennes avec des données modernes est par conséquent d'autant plus pertinente que ces données sont de très bonne qualité. En outre, de nouvelles perspectives méthodologiques comme l'extraction d'ADN à partir de différents substrats ou le développement de nouvelles méthodologies sont envisageables grâce à la préservation des corps et la bonne conservation de l'ADN.

II L'ADN ANCIEN

II.1 Les champs d'étude de l'ADN ancien

II.1.1 Historique de la discipline

II.1.1.1 Les premières études

La découverte par l'équipe de Wang *et al.* en 1981 de la conservation de la molécule d'ADN dans les tissus hépatiques d'une momie chinoise âgée de plus de 2000 ans permit l'émergence d'une nouvelle discipline : l'anthropologie moléculaire. Il fallu cependant attendre 1984 et les travaux de Higuchi *et al.* pour que le pas soit franchi entre la simple extraction d'acide nucléique et l'amplification par clonage suivie du séquençage d'un fragment d'ADN ancien. L'analyse d'une séquence de 112pb d'une espèce éteinte, le couagga (*Equus quagga*), à partir de muscle d'un sujet naturalisé au XIX^{ème} siècle démontra que les informations contenues dans l'ADN présent dans les tissus anciens pouvaient être étudiées avec succès. L'application à des restes humains ne tarda pas puisqu'en 1985, Svante Pääbo arriva pour la première fois à cloner un fragment d'ADN nucléaire extrait à partir de tissus mous d'une momie d'enfant égyptien âgé de 2400 ans et contenant plusieurs répétitions de séquences *Alu* (Pääbo, 1985). En outre, cette étude mit en évidence les difficultés inhérentes à l'étude de l'ADN ancien puisque parmi les 23 sujets analysés, seul un contenait de l'ADN étudiable. Ces difficultés liées aux dégradations chimiques et physiques de la molécule d'ADN feront l'objet d'une description dans un chapitre ultérieur (II.3.1 Les dégradations).

La découverte de la technique de PCR (*Polymerase Chain Reaction*) (Saiki *et al.*, 1985) représenta une alternative de choix au clonage bactérien puisque cette méthode allait permettre d'amplifier un fragment d'acide nucléique cible à partir d'un faible nombre, voire d'une seule copie d'ADN. La PCR eut pour conséquence une augmentation du nombre des études sur l'ADN ancien (Pääbo & Wilson 1988 ; Pääbo 1989; Pääbo *et al.* 1989; Thomas *et al.* 1989). Ces études sur les tissus mous, relativement rares lors des découvertes archéologiques, furent complétées par la suite par l'analyse de substrats osseux et dentaires. Les analyses effectuées par les équipes d'Hagelberg (Hänni, 1989), Hummel et Herrman (1991) et Tuross (1994) montrèrent que non seulement ces substrats durs pouvaient être utilisés comme source d'ADN mais que la conservation de la molécule d'ADN était meilleure que dans les tissus mous en raison de l'adsorption problable de l'ADN sur les cristaux d'hydroxyapatite (cf chapitre II.2 Les sources potentielles d'ADN ancien).

La PCR permit de confirmer les difficultés liées à la dégradation de la molécule d'ADN mais également de mettre en évidence l'un des problèmes majeurs de la discipline : la contamination par de l'ADN exogène. En effet, l'amplification par PCR est particulièrement sensible aux contaminations puisque les molécules d'ADN intactes sont préférentiellement amplifiées au détriment des molécules d'ADN endogènes souvent altérées.

Ces premières découvertes de la paléogénétique ouvrirent de nouvelles perspectives dans différents domaines comme l'anthropologie, l'archéologie et la génétique des populations. La comparaison de caractères morphologiques entre des sujets actuels et les fossiles allait ainsi pouvoir être transposée au niveau génétique. Jusqu'alors les études génétiques portant sur l'évolution des espèces ou des populations n'étaient basées que sur des données contemporaines qui n'apportaient que des preuves indirectes des changements survenus dans le passé. L'ADN ancien, par un accès direct aux données moléculaires des populations du passé, allait permettre de visualiser « en temps réel » les modifications et ainsi de mieux comprendre les liens entre les individus d'un même ensemble funéraire et, à une échelle plus vaste, les relations entre populations.

II.1.1.2 Les premières désillusions

Les perspectives offertes par cette nouvelle discipline ainsi que l'absence de lignes directrices entraînèrent, dans les années qui suivirent les premières découvertes, la publication de nombreux articles dont les résultats sont aujourd'hui contestés. Les annonces spectaculaires portant sur de l'ADN datant de millions d'années extrait à partir de plantes (Golenberg *et al.* 1990 ; Soltis *et al.* 1992 ; Poinar *et al.* 1993), d'os de dinosaures (Woodward *et al.* 1994) ou encore d'insectes piégés dans de l'ambre (Cano *et al.* 1992a, b, 1993 ; DeSalle *et al.* 1992, 1993, 1994) se multiplièrent au début des années 1990. La plupart de ces études portant sur de l'ADN datant de centaines de milliers voire de millions d'années se sont révélées être liées à des contaminations d'origine bactérienne ou humaine (Zischler *et al.* 1995a ; Gutierrez et Marin 1998). De plus, dans la majorité des cas, les analyses effectuées à l'époque n'ont pas pu être répétées rendant impossible toute vérification des résultats. Par la suite, il fut démontré qu'à l'exception des restes retrouvés dans le permafrost, la molécule d'ADN ne pouvait se conserver plus de 130 000 ans (Wayne *et al.*, 1999) réfutant ainsi l'authenticité des données obtenues à partir d'ADN antédiluvien.

En dépit des premiers essais aujourd’hui controversés, l’étude de l’ADN ancien est aujourd’hui reconnue comme une discipline scientifique à part entière. Cette évolution a été rendue possible par une meilleure connaissance des altérations subies par la molécule d’ADN, des mesures à mettre en place pour éviter les contaminations mais également grâce aux nombreux progrès méthodologiques.

II.1.2 La réaction d’amplification en chaîne : *Polymérase Chain Reaction* (PCR)

La découverte de la technique d’amplification de l’ADN par PCR (Saïki *et al.*, 1985) et l’optimisation de la méthode (Mullis et Falloona 1987, Saïki *et al.*, 1988) s’est révélée d’une importance capitale pour l’étude de l’ADN ancien. Les premières études avaient, en effet, démontré les difficultés rencontrées lors des tentatives de clonage bactérien de molécules d’ADN altérées et en faible nombre de copies. Les modifications *postmortem* subies par l’ADN avaient pour conséquence soit une absence de réPLICATION dans les bactéries, soit l’apparition d’erreurs lors de la réPLICATION liées aux processus de réparation de l’ADN (Pääbo *et al.*, 1989b). Par exemple, les deux remplacements de bases azotées détectés dans la séquence d’ADN mitochondrial du couagga lors de l’analyse complémentaire du sujet auraient été générés lors du processus de clonage. La PCR apparaît comme la méthode idéale pour l’étude de l’ADN ancien puisque cette technique permet de générer un nombre quasi illimité de copies d’un fragment d’ADN cible à partir d’un faible nombre de copies intactes. Par ailleurs, cette méthode est beaucoup moins sensible aux modifications *postmortem* que le clonage.

Le principe de la PCR (cf fig. 8) est basé sur la répétition de cycles thermiques comprenant des étapes de dénaturation, d’hybridation et d’elongation. L’amplification d’une région donnée est réalisée grâce à l’emploi d’amorces spécifiques qui s’hybrident de part et d’autre de la séquence cible. Si l’efficacité de la réaction est de 100%, le nombre de copies obtenues à partir d’une molécule d’ADN au bout de n cycles sera de 2^n . L’augmentation du nombre de cycles permettrait, en théorie, l’amplification d’une seule copie d’ADN cible (Gill *et al.*, 2000). Cependant, la très grande sensibilité de cette méthode entraîne une amplification préférentielle de toute molécule intacte présente. Cette propriété aura pour conséquence l’amplification de molécules d’ADN exogènes présentes dans le milieu réactionnel au

détriment de l'ADN du sujet étudié. Une très grande vigilance lors de toutes les étapes pré-PCR est donc nécessaire afin d'éviter toute amplification de molécules d'ADN contaminant. Ce risque important et les conséquences pour l'interprétation des résultats conduisirent en 2000 à la publication d'une liste de critères d'authentification des résultats par Cooper et Poinar qui seront exposés et discutés dans le chapitre II.4.

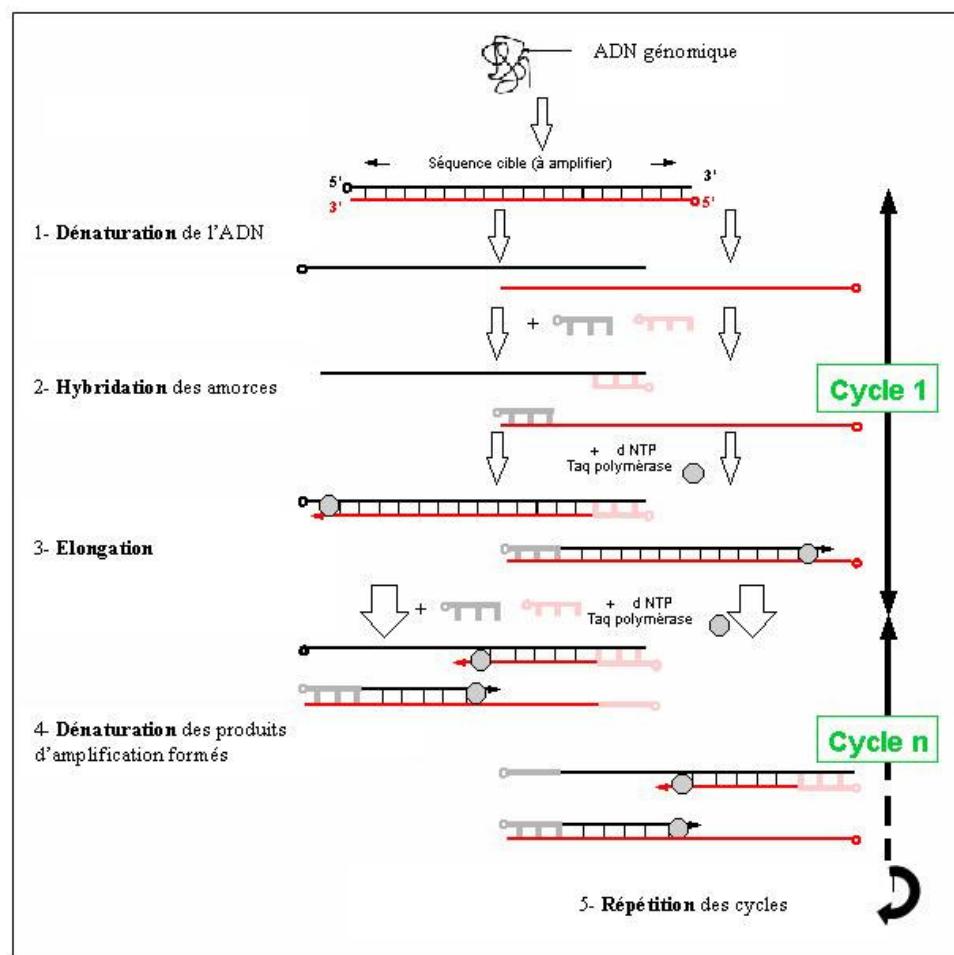


Figure 8 : Schéma des différentes étapes de l'amplification par PCR.

II.1.3 Les champs d'application

Les possibilités offertes par l'analyse du génome ancien intéressent des domaines d'étude très divers allant de la détermination du sexe d'individus isolés à la reconstruction phylogénétique des espèces éteintes ou la paléopathologie. Nous nous limiterons, cependant, dans ce chapitre à la présentation des applications relatives à la lignée humaine.

II.1.3.1 Etudes de restes humains

II.1.3.1.1 Détermination du sexe des individus

La détermination du sexe est une information importante dans l'étude des populations du passé. En effet, il est nécessaire de prendre en compte le sexe des sujets lors de l'étude des pratiques funéraires (infanticide sélectif, mode de recrutement, organisation des ensembles funéraires, ...) mais aussi des informations sur la biologie des populations anciennes telles que la stature ou les pathologies. En outre, la correspondance entre la détermination du sexe morphologique et génétique est considérée comme un critère d'authentification des résultats (Meyer *et al.*, 2000).

En anthropologie classique, la diagnose sexuelle des sujets matures est principalement basée sur le dimorphisme existant au niveau des os coxaux. La méthode de Bruzek (Bruzek, 2002) permet, par l'évaluation d'un ensemble de caractères morphométriques, de différencier de manière fiable (environ 95% de réussite) les deux sexes. D'autres méthodes basées sur des critères de robustesse du crâne (Ferembach, 1979) ou des os longs ont également été proposées, cependant ces méthodes sont peu fiables en raison de la variabilité importante de ces critères. La méthode de Bruzek n'est malheureusement pas toujours applicable. En effet, une mauvaise conservation des os coxaux, la découverte de squelettes incomplets ou de fragments osseux sont des facteurs limitants. Enfin, pour les sujets immatures, il n'existe pour l'instant aucune méthode ostéologique fiable, la biologie moléculaire représente donc la seule possibilité pour la diagnose sexuelle.

L'étude du gène de l'amélogénine est la méthode la plus fréquemment employée. Ce gène, impliqué dans la synthèse de l'émail dentaire et localisé sur les chromosomes X et Y, présente

des délétions qui permettent de différencier ces deux chromosomes (Nakahori *et al.*, 1991a et 1991b). La région analysée correspond au premier intron du gène qui présente une délétion de 6pb sur le chromosome X (Sullivan *et al.*, 1993). La taille des produits d'amplification sera donc de 106pb pour le chromosome X et de 112pb pour le chromosome Y. L'application de cette technique est quasi systématique dans les études portant sur l'ADN ancien (Faerman *et al.*, 1995 ; Stone *et al.*, 1996 ; Ovchinikov *et al.*, 1998 ; Cipollaro *et al.*, 1998 ; etc.). L'analyse moléculaire permet d'apporter des informations pour compléter les données de l'anthropologie classique comme dans le cas d'Ashkelon en Israël (Faerman *et al.*, 1998). Dans cette étude la détermination génétique du sexe de 43 sujets immatures répartis sur une période allant du IV^{ième} au VI^{ième} siècle AD permit de mettre en évidence un infanticide préférentiel des sujets masculins. L'étude du gène de l'amélogénine a également permis de réviser les conclusions apportées par l'analyse morphométrique de sujets problématiques comme dans le cas du cimetière de Aegerten en Suisse (Lassen *et al.*, 2000). L'analyse de 121 sujets immatures de ce site mit en évidence une sous estimation du nombre de sujets masculins lors de la détermination morphologique du sexe. L'utilisation de l'amélogénine ainsi qu'une nouvelle analyse morphométrique permit de réfuter l'hypothèse d'une pratique d'infanticide des sujets féminins proposée lors de la première étude du site. L'analyse d'autres régions spécifiques des hétérosomes (notamment les STR) permet, en plus de livrer des informations concernant les lignées maternelles et paternelles, de déterminer le sexe d'un sujet (Schultes *et al.*, 2000). Cette méthode est souvent utilisée comme contrôle, en parallèle de l'étude de l'amélogénine.

II.1.3.1.2 Etude des relations de parenté

Les relations de parenté au sein des ensembles funéraires sont classiquement abordées par l'étude des caractères discrets (Crubézy *et al.*, 2002). Le déterminisme de ces caractères phénotypiques, codés comme présent ou absent, est probablement génétique. Cependant leur mode de transmission exact est encore assez mal connu et la possible incidence de plusieurs facteurs sur leur apparition peut être un élément de biais. De plus, leurs fréquences n'étant pas similaires dans toutes les populations l'analyse des caractères discrets n'est pas toujours possible.

Pour l'étude des relations entre les individus, l'analyse de marqueurs autosomaux à transmission mendéienne ou de marqueurs haplotypiques à transmission uniparentale comme

ceux présents sur la partie non recombinante du chromosome Y et les régions hyper-variables (HV) de l'ADN mitochondrial est généralement employée. Ces marqueurs moléculaires permettent de retracer finement les relations de parenté ainsi que les lignées paternelles ou maternelles mais aussi d'appréhender les modes de recrutement des ensembles funéraires, l'organisation sociale et les schémas maritaux d'un groupe d'individus.

L'analyse des STR autosomaux, couramment utilisés pour l'identification génétique en médecine légale, permet d'étudier les relations de proche parenté (parents/enfants, grands-parents/petits-enfants, oncle/neveu, frère/frère, etc.). Pour établir les liens entre parents plus éloignés l'utilisation de ces seuls marqueurs ne suffit plus, aussi est-il nécessaire d'employer des marqueurs à transmission uniparentale. L'exemple le plus connu, et peut-être le plus controversé, est celui de la famille Romanov. L'équipe de Gill *et al.* réalisa en 1994 l'analyse génétique de 9 sujets retrouvés à Ekaterinburg en Russie et correspondant potentiellement à la famille du Tsar Nicolas II. Les relations de parenté entre les différents sujets furent confirmées par l'analyse des STR autosomaux et l'haplotype mitochondrial de ces sujets correspondait avec celui d'un des descendants de la famille.

L'étude conjointe des trois marqueurs est particulièrement utile pour l'analyse du mode de recrutement des grands ensembles funéraires. Ainsi l'étude de la nécropole d'Egyin Gol, qui couvre une période allant du III^e siècle avant JC au IInd siècle AD reflète bien les possibilités d'étude offertes par l'analyse de l'ADN ancien (Keyser *et al.*, 2003). L'analyse de 62 sujets grâce à des marqueurs moléculaires possédant des taux de mutations différents, STR du chromosome Y et région HV1 de l'ADN mitochondrial permit de compléter les connaissances existantes sur les pratiques funéraires des Xiongnu. En effet, les auteurs démontrent que l'organisation de la nécropole était basée sur les relations de proches parentés, l'appartenance à une même famille étendue ou au statut social.

II.1.3.1.3 Etude des mouvements de populations

Les mouvements des populations anciennes ainsi que la continuité des peuplements sont des questions abordées de longue date par les archéologues, les anthropologues et les généticiens. Les particularités génétiques étant transmises au fil des générations, les groupes d'individus actuels sont supposés présenter des fréquences similaires à celles de leurs ancêtres pour un marqueur génétique donné. Par inférence, il serait possible de retracer les événements passés par l'étude des données génétiques actuelles. Les études pluridisciplinaires combinant les

données archéologiques, génétiques et linguistiques ont permis de proposer des hypothèses sur la plupart des grandes migrations humaines (Ammerman et Cavalli-Sforza, 1984 ; Underhill *et al.*, 2001 ; Forster, 2004 ; etc.). Cependant, les phénomènes de dérive, les réductions de populations ou les migrations vont introduire des biais dans ces analyses. Il semble par conséquent difficile de baser ces hypothèses sur les seules données actuelles (Kolman *et al.*, 2000 ; Relethford 2001).

L'étude de l'ADN ancien semble être un outil particulièrement bien adapté pour répondre à ces questions puisqu'il permet d'accéder directement aux données génétiques anciennes et de les comparer avec les populations actuelles. L'effectif réduit des sujets anciens représente toutefois une limitation importante à ce type d'étude et il est indispensable de disposer d'un nombre de sujets conséquent (e.g. une trentaine de sujets) afin de s'affranchir, ou du moins de réduire, les biais d'échantillonnage.

L'origine des Européens est une question débattue de longue date en anthropologie aussi bien qu'en génétique des populations. L'étude de 57 sujets néolithiques (7000/7500 BP) appartenant à 16 sites répartis en Allemagne, Autriche et Hongrie par Haak *et al.* (2005) a permis d'évaluer la contribution des fermiers néolithiques au peuplement actuel de l'Europe. La première région hypervariable de l'ADN mitochondrial a été analysée grâce à l'utilisation de 4 fragments d'une centaine de paire de bases se chevauchant. Des résultats reproductibles ont été obtenus pour 24 sujets et, six d'entre eux se sont avérés appartenir à l'haplogroupe rare N1a. La fréquence élevée de cet haplogroupe, présent chez des sujets répartis sur l'ensemble des sites étudiés, suggère que les fermiers néolithiques n'auraient eu qu'un impact limité sur les lignées maternelles actuelles de l'Europe. Les chasseurs cueilleurs auraient adopté la culture des nouveaux arrivants dont les lignées maternelles se seraient "diluées" au cours du temps (Haak *et al.*, 2005). Il serait cependant nécessaire de déterminer la contribution des fermiers néolithiques par l'étude des marqueurs du chromosome Y afin d'évaluer l'influence de ces populations sur les lignées paternelles retrouvées aujourd'hui. De plus, certains aspect de cette étude ont été remis en cause (Ammerman *et al.*, 2006). En effet, la possibilité que les lignées maternelles des sujets analysés proviennent de femmes appartenant aux chasseurs-cueilleurs . Un écart chronologique important existe entre les échantillons de cette étude ce qui pourrait directement remettre en cause les conclusions de cette étude puisque les lignées maternelles auraient perduré en Europe Centrale durant 1700 ans (Ammerman *et al.*, 2006).

II.1.3.1.4 La lignée néandertalienne

Du fait de leur contemporanéité, la relation entre les Néandertaliens et les *Homo sapiens sapiens* est un sujet débattu depuis la découverte en 1856 de l'Homme de Néanderthal, dans la grotte de Feldhofer, près de Dusseldorf en Allemagne. A la suite de cette découverte, de nombreux restes ont été mis au jour du Moyen-Orient à l'Europe. De ce fait, les hommes de Néanderthal représentent actuellement le groupe d'Hommes fossiles le mieux connu puisque plus de 80 sites européens ont livré des restes de Néandertaliens et de pré-Néandertaliens ce qui représente environ 200 individus (Crubézy *et al.*, 2002).

Le débat opposant les partisans de la théorie multi-régionaliste à ceux de l'hypothèse "*Out of Africa*", partage les paléoanthropologues notamment depuis la découverte de fossiles dits intermédiaires comme l'enfant de Largar Velho. L'opportunité d'étudier directement le génome de ces hommes fossiles et de le comparer à celui des Hommes modernes a bien entendu été perçue comme une chance unique d'apporter une réponse à la question du métissage entre *Homo sapiens sapiens* et *Homo neanderthalensis*.

Depuis les premières études en 1997 (Krings *et al.*, 1997) pratiquées sur le sujet de la grotte de Feldhofer jusqu'à l'analyse des sujets d'El Sidron (Lazuela-Fox *et al.*, 2005) et de la grotte de Scladina (Orlando *et al.*, 2006), l'ADN de 11 Néandertaliens provenant de 5 sites différents a été analysé (Feldhofer ; Mezmaiskaya (Ovchinnikov *et al.*, 2000) ; Vindija (Krings *et al.*, 2000 et Serre *et al.*, 2004) ; Engis 2 (Serre *et al.*, 2004) ; la Chapelle-aux-Saints (Serre *et al.*, 2004) et Les Rochers de Villeneuve (Beauval *et al.*, 2005)). Les sujets étudiés couvrent une période allant de 29 000 à 100 000 ans et sont répartis sur une vaste zone géographique allant du Moyen-Orient à l'Europe Occidentale (cf Figure 9).

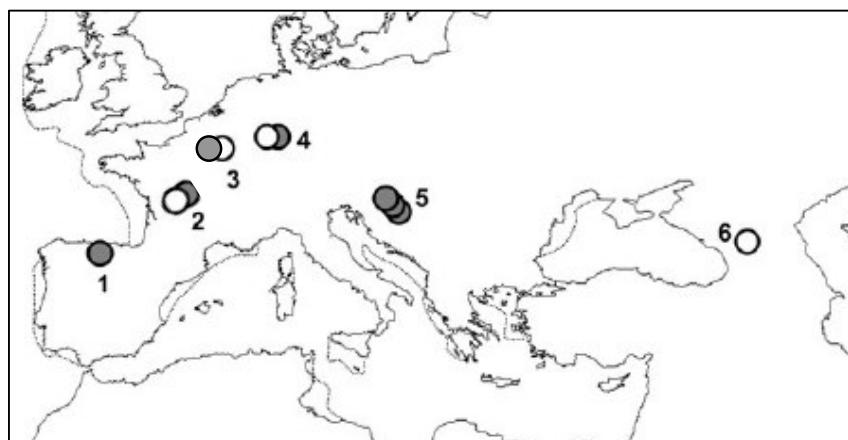


Figure 9 : Répartition géographique des sites ayant livré des restes de Néandertaliens analysés au niveau moléculaire. 1 : El Sidron ; 2 : La Chapelle-aux-Saints et Les Rochers-de-Villeneuve ; 3 : Engis et Scladina ; 4 : Feldhofer ; 5 : Vindija ; 6 : Mezmaiskaya. (Modifié à partir de Lalueza-Fox *et al.*, 2005).

Les études portant sur les sujets les plus récents sont assez homogènes quant aux résultats obtenus (Serre *et al.*, 2004). En effet, les séquences d'ADN mitochondrial provenant des différents sujets présentent des mutations qui sont retrouvées chez la plupart des individus analysés. La séquence HV1 de 123pb clonée par l'équipe d'Orlando *et al.* (2006) semble toutefois s'écartez des séquences précédentes. Cette courte séquence présente une position polymorphe commune à l'ensemble des sujets étudiés mais possède 3 nouvelles mutations qui n'avaient pas encore été décrites. Les auteurs suggèrent que la population des Hommes de Néanderthal aurait subi un phénomène de dérive ou de goulot d'étranglement conduisant à une réduction de la variabilité génétique entre 100 000 et 42 000 ans. Quelqu'en soit la cause, cohabitation avec *Homo sapiens sapiens*, changements climatiques ou division de la population, la diversité génétique aurait été sous estimée dans les études antérieures (Orlando *et al.*, 2006).

Cependant, les reconstitutions phylogénétiques sont assez homogènes et placent les Néandertaliens dans un cluster séparé de celui des Hommes anatomiquement modernes. De plus, la comparaison de ces séquences avec des données obtenues sur des *Homo sapiens sapiens* du Paléolithique Supérieur (Caramelli *et al.*, 2003 ; Serre *et al.*, 2004) ou sur des séquences actuelles tendent à montrer que la participation des hommes de Néanderthal au pool génétique des Européens actuels aurait été faible voire inexisteante (Knight *et al.*, 2003 ;

Serre *et al.*, 2004). Il est important de nuancer ces résultats en considérant le nombre de sujets étudiés, à la fois pour les Néandertaliens et pour les *Homo sapiens sapiens* contemporains. La publication de deux articles portant sur le séquençage de 65000pb (Noonan *et al.*, 2006) et 1 million de paires de bases(Green *et al.*, 2006) de l'ADN nucléaire d'un sujet de la grotte Vindija représente une avancée exceptionnelle dans la compréhension des relations entre ces Hommes fossiles et nos ancêtres. Outre les possibilités d'analyse des dégradations de l'ADN (Stiller *et al.*, 2006), la technique de pyroséquençage (Margulies *et al.*, 2005) utilisée dans ces deux études semble ouvrir la voie à de nouvelles études et à une meilleure compréhension des relations entre ces deux lignées d'Hominidés. L'analyse des résultats obtenus n'en est qu'aux premières étapes d'interprétation, cependant, ces deux équipes proposent des dates de divergences entre les Néandertaliens et les Hommes anatomiquement modernes, antérieures à l'émergence de ces derniers. De plus, ces premiers résultats semblent indiquer une absence de mélange génétique entre les Néandertaliens et les Hommes modernes en Europe, bien que Noonan et collaborateurs soulignent l'importance de l'étude d'autres sujets.

Soulignons cependant que lors des analyses, toute séquence présentant des similitudes avec des séquences humaines est écartée en raison du risque possible de contamination. Ce choix, bien que nécessaire afin d'éviter toute interprétation fallacieuse liée à la prise en compte de résultats erronés, peut néanmoins biaiser les conclusions en excluant d'emblée des séquences originales proches des séquences humaines (Nordborg, 1998 ; Trinkaus, 2001).

Les applications de l'ADN ancien ne se limitent pas à la lignée humaine. En effet, les analyses génétiques d'espèce disparues permettent de replacer ces dernières phylogénétiquement, l'étude des microorganismes et des pathogènes permet de compléter les diagnostics proposés à partir des restes osseux mais également d'étudier l'évolution des agents pathogènes, enfin l'impact de la domestication sur le génome d'espèces animales ou végétales peut être évalué par l'analyse des génomes anciens.

II.2 Les sources potentielles d'ADN ancien

L'ensemble des substances d'origine organique peut être considéré comme une source potentielle d'ADN ancien (O'Rourke *et al.*, 2000). Les substrats les plus couramment utilisés sont les os, les dents et les tissus mous. Il est également possible d'analyser des cheveux ou des coprolithes mais leur emploi est moins fréquent. Toutefois en raison de la conservation différentielle des tissus après le décès et des propriétés des différents tissus, certains substrats permettent de meilleurs résultats que d'autres.

II.2.1 Les os

Les os représentent le type de substrat le plus fréquemment employé dans les études de paléogénétique principalement en raison de leur meilleure conservation ce qui rend plus abondants que les autres substrats, mais également car il semble que l'ADN soit mieux conservé dans les tissus durs (os et dents) et que le risque de contamination soit plus faible (Hagelberg *et al.*, 1989 ; Cooper *et al.*, 1992).

II.2.1.1 Structure et propriétés

L'os est un matériau composite formé par l'assemblage d'une trame organique (90% de collagène et 10% de glycoprotéines et protéoglycans) sur laquelle se forme une phase minérale. Cette phase minérale est composée par des cristaux d'hydroxyapatite de très petite taille ce qui favorise les réactions avec le milieu (*e.g.* dissolution) si un équilibre n'est pas maintenu par des processus physiologiques. Toutefois, l'imbrication des deux phases confère à l'os une résistance face aux dégradations pouvant survenir après la mort (Trueman et Martill, 2002). Le collagène étant une molécule non soluble, les fibrilles fournissent une protection aux cristaux d'hydroxyapatite et ces derniers empêchent l'accès des collagénases secrétées par les microorganismes (Lees, 1989). L'os compact est organisé de façon hiérarchique en plusieurs niveaux (cf Figure 10). A l'échelle macroscopique, l'os cortical est formé par les ostéons qui représentent l'unité élémentaire. Chaque ostéon est composée de cercles concentriques constitués par les canaux de Havers. Les différents ostéons sont reliés entre eux par des lamelles interstitielles formées par les restes d'ostéons antérieurs donnant

ainsi une structure compacte. Le niveau inférieur est représenté par les fibres de collagène composées de fibrilles. Enfin au niveau moléculaire on distingue le collagène et les cristaux d'hydroxyapatite.

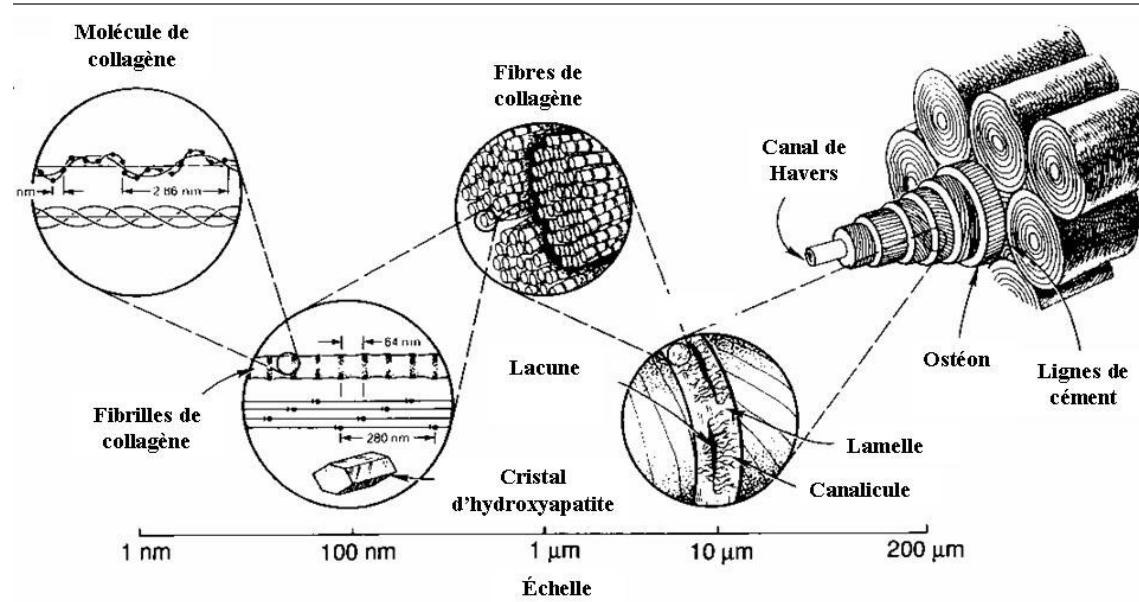


Figure 10 : Structure hiérarchique de l'os compact humain (modifié d'après Lakes, 1994).

II.2.1.2 Diagenèse des tissus osseux et préservation de l'ADN

La diagenèse des os est un phénomène complexe qui peut varier au sein d'un même site (Jans *et al.*, 2002). La dégradation des ossements est liée (i) à des facteurs intrinsèques comme l'âge biologique du sujet étudié (Gordon et Buikstra, 1981) et l'espèce (Nicholson, 1998) (ii) mais aussi aux facteurs environnementaux comme le pH et le potentiel d'oxydo-réduction du sol (Gordon et Buikstra, 1981 ; Stephan, 1997), la température, la présence d'eau, les dégradations dues aux microorganismes (Hedges, 2002) (iii) et enfin aux pratiques funéraires telles l'exposition du corps, le décharnement ou la profondeur d'ensevelissement (Hedges, 2002).

L'ensemble des phénomènes de dégradation de l'os va intervenir dans les chances de retrouver de l'ADN mais d'autres facteurs peuvent influer sur la qualité des acides nucléiques. L'adsorption de l'ADN sur les cristaux d'hydroxyapatite combinée à la conformation tridimensionnelle de la molécule dans l'os permettrait de ralentir les lésions subies par les

acides nucléiques (Collins *et al.*, 2000). Cette adsorption est certainement liée à une interaction entre les groupements phosphates chargés négativement et les ions calcium chargés positivement (Okazaki *et al.*, 2001) (Figure 11). Les interactions entre la surface des cristaux d'hydroxyapatite, leur conformation et les conditions favorisant la fixation de la molécule d'ADN restent toutefois à établir.

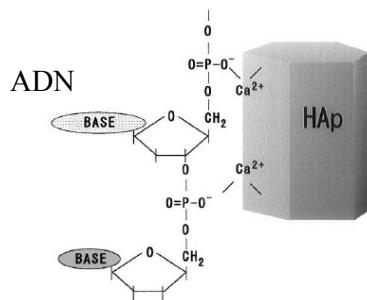


Figure 11 : Modèle schématique de fixation de l'ADN à la surface d'un cristal d'hydroxyapatite (HAp). (Modifié d'après Okazaki *et al.*, 2001).

L'équipe de Götherström *et al.* (2002) a mis en évidence que la présence d'ADN dans les os est inversement corrélée au degré de cristallisation de l'hydroxyapatite. De plus, la liaison entre les biomolécules (comme le collagène) et l'hydroxyapatite empêche les phénomènes de dissolution et recristallisation (Long *et al.*, 1998). Les conditions qui entraînent une dégradation de ces molécules (attaques par des microorganismes, température élevée, pH acide) vont donc favoriser la recristallisation de l'hydroxyapatite ce qui aura pour conséquence d'augmenter le risque dégradation de l'ADN (Götherstörn *et al.*, 2002).

Il semble que la présence dans le tissu osseux d'agrégats de cristaux d'hydroxyapatite offre une meilleure protection de l'ADN vis à vis des facteurs environnementaux que l'adsorption sur des cristaux isolés (Salamon *et al.*, 2005). L'analyse de l'ADN fixé sur ces agrégats après un traitement à l'hypochlorite de sodium permet, selon les auteurs, d'obtenir de l'ADN de meilleure qualité et de s'affranchir de la présence des contaminants ainsi que de l'action des inhibiteurs.

Les prélèvements sur les échantillons osseux sont préférentiellement réalisés à partir de l'os compact qui présente une densité minérale plus importante que l'os spongieux. Cette densité minérale est certainement en relation avec les propriétés de conservation de l'ADN. Nous avons d'ailleurs constaté que, pour les échantillons provenant de nouveaux-nés ou de jeunes enfants, pour lesquels le tissu osseux n'est pas encore organisé de façon lamellaire (tissu osseux dit tissé), la probabilité de retrouver de l'ADN était réduite (cf résultats).

II.2.2 Les dents

Les dents représentent un substrat intéressant du fait de leur grande longévité (Smith *et al.*, 1993) et de leur résistance face aux facteurs environnementaux tels que la pression, l'humidité (Sweet *et al.*, 1995) ou les températures très élevées (Nossintchouk *et al.*, 2002). Néanmoins ce substrat est moins souvent utilisé que les os car une seule dent ne permet pas de générer autant de matériel et par conséquent la répétition des analyses, nécessaire à la validation des résultats, est moins aisée (Lazuela-Fox *et al.*, 2003). En revanche, si plusieurs dents d'un même sujet sont disponibles, la comparaison des résultats peut être très informative puisque chaque dent peut être considérée comme un échantillon indépendant (O'Rourke *et al.*, 2000).

II.2.2.1 Structure

La dent est constituée par une chambre pulpaire dans laquelle se trouvent les vaisseaux sanguins ainsi que les nerfs. Cette cavité est entourée de dentine (ivoire) recouverte par de l'émail au niveau de la couronne et de cément dans la partie racinaire (cf figure 12).

L'émail est le tissu le plus minéralisé de l'organisme. Il est composé pour 96% de son poids tissulaire d'une phase minérale, pour 2% d'eau et 2% de matrice organique et d'"impuretés" (CO_3 , Na, Mg, Cl, K, F). La phase minérale est constituée de phosphates de calcium, en majorité organisés en cristaux d'apatite tout comme dans le tissu osseux (Triller, 1996). L'émail dentaire confère aux dents une protection vis à vis des contaminants extérieurs en raison de l'imperméabilité de ce tissu (Hillson, 2005).

La pulpe dentaire est un tissu conjonctif spécialisé, fibreux, riche en vaisseaux sanguins et lymphatiques et en fibres nerveuses. En périphérie se trouve une couche d'odontoblastes, cellules différencierées qui assurent la sécrétion des composants nécessaires à la minéralisation de la dentine. La pulpe dentaire, riche en cellules nucléées, représente potentiellement la principale source d'ADN nucléaire de l'organe dentaire.

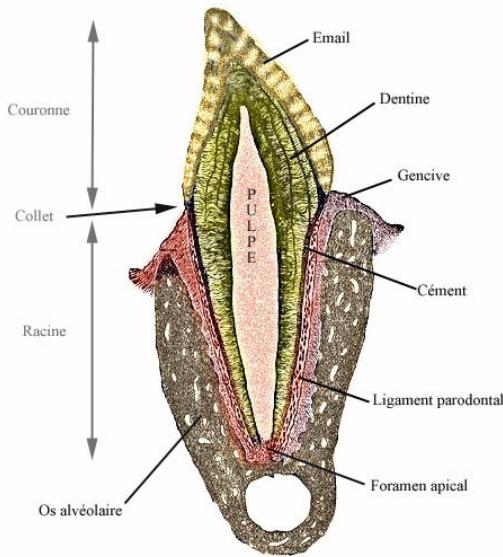


Figure 12: Schéma de coupe longitudinale d'une dent. (D'après Gray, 1918).

La dentine représente en volume le tissu dentaire le plus important. Ce tissu est fortement minéralisé mais sa composition diffère de celle de l'email, puisqu'elle est constituée d'une fraction minérale représentant 70% en poids, d'une fraction organique (20%) et enfin d'eau (10%). La dentine est traversée sur toute son épaisseur par des tubuli, petits canaux qui renferment les prolongements cellulaires des odontoblastes. L'ADN mitochondrial serait localisé au niveau de ces prolongements (Chomette *et al.*, 1992). En outre ce tissu riche en hydroxyapatite doit permettre, tout comme dans le tissu osseux, une adsorption et une protection de l'ADN au cours du temps.

II.2.2.2 Utilisation des dents en paléogénétique

Les dents représentent un substrat de choix pour les études de paléogénétique du fait de la bonne conservation de l'ADN endogène (Ricaut *et al.*, 2004 ; Amory *et al.*, 2006 ; Lazuela-Fox *et al.*, 2006) et de sa protection vis à vis des contaminants extérieurs. Cependant, les racines peuvent constituer une voie de contamination possible en raison de leur porosité et de la présence du canal racinaire. En outre les lésions *antemortem* (caries, fractures, usure extrême de la couronne) ou *postmortem* (cassures, ouvertures des canaux racinaires) représentent également des sources possibles de contaminations (Grimoud *et al.*, 2004). Il est important que les échantillons dentaires soient protégés immédiatement après leur extraction

et surtout qu'aucun lavage ne soit effectué hors d'un laboratoire permettant d'assurer les conditions de stérilité requises (Sampietro *et al.*, 2006).

Les méthodes de prélèvement les plus couramment utilisées sont : (i) le broyage intégral de la dent qui présente l'avantage de générer une quantité relativement importante de matériel destiné à l'extraction, (ii) la section, horizontale ou longitudinale, puis la récupération du matériel présent dans la cavité pulpaire et (iii) la section de la dent en dessous du collet puis le prélèvement des racines incluses dans les maxillaires ou la mandibule puis le broyage de cette partie. La deuxième méthode permet d'obtenir de l'ADN de meilleure qualité que la première mais en moins grande quantité (Smith *et al.*, 1993). En outre, la technique de broyage de la dent implique bien entendu la destruction complète de l'échantillon ce qui représente une perte d'information anthropologique. La trépanation de la couronne permet quant à elle de reconstituer la dent en limitant l'altération physique (Shiroma *et al.*, 2004).

II.2.3 Les cheveux

Ces dernières années, quelques études ont été menées sur des cheveux humains anciens (Baker, 2001 ; Gilbert *et al.*, 2004) ainsi que sur des poils d'animaux (Bonnichsen *et al.*, 2001 ; Gilbert *et al.*, 2004). Toutefois ce type d'échantillon est bien moins utilisé que les tissus durs pour deux raisons : (i) la quantité d'ADN présente dans les cheveux est relativement faible comparée aux échantillons osseux ou dentaire (Higuchi *et al.*, 1988 ; Allen *et al.*, 1998) ; (ii) de plus, les éléments pileux ne peuvent se conserver sur une longue durée que dans des conditions particulières (e.g. permafrost, momification) ce qui limite drastiquement le nombre de sujets étudiables. Pourtant l'étude de ce substrat peut se révéler intéressante du fait du caractère peu invasif des techniques d'échantillonnage mises en oeuvre (Gilbert *et al.*, 2006) et de certaines propriétés physico-chimiques intrinsèques en relation avec la structure et la physiologie des cheveux.

II.2.3.1 Structure du cheveu

Le cheveu peut être séparé en deux zones : le follicule et la tige. Le follicule est généralement la source d'ADN nucléaire utilisée, par exemple, pour l'identification des individus en médecine légale. Le bulbe contient en effet de nombreuses cellules nucléées qui forment les différentes gaines entourant la racine du cheveu, les mélanocytes présents à la base du cheveu ainsi que des cellules non différencierées (cf figure 13).

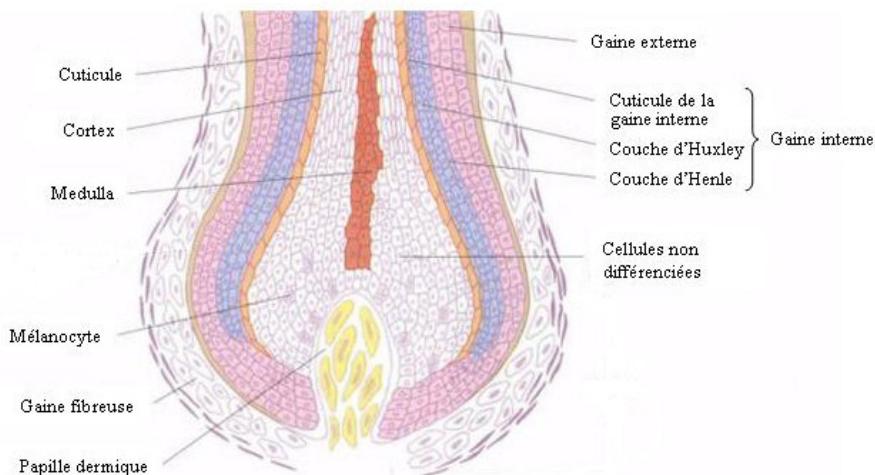


Figure 13 : Structure du bulbe capillaire (modifié à partir de Whiting et Howsden, 1996).

La partie supérieure du follicule représente une zone de transition entre la zone contenant les cellules en division responsables de la croissance du cheveu et la tige dont les cellules kératinisées sont majoritairement sénescentes.

La tige comprend trois couches cellulaires distinctes : la medulla, le cortex et la cuticule (Ebling, 1980, Harding et Rogers, 1999) (Figure 14). Les cellules du cortex commencent à synthétiser de la kératine rapidement après leur différenciation au sein du follicule (Linch *et al.*, 1993). Cette protéine formant 90% du cheveu dans des conditions d'humidité normale (Robbins, 1985) représente, sous forme de fibrilles, l'élément structural majoritaire du cheveu. La kératinisation va entraîner une cytolysé et une perte d'ADN nucléaire due à l'action de DNases au niveau des cellules corticales (Ackerman *et al.*, 1993 ; Linch *et al.* 1993) ; les organites intracellulaires et en particulier les mitochondries peuvent cependant résister à cette dégradation (Hacker *et al.*, 2000).

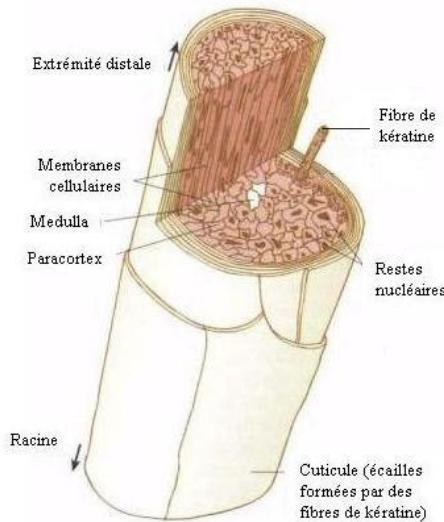


Figure 14 : Schéma d'une tige de cheveu (modifié à partir de Whiting et Howsden, 1996).

Les noyaux de ces cellules seront en partie piégés dans les filaments de kératine (Harding et Rogers, 1999) ce qui explique la nécessité d'une phase de dégradation protéique prolongée lors de l'extraction d'ADN (Marshall, 1985).

II.2.3.2 Propriétés structurales et l'analyse moléculaire des cheveux

Les caractéristiques structurales et histologiques du cheveu sont responsables de certaines particularités qui se retrouvent dans l'étude moléculaire de ce substrat.

La structure du cheveu ainsi que la nature hydrophobe des protéines des cellules de la cuticule (Gilbert *et al.* ; 2006b) et la présence de kératine (Fraser *et al.*, 1972 ; Valkovic, 1988) semble protéger l'ADN endogène vis à vis des contaminants extérieurs.

En outre, il semble qu'il existe une corrélation entre l'intégrité structurale et la qualité des résultats qu'il est possible d'obtenir (Gilbert *et al.*, 2006a). Cette étude a mis en évidence que les dégradations subies par l'ADN étaient d'autant plus importantes que la structure histologique des cheveux était endommagée. Une analyse microscopique est par conséquent recommandée avant d'entamer toute analyse afin de s'assurer de la qualité du matériel étudié.

Plusieurs études ont montré que lors de l'analyse de séquences correspondant à de l'ADN extrait à partir de cheveux, la présence simultanée de plusieurs populations de molécules d'ADN mitochondrial présentant des compositions nucléotidiques légèrement différentes était observée plus fréquemment que dans les autres types de substrats (Grzybowski *et al.*, 2000, 2003 ; Budowle *et al.*, 2002 ; Brandstätter *et al.*, 2003 ; Tuly *et al.*, 2004). L'importance de ce phénomène, appelé hétéroplasmie, est en relation avec l'origine histologique des cellules corticales de la tige. Deux populations cellulaires peuvent être à l'origine de ces cellules : les cellules germinales non différenciées et les mélanocytes (Linch *et al.*, 2003). De plus, il semble que la fréquence des hétéroplasmies soit variable en fonction de la zone de la tige considérée. Les hétéroplasmies seraient plus nombreuses à mesure que l'on s'approche de l'extrémité distale du cheveu. Ce phénomène serait lié à la production constante de mitochondries par les mélanocytes qui pourrait induire une dérive dans les proportions respectives de chaque population de mitochondries (Linch *et al.*, 2003).

II.2.4 Les autres substrats

D'autres types de substrats peuvent être analysés comme alternative ou complément à ceux précédemment décrits. Les tissus mous tels que le muscle (Higuchi *et al.*, 1984), la peau (Thomas *et al.*, 1990 ; Pääbo, 1985a) ou différents organes (Pääbo, 1985b) ont été utilisés dès les premières études de paléogénétique. Il semble que les échantillons de tissus mous les plus appropriés soient les tissus ayant subi une dessiccation rapide qui est supposée diminuer les altérations hydrolytiques (Pääbo, 1989) mais ne réduit pas les effets des autres types de dommages. Par conséquent il serait nécessaire de prélever les tissus superficiels ou périphériques. Or les risques liés aux contaminants pouvant être présents à la surface des échantillons impliquent que les prélèvements soient réalisés en profondeur. Les quantités d'ADN extrait à partir de tissus mous sont cependant plus faibles que celles obtenues à partir d'os. De plus, la co-extraction d'inhibiteurs étant plus importante, ce type d'échantillons est dorénavant moins utilisé au profit des substrats osseux ou dentaires.

Les coprolithes représentent un autre type de substrat pouvant être analysé dans le but d'étudier l'ADN du sujet mais également pour identifier le régime alimentaire de l'espèce (Höss *et al.*, 1992 ; Sobolik *et al.*, 1996 ; Poinar *et al.*, 1998, 2001 ; Hofreiter *et al.*, 2000 ; Loreille *et al.*, 2001 ; Rollo *et al.*, 2002).

L'ADN présent sur l'ensemble des produits de l'industrie humaine peut être analysé afin de mieux appréhender les pratiques culturelles des utilisateurs en identifiant les espèces animales domestiquées ainsi que les plantes cultivées ou utilisées dans la pharmacopée. Ainsi, les résidus organiques retrouvés sur des objets usuels tels les outils lithiques (Hardy *et al.*, 1997), la vaisselle (Burger *et al.*, 2000), les papyrus (Marota *et al.*, 1997) ou les parchemins (Bar-Gar *et al.*, 1995) ont été utilisés dans quelques études. Cependant ces analyses restent plutôt anecdotiques en comparaison avec les études menées sur les substrats "traditionnels" et les résultats de certaines de ces études ont été remis en cause. Une étude portant sur la vitesse de dégradation de l'ADN dans les papyri égyptiens a mis en évidence que les échantillons âgés de plus de 700 à 800 ans ne pouvaient contenir de l'ADN endogène (Marota *et al.*, 2002). A la suite de cette publication une controverse sur les résultats des différentes études portant sur les échantillons égyptiens est apparue conduisant à différentes publications contradictoires (Zink et Nerlich, 2003 ; Gilbert *et al.* 2005).

Quelle que soit la problématique envisagée ou le substrat utilisé les études de paléogénétiques vont invariablement être confrontées à des difficultés liées à la conservation des acides nucléiques au cours du temps. Ces difficultés sont en relation avec les dégradations subies au cours du temps ou à des substances chimiques qui peuvent inhiber les réactions d'amplification compliquant considérablement l'accès à l'information génétique.

II.3 Les spécificités des études sur l'ADN ancien

II.3.1 Les dégradations de la molécule d'ADN

L'instabilité post-mortem des acides nucléiques représente une des difficultés principales rencontrée lors de l'analyse de l'ADN ancien. Dans les cellules vivantes, l'intégrité de l'ADN est continuellement maintenue par des processus de réparation enzymatiques (Lindahl *et al.*, 1993). Après la mort de l'individu, les phénomènes spontanés d'hydrolyses et d'oxydation ne sont plus contrebalaancés. De plus, les enzymes cataboliques (nucléases et protéases lysosomiques) sont libérées dans la cellule entraînant une dégradation rapide des acides nucléiques par l'attaque de différents sites (cf figure 15). La molécule d'ADN est, en outre, soumise à un ensemble de dégradations causées par l'action des insectes, des champignons et des bactéries qui se succèdent après la mort. Dans certains cas, comme une dessiccation rapide des tissus (Pääbo, 1989), l'adsorption sur la matrice des cristaux d'hydroxyapatite formant la structure des os (Tuross, 1994), ou la conservation dans des environnements particuliers, les acides nucléiques peuvent échapper à ces dégradations. Cependant les processus de réparation intra-cellulaires ne sont plus actifs, par conséquent la perte d'intégrité des acides nucléiques au cours du temps est inéluctable même si elle peut être ralentie (Pääbo *et al.*, 2004).

II.3.1.1 Nature des dégradations

II.3.1.1.1 Fragmentation du brin d'ADN

La dégradation la plus flagrante subie par l'ADN est sa réduction en fragments variants de 100 à 500 pb (Pääbo, 1989 ; Handt *et al.*, 1994 ; Höss *et al.*, 1996). Cette fragmentation est due à la fois aux réactions enzymatiques qui surviennent peu après la mort ainsi qu'à l'hydrolyse qui entraîne la rupture des liaisons phosphodiester générant des coupures sur un des brins (cf Fig.16a). Les liaisons N-glycosidiques entre les bases azotées et le squelette désoxyribophosphate sont également soumises à des lésions hydrolytiques (cf Fig. 16b). Ces lésions non enzymatiques auront pour effet des pertes de bases azotées qui favorisent les

ruptures de brin. Lors des processus de décomposition, l'action des microorganismes agrave encore la fragmentation de la molécule d'ADN.

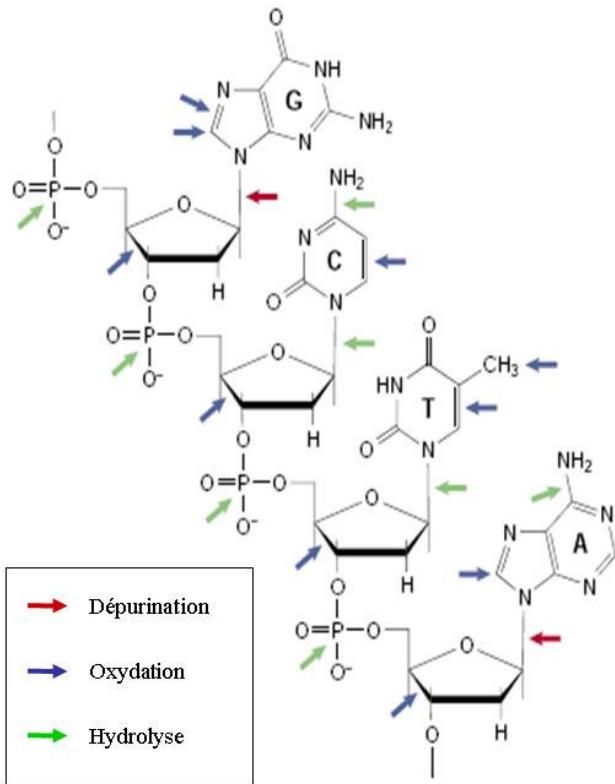


Figure 15 : Sites de lésions sur la molécule d'ADN par les différent type de dégradation (modifié à partir de Hofreiter *et al.*, 2001).

D'autre part, les peroxydes (composés à groupement sulfhydryl (cystéine)) et certains ions métalliques (Fe^{2+} et Cu^{2+}) provoquent des ruptures des liaisons phosphodiester. Les radiations ionisantes produisent aussi des cassures soit par l'action d'électrons secondaires produits par l'impact de particules b- ou de photons X, soit par l'ionisation de l'eau et la production de radicaux libres qui sont capables d'attaquer les liaisons phosphodiesters (Spotheim-Maurizot *et al.*, 2003).

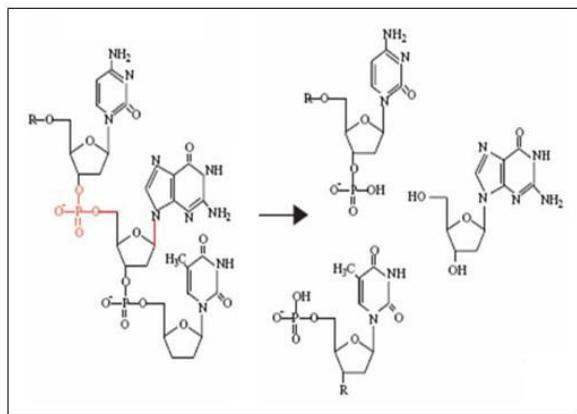


Figure 16a : Rupture de brin par hydrolyse de liaison phosphodiester (modifié à partir de Willerslev et Cooper, 2005)

Les ruptures de brins sont un des facteurs principaux qui limitent l'amplification et par conséquent la longueur des séquences étudiabiles. Cependant d'autres lésions interviennent également en bloquant l'elongation de la Taq polymérase.

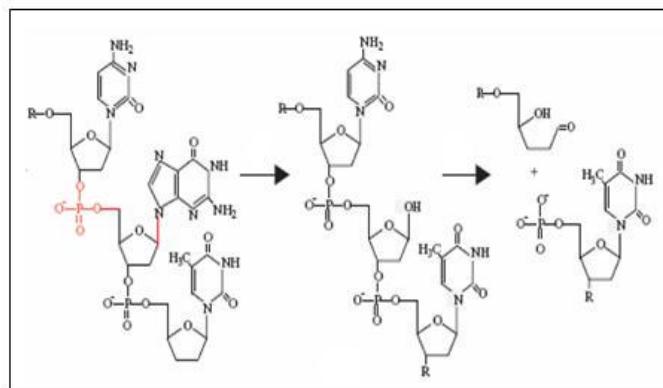


Figure 16b : Rupture de brin liée à la perte de base azotée (modifié à partir de Willerslev et Cooper, 2005)

II.3.1.1.2 Hydrolyse

Les lésions hydrolytiques entraînent la perte des groupements amines des bases azotées (Hofreiter *et al.*, 2001). Ainsi l'adénine est transformée en hypoxanthine, la cytosine en uracile (la plus fréquente) et la guanine en xanthine (Lindhal, 1993). Les bases modifiées n'ont pas les mêmes spécificités d'hybridation que les bases dont elles dérivent (Eglinton et Logan, 1991). La xanthine garde cependant la même spécificité de liaison à la cytosine mais

avec deux liaisons hydrogène seulement. Ces changements ont pour conséquence l'incorporation de bases incorrectes (A/G et C/T) lors de l'amplification par PCR ce qui peut conduire à des erreurs lors de l'analyse et de l'interprétation des résultats.

Ce type d'erreurs explique la nécessité de répéter plusieurs fois les analyses à partir d'extraits d'ADN différents afin de détecter d'éventuelles erreurs survenues durant la PCR.

II.3.1.1.3 Oxydation

Les lésions oxydatives entraînent des modifications des bases azotées. Les principales mutations sont responsables de l'apparition d'hydroxyguanines (Lindahl *et al.*, 1993) et d'hydantoïnes (Höss *et al.*, 1996). En outre, ces lésions vont agir sur les liaisons chimiques des résidus ribosiques ce qui aura pour conséquence une fragmentation du brin d'ADN. Les oxydations affectant la molécule d'ADN ont pour effet un blocage de la polymérase et donc l'impossibilité d'amplifier un fragment contenant des bases modifiées, notamment les hydantoïnes (Hoss *et al.*, 1996).

Dans certains cas, les lésions oxydatives favorisent la formation de séquences chimériques lors de l'amplification par le phénomène de "*jumping PCR*" (Pääbo *et al.*, 1989). Ces séquences mosaïques correspondent à la recombinaison de molécules dégradées, de contaminants et de fragments de la séquence cible générant un grand nombre de séquences différentes (Gilbert *et al.*, 2003).

Il est important de noter que l'ADN mitochondrial semble être plus sensible aux phénomènes d'oxydation attendu que les mitochondries sont le siège du métabolisme cellulaire. A l'inverse l'ADN nucléaire serait mieux protégé du stress oxydatif puisque le noyau cellulaire est un milieu anoxique. Le taux de mutation plus élevé de l'ADN mitochondrial pourrait être en relation avec cette différence de niveau d'oxydation (Wallace, 1989).

II.3.1.1.4 Liaisons croisées

La formation de liaisons croisées représente un autre type de dégradation pouvant induire un blocage de la Taq polymérase et donc une absence d'amplification. Ces modifications peuvent être de deux types : liée au phénomène d'alkylation ou à la formation de complexes de Maillard (Figure 17).

L'alkylation se caractérise par le transfert de groupements méthyles ou éthyles sur des bases ou des phosphates de l'ADN. Ainsi la guanine peut être transformée en O6-méthylguanine, ce qui entraîne la rupture de liaisons hydrogène entre la guanine et la cytosine. Par ailleurs la O6-méthylguanine peut s'apparier par erreur avec la thymine, ce qui pourrait conduire à une transition G-C→T-A. Les éléments alkylants sont présents de manière endogène, mais aussi dans le milieu : le chlorométhane produit par les plantes et les champignons est le plus abondant (Drablos *et al.*, 2004).

Les complexes de Maillard correspondent à des liaisons qui se forment entre les sucres et les groupements amines des acides nucléiques et des protéines. La formation de produits de Maillard a pour conséquence une séquestration des molécules d'ADN et donc une absence d'amplification. La présence de produits de Maillard dans les extraits d'ADN ancien a été mise en évidence par Poinar *et al.* en 1998. Les résultats de cette étude ont montré que l'utilisation de Bromure de N-Phényl Thiazolium (PTB) pouvait, dans certains cas, permettre l'amplification de séquences anciennes à partir d'échantillons problématiques comme des coprolithes datés de 20 000 ans (Poinar *et al.*, 1998) ou encore des ossements de Néandertaliens (Krings *et al.*, 2000).

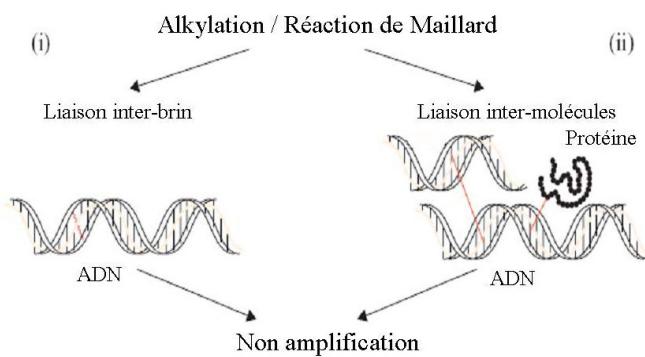


Figure 17 : Différents types de réactions chimiques conduisant à la formation de liaisons croisées.

Une étude récente (Hansen *et al.*, 2006) a montré que la formation de liaisons croisées, et non la fragmentation, était le principal facteur limitant de l'analyse des échantillons anciens conservés dans le permafrost. En effet, la fréquence de formation des liaisons entre les molécules d'ADN serait beaucoup plus importante que celle des ruptures de brin dans ce milieu.

II.3.1.2 Sources de dégradations et facteurs environnementaux

Les dégradations subies par l'ADN sont, dans un premier temps, liées aux enzymes endogènes ou à des phénomènes se produisant rapidement après la mort, lors de la décomposition. Cependant sur de longues périodes de temps les facteurs environnementaux deviennent les acteurs principaux de la conservation/dégradation des molécules d'ADN. exception faite du permafrost, il apparaît que l'ADN ne peut se conserver sur une période de temps supérieure à 100 000 ans (Wayne *et al.*, 1999). Toutefois, suivant les conditions taphonomiques la durée de préservation des biomolécules dans les tissus peut varier de quelques dizaines à des milliers d'années.

II.3.1.2.1 La température

De nombreuses études ont démontré le caractère prépondérant de la température dans la conservation de l'ADN au cours du temps (Lindahl 1993, Hofreiter *et al.*, 2001 ; Smith *et al.* 2001, 2003 ; Willerslev *et al.*, 2004b) et, par conséquent, l'avantage de travailler sur des échantillons provenant de zones froides voire sur des sujets retrouvés dans le permafrost ou inhumés dans des tombes gelées (Keyser *et al.*, 2003 ; Ricaut *et al.*, 2005a, 2005b). Les dégradations subies par l'ADN des sujets conservés dans le permafrost seraient uniquement liées à des dégradations entropiques et non pas aux processus d'altération décrits plus haut (Mitchell *et al.*, 2005). Les restes les plus anciens étudiés avec succès ont été obtenus à partir de sujets retrouvés dans le permafrost (Hoss *et al.*, 1994 ; Gilbert *et al.*, 2004 ; Shapiro *et al.*, 2004). De plus, de l'ADN extrait à partir d'ossement datant de l'Holocène et du Pléistocène ont permis l'amplification de fragments compris entre 900 et 1000pb révélant l'excellente qualité de l'ADN endogène de ces sujets (Barnes *et al.*, 2001 ; Lambert *et al.*, 2002).

Il est important de noter qu'une absence ou une faible amplitude thermique annuelle du lieu de conservation est également cruciale. Smith *et al.*, (2003) ont proposé un modèle d'estimation de la durée de conservation possible de l'ADN en fonction de la taille des fragments, de la température moyenne au cours du temps, du taux de dépurination et du pH. Cependant il apparaît nécessaire d'évaluer le type de dégradations pouvant survenir dans les différents milieux plutôt que de proposer un modèle général puisque les dégradations sont fortement corrélées aux différents facteurs environnementaux.

II.3.1.2.2 Le pH

Les milieux dont le pH est neutre ou légèrement alcalin présentent les conditions les plus favorables à la conservation des acides nucléiques (Lindahl 1993). Le permafrost présent en Sibérie possède un pH neutre ce qui pourrait influer positivement sur la conservation de l'ADN des restes retrouvés dans cette région (Willerslev *et al.* 2004). A l'inverse les milieux acides auraient pour effet une altération du squelette d'hydroxyapatite et par conséquent une dégradation accélérée des restes osseux et dentaires. L'utilisation de natron sur les momies égyptiennes lors de l'embaumement permet une dessiccation rapide ainsi qu'une augmentation du pH qui pourrait entraîner une meilleure préservation de l'ADN (Gilbert *et al.*, 2005).

II.3.1.2.3 L'humidité

La présence d'eau ou un degré d'humidité important sont des facteurs qui favorisent les réactions d'hydrolyse ainsi que l'activité des microorganismes (Lindhal, 1993). Les milieux secs et tempérés qui permettent une dessiccation des tissus sont donc plus favorables à la conservation de l'ADN en raison de la diminution des lésions hydrolytiques (Pääbo, 1984). Le permafrost sibérien correspond bien à ces critères puisque c'est un milieu relativement sec. En effet, seul 3-8% de l'eau est présente sous forme liquide suivant le type de sédiment et la température (Willerslev *et al.*, 2004).

II.3.1.2.4 La quantité d'oxygène

Les processus d'oxydation ainsi que la prolifération de microorganismes aérobies sont directement liés à la teneur en oxygène de l'environnement dans lequel les restes sont conservés. Une meilleure préservation des acides nucléiques sera donc observée dans les milieux anoxiques tels certaines tourbières (Hauswirth *et al.*, 1994), des carrières à craie (Colson *et al.*, 1997) ou encore les gisements de goudron naturel (Janczewski *et al.*, 1992). Dans le cas de notre étude, le permafrost sibérien présente la particularité d'être un milieu relativement anaérobie. En effet, la teneur en méthane ainsi que le potentiel redox mesuré dans le permafrost du nord-est de la Sibérie semblent indiquer une faible concentration en oxygène (Vishnivetskaya *et al.*, 2001 *in* Willerslev *et al.*, 2004).

II.3.1.2.5 Le rayonnement ultraviolet

L'exposition de l'ADN aux radiations UV va générer des dimères de cyclobutane-pyrimidines, la formation de liaisons croisées ADN-ADN ou ADN-protéines ou encore des ruptures de brin (Hall et Ballantyne, 2004). L'exposition de restes osseux au soleil durant une période prolongée aura pour conséquence des modifications chimiques des bases azotées ainsi que des ruptures de brin de la molécule d'ADN (Hussein, 2005). Les ossements ayant été exposés au soleil sont aisément reconnaissables. En effet, au niveau macroscopique il est possible d'observer des modifications dans la structure de l'os : blanchissement de la surface exposée, friabilité, fentes, etc. Il est cependant parfois possible d'obtenir des résultats à partir d'un os retrouvé en surface en extrayant l'ADN à partir de matériel osseux prélevé sur la surface non exposée. Un enfouissement trop superficiel peut donc être responsable d'une mauvaise préservation de l'ADN (Zink et Nerlich, 2003).

Les différents facteurs environnementaux sont les principaux acteurs qui vont déterminer la bonne conservation de l'ADN (Gilbert *et al.*, 2003). Cependant d'autres éléments entrent en jeu comme (i) le type de substrat : les substrats dits durs comme les os et les dents semblent plus favorables à une conservation de l'ADN sur de longues périodes ; (ii) les conditions de stockage des échantillons ; (iii) le temps écoulé entre le décès et l'inhumation dans un milieu qui permettra la stabilisation des conditions de dégradations (iv) ou les pratiques funéraires qui peuvent influer directement comme les processus de momifications artificielles.

II.3.1.3 *Conditions environnementales et taphonomiques dans notre étude*

Comme nous l'avons vu précédemment, le permafrost du nord-est sibérien semble être un milieu particulièrement favorable à la conservation de l'ADN car il regroupe nombre des conditions requises. Cet environnement présente tout d'abord une température basse et surtout relativement stable au cours de l'année. Le pH, l'humidité et la concentration en oxygène sont également à des niveaux qui limitent les dégradations pouvant survenir sur la molécule d'ADN.

Certaines pratiques funéraires des Yakoutes pourraient également favoriser la conservation de l'ADN. En effet, lors des inhumations l'intérieur et l'extérieur des coffres ou des cercueils étaient généralement recouverts d'une couverture en écorce de bouleau. Ce matériau est celui qui était traditionnellement utilisé pour couvrir les *uraha* (grandes tentes coniques servant à

l'habitat d'été). D'après les observations réalisées sur le terrain, cette couverture permet une bonne isolation par rapport aux eaux de ruissellement qui pourraient s'infiltrer durant les précipitations survenant durant les mois d'été (cf Figure 18).



Figure 18 : Différentes couvertures en écorce de bouleau sur la partie supérieure externe du coffre, partie intérieure recouvrant le corps et sur le fond du coffre. Tombe de *Munur Urekha*.

II.3.1.4 Influence des dégradations post-mortem sur l'analyse et l'interprétation des résultats

Des conditions taphonomiques va dépendre la qualité de l'ADN qu'il sera possible d'extraire des échantillons anciens. Comme nous l'avons vu précédemment, les lésions subies par l'ADN au cours du temps vont conduire à des fragmentations du brin d'ADN, à la formation de sites abasiques, de liaisons croisées ou encore à la modification des thymines et cytosines en hydantoïnes. Ces différents types de dégradations ont pour effet de bloquer l'activité de la polymérase et d'empêcher l'amplification. Cependant, une partie des modifications *postmortem* n'influe pas sur l'amplification mais conduit à des erreurs (Pääbo, 1989). Ces lésions modifient la séquence répliquée (Fattorini *et al.*, 1999) et peuvent être à l'origine d'interprétations erronées lors de l'analyse des résultats (Gilbert *et al.*, 2003b). Ces erreurs vont se produire en plus des erreurs de réPLICATION dues à la Taq polymérase indépendamment de la qualité de la molécule d'ADN de départ.

Les dégradations au cours du temps ne semblent pas se produire de manière aléatoire tant au niveau du type de lésions que de leur localisation. Les changements les plus fréquemment rencontrés sont générés par les déaminations de la cytosine en uracile (anologue de la thymine) et de l'adénine en hypoxanthine (anologue de la guanine) (Hansen *et al.*, 2001 ; Hofreiter *et al.*, 2001). Suivant la nomenclature proposée par Hansen (2001) en tenant compte du brin complémentaire, les changement de type A→G/T→C sont désignés comme "type 1" et les changements C→T/G→A sont de "type 2". Les dégradations de type 1 sont observées plus fréquemment dans les échantillons anciens (Gilbert *et al.*, 2003b). Ces erreurs pourront être transmises lors de l'amplification à l'ensemble des amplicons en particulier si le nombre de molécules de départ est faible.

Concernant l'ADN mitochondrial, il semble que des lésions se produisent de manière préférentielle dans la région HV1 par rapport aux régions codantes (Gilbert *et al.*, 2003a). De plus les mutations surviennent avec des fréquences plus élevées pour certaines positions qualifiées de « *hot spots* » (Gilbert *et al.*, 2003a). Cette vulnérabilité plus importante serait directement liée à la structure secondaire de la région HV1 qui la rendrait plus sensible aux dégradations. Cette région étant largement étudiée dans les études paléoanthropologiques en raison de son héritabilité strictement maternelle, de l'absence de recombinaison et de la facilité de comparaison des données anciennes avec les bases de données actuelles, rendent encore plus critiques les erreurs d'interprétation.

Pour Svante Pääbo (1989b, 1990), une autre source d'erreur d'interprétation serait représentée par les amplicons générés lors d'évènement de *jumping PCR*. Il semblerait qu'une corrélation positive puisse exister entre l'état de dégradation des molécules d'ADN et ce phénomène. Il apparaîtrait possible de déterminer dans quel cas un évènement de *jumping PCR* soit survenu en comparant les mutations présentes sur les séquences de plusieurs clones (Pääbo, 1990).

II.3.1.5 Les méthodes de "réparation" de l'ADN

Plusieurs méthodes ont été testées afin d'essayer de dépasser les problèmes liés aux lésions post-mortem. Ces méthodes visent soit à améliorer la qualité de l'ADN présent dans l'extrait, soit à contourner les problèmes d'amplification liés à la nature dégradée de la matrice. Le paragraphe suivant ne présente pas une liste exhaustive de l'ensemble des méthodes mais propose un aperçu des différentes techniques les plus couramment rencontrées.

L'action conjointe de la polymérase I d'*E. coli* et de la T4 ADN ligase peut permettre de réparer les ruptures double-brin causées par les lésions hydrolytiques, oxydatives ou enzymatiques (Pääbo, 1989a; Pusch *et al.* 1998 ; di Bernardo *et al.*, 2002 ; Pusch *et al.*, 2002). La polymérase I d'*E. coli* va déplacer les sites abasiques alors que la T4 ADN ligase permettra de combler les espaces entre les extrémités 3'OH et 5'P. Les sites comportant ces extrémités étant peu fréquents, la quantité de sites réparés sera relativement faible (Hansen *et al.*, 2006). De plus, la faible fiabilité de la polymérase a pour conséquence de générer des erreurs lors des amplifications (Pusch *et al.*, 2002).

Une autre technique permettant de réparer les ruptures de brins a été proposée récemment (Hansen *et al.*, 2006). Cette méthode est basée sur un premier traitement enzymatique par la polynucléotide kinase T4 et l'AP endonucléase I qui va permettre l'ajout de groupements phosphates et hydroxyles sur les extrémités 5' et 3'. Ce traitement est complété par l'utilisation d'une polymérase du type ADN polymérase II d'*E. coli* et de la T4 ADN ligase.

Le taux d'erreur d'incorporation de nucléotides par l'ADN polymérase est de l'ordre de 10^{-6} en moyenne et est insensible aux légères variations de stringence du milieu de réaction. L'utilisation de polymérases possédant des taux d'erreur de réPLICATION très faible (*e.g.* Taq

Platinum Hifidelity (Invitrogen) permet de s'affranchir des erreurs pouvant survenir lors de l'amplification et qui ne seraient pas liées à la qualité de l'ADN présent dans l'extrait (Gilbert *et al.*, 2003). L'utilisation de ce type de polymérase permet donc de minimiser les erreurs mais aussi d'augmenter l'efficacité d'amplification (Willerslev *et al.*, 1999 ; Cooper *et al.*, 2001).

Afin de contourner les erreurs possibles liées à la déamination de la cytosine, il est possible d'utiliser l'uracile-N-glycosilase d'*E. coli* (Hofreiter *et al.*, 2001). Cette enzyme va supprimer les uraciles de l'ADN ce qui entraîne la formation de sites abasiques qui sont hydrolysés par β -élimination, ce qui a pour conséquence une rupture de brin. Ce traitement permet donc d'éliminer efficacement les bases responsables des erreurs de type G→A.

L'utilisation de PTB (Bromure de Phénacyl Thiazolium) peut être considéré comme une méthode de réparation puisque ce réactif permet de cliver les liaisons croisées entre l'ADN et les protéines (Vasan *et al.*, 1996). Le mode d'action de ce réactif n'est pas clairement établi cependant son efficacité a été confirmée par plusieurs publications (Poinar *et al.*, 1998, Hofreiter *et al.*, 2000 ; Krings, *et al.*, 2000) ainsi que par nos travaux (cf chapitre IV.3.2.3 Influence du PTB).

L'amplification à l'aide d'amorces dégénérées et de températures d'hybridation basses (*Degenerated Oligonucleotide-Primed* ou DOP-PCR) vise à augmenter le nombre de copies d'ADN initialement amplifiées (Telenius *et al.*, 1992). Une fois cette première étape effectuée, une amplification du fragment d'intérêt est réalisée dans les conditions normales. Cette méthode aurait tendance à favoriser l'amplification de contaminants (Kaestle et Horsburg, 2002).

Méthode d'amplification du génome entier (*whole genome amplification*) :

La perspective d'augmenter la quantité globale d'ADN présente dans un extrait représenterait une avancée incroyable dans les études en paléogénétique. En effet, il serait possible pour les échantillons extrêmement précieux de répéter un grand nombre de fois les analyses sans porter préjudice au sujet par le prélèvement de grandes quantités de matériel. Cette méthode permettrait d'éviter les artefacts fréquemment rencontrés lors de l'analyse d'extraits contenant un faible nombre de copies d'ADN.

Les méthodes d'amplification totale du génome sont basées sur l'utilisation de polymérases possédant une activité de déplacement de brin en combinaison avec des hexamères qui vont s'hybrider aléatoirement dans tout le génome. Par exemple, la polymérase Phi29 utilisée dans le kit Genomiphi (Amersham) possède une grande processivité ainsi qu'un taux d'erreur 100 fois inférieur à celui de la Taq polymérase (Esteban *et al.*, 1993).

II.3.2 Les inhibitions

Les lésions subies par l'ADN ancien au cours du temps ne sont pas les seules difficultés rencontrées lors des études en anthropologie moléculaire. L'amplification des acides nucléiques peut être grandement compliquée par la présence d'inhibiteurs. Ces substances regroupent des composés organiques ou inorganiques qui sont co-purifiés avec l'ADN lors de l'extraction et qui exercent une action néfaste sur l'efficacité de la Taq polymérase. L'inhibition peut être totale ou partielle et peut conduire à des résultats faussement négatifs si aucun témoin n'est inclus dans la réaction d'amplification.

II.3.2.1 Les différents types d'inhibiteurs

Les inhibiteurs les plus fréquemment rencontrés sont des substances contenues dans le sol comme les acides humiques et fulviques, les tannins ainsi que les ions métalliques (Hummel *et al.*, 1992 ; Tuross, 1994). Les acides humiques et fulviques entraînent la chélation des ions Mg^{2+} (Tsai et Olson 1991) ce qui diminue l'efficacité de la PCR. La concentration de ce type d'inhibiteur est principalement liée aux plantes présentes dans le milieu et aux facteurs environnementaux comme la température et le pH (Hummel, 2003). Un inhibiteur, dont la nature reste à déterminer, a été retrouvé dans la fraction microbienne du sol. Cette substance diffère des acides humiques et fulviques et pourrait constituer le principal inhibiteur présent dans plusieurs types de sol (Watson et Blackwell, 2000). La présence de ce type d'inhibiteurs est généralement marquée par une coloration brunâtre de l'extrait d'ADN (Cooper *et al.*, 1992).

Les substances produites par le corps constituent également une source importante d'inhibiteurs. Les composés présents dans les fluides corporels tels que l'urée (Kang *et al.*, 1991), l'héparine (Beutler *et al.*, 1990) ou les humeurs aqueuse ou vitrée (Limpens et Kijlstra, 1993, Wiedbrauk *et al.*, 1995) peuvent être responsables d'une inhibition. Les porphyrines, comme l'hème de l'hémoglobine (Akane *et al.*, 1994) et d'autres ions métalliques, les cytochromes, le collagène de type I (Scholz *et al.*, 1998) ont également une action inhibitrice sur la Taq polymérase. La mélanine, protéine responsable de la pigmentation des cheveux, est

une des substances inhibitrices rencontrées dans les études sur l'ADN ancien ou l'identification des individus en médecine légale (Uchihi *et al.*, 1992).

L'ADN endogène fortement fragmenté (Pääbo *et al.*, 2004, Pusch *et al.*, 1998), l'ADN bactérien (Pääbo, 1989) et les complexes sucres-protéines, ou complexes de Maillard, constituent également des inhibiteurs de la réaction de PCR.

II.3.2.2 Mode d'action des inhibiteurs

Les inhibiteurs peuvent perturber la réaction d'amplification par leurs interactions avec les molécules d'ADN ou par leur actions sur la Taq polymérase.

II.3.2.2.1 Action sur la molécule d'ADN

La séquestration de la molécule d'ADN au sein des complexes de Maillard peut entraîner une impossibilité d'amplification. La présence de débris de bactéries, de protéines ou de polysaccharides peut avoir des effets physiques qui empêchent la polymérase de s'hybrider à sa cible (Wilson, 1997). La présence en concentration importante de courts fragments d'ADN endogène ou contaminant peut entraîner une inhibition de la réaction en limitant la possibilité que deux amorces complémentaires s'hybrident sur le même fragment (Pusch *et al.*, 1998).

II.3.2.2.2 Action sur la Taq polymérase

Les inhibiteurs peuvent influer négativement sur l'amplification par PCR en agissant sur la liaison de la Taq polymérase sur la matrice comme les composés phénoliques (Simon *et al.*, 1996), en ralentissant son activité ou encore en la dégradant. Les acides humiques notamment vont interférer lors de la liaison de l'enzyme à sa cible (Tebbe et Vahjen, 1993) et engendrent une inhibition à partir de concentration très faible dans l'extrait ($0,1\mu\text{g/ml}$; Tsai *et al.* 1991).

La mélanine présente dans les cheveux conduit à la formation de complexes entre la Taq polymérase et d'autres protéines ce qui a pour conséquence de diminuer la capacité d'elongation de l'enzyme (Eckart *et al.*, 2000). L'influence de cet inhibiteur est donc plus marqué lors de l'amplification de longs fragments (Eckart *et al.*, 2000 ; Amory *et al.*, 2006).

II.3.2.3 Les méthodes de détection des inhibiteurs

Les échecs d'amplification peuvent être liés au fait que l'ADN endogène est trop fortement dégradé ou présent en trop faible quantité. Cependant la présence d'un inhibiteur peut toujours être suspectée. Il est donc nécessaire de s'assurer de la présence ou de l'absence de substances inhibitrices dans l'extrait analysé. Pour ce faire plusieurs méthodes existent, les plus couramment utilisées étant : la coloration de l'extrait, le dépôt d'une fraction de l'extrait sur gel ou la quantification par PCR en temps réel.

Il est possible de détecter la présence d'inhibiteurs après l'étape d'extraction phénolique grâce à la coloration du surnageant. Une couleur allant du jaune au brun foncé peut indiquer la présence de substances inhibitrices telles que des acides humiques ou des produits de Maillard (Cooper *et al.*, 1992). Cette méthode, bien que subjective, c'est révélée fiable pour certains de nos échantillons, notamment YAKa33.

Une fois l'extraction effectuée, le dépôt de l'extrait d'ADN sur un gel d'électrophorèse ne contenant pas de BET permet de détecter la présence d'inhibiteurs dans l'extrait si une coloration bleutée apparaît (Pääbo, 1989 ; Cooper *et al.*, 1992).

La quantification réalisée après chaque extraction à l'aide du kit QuantifilerTM (Applied Biosystems) (cf III.3.1 Quantification de l'ADN par PCR en temps réel) permet de distinguer la présence d'inhibiteurs de manière fiable grâce à la présence dans le mélange réactionnel d'un témoin interne de PCR (IPC : *Internal PCR Control*). Cet IPC est une courte séquence nucléotidique totalement synthétique ne correspondant à aucune séquence retrouvée dans la nature. L'amplification de ce témoin possède une cinétique connue, par conséquent, tout retard d'amplification indique la présence d'un facteur interférant lors de la PCR. L'intérêt de cette méthode est qu'elle rend possible la visualisation des inhibitions avant de commencer les analyses sur les différents marqueurs moléculaires, ce qui implique une économie de matériel génétique précieux et de réactifs onéreux. En outre, il est possible d'estimer la teneur approximative de l'extrait en inhibiteur en fonction du retard accusé par l'IPC lors de l'amplification.

II.3.2.4 Les méthodes de purification des inhibiteurs

La présence d'inhibiteurs dans les extraits d'ADN ancien est un facteur bien connu, par conséquent les protocoles appliqués prennent en compte ce paramètre afin de limiter au maximum la propagation des substances inhibitrices au cours des différentes étapes de l'analyse.

II.3.2.4.1 L'extraction

L'extraction constitue la première étape permettant de purifier l'ADN des éventuels inhibiteurs. En effet, c'est durant cette phase que l'ADN doit être isolé de tous les autres composés pouvant être présents dans l'échantillon.

Certains réactifs chimiques (DMSO, bleu de dextran, NaOH, formamide collagénase) ont des actions spécifiques sur des inhibiteurs donnés. Cependant il est difficile de déterminer la substance présente dans l'échantillon et donc le réactif à employer. De plus, certains de ces produits peuvent avoir un impact néfaste et devenir à leur tour des inhibiteurs si leur concentration n'est pas dosée avec précision (spermidine, formamide, bleu de dextran) (Wilson *et al.*, 1997 ; Scholz *et al.*, 1998 ; Kalmar *et al.*, 2000). L'utilisation de ces réactifs est par conséquent à éviter si une autre solution est envisageable.

Selon Kathryn Watt (2005), la décontamination des échantillons par l'utilisation d'hypochlorite de sodium (NaOCl) permettrait une diminution sensible de la concentration en inhibiteurs dans les extraits d'ADN. Les conclusions de cette étude, portant sur un nombre relativement important de sujets (41 échantillons), ont été confirmés par Salamon *et al.* (2005) qui ont également observé une réduction de l'incidence des inhibiteurs après un traitement prolongé avec de l'hypochlorite de sodium.

Comme mentionné précédemment, l'ajout de PTB dans le tampon de lyse va permettre de dissocier les complexes de Maillard qui se forment entre les sucres et les protéines (Poinar *et al.*, 1998). L'efficacité de ce réactif a été évaluée au cours de cette étude sur différents substrats (cf IV.1.2 Utilisation du PTB et IV.3.2.3 Influence du PTB).

L'extraction phénol-chloroforme permet de séparer les acides nucléiques des résidus protéiques, des divers composés phénoliques (acides humiques et fulviques, tannins) ainsi que des autres substances présentes dans le tissu source (O'Rourke *et al.*, 2000). Cette étape peut être répétée plusieurs fois si nécessaire pour les cheveux par exemple ou dans le cas

d'échantillons particulièrement susceptibles de contenir des inhibiteurs (*e.g.* ossements enfouis dans un sol riche en matières organiques végétales).

Plusieurs techniques de purification post-extraction sont envisageables afin de compléter les étapes précédentes. L'utilisation de colonnes à membrane de silice permet de fixer de manière spécifique les acides nucléiques en raison de l'affinité de la silice pour l'ADN (Hoss et Pääbo, 1993). Cette méthode est très efficace, rapide et peu onéreuse mais elle comporte cependant un inconvénient : l'ADN exogène a également tendance à se fixer sur la silice ce qui peu représenter une source de contamination importante. Les colonnes avec filtre de Sephadex peuvent également être utilisées pour la purification des échantillons et donnent des résultats satisfaisants (Miller *et al.*, 1999). Les méthodes utilisant le Chelex (Walsh *et al.*, 1991) se sont révélées efficaces sur certains types de substrats, notamment les cheveux (Suenaga et Nakamura, 2005). Néanmoins certains auteurs (Miller *et al.*, 1999 ; Kalmar *et al.*, 2000 ; Castella *et al.*, 2006) ont décrit cette technique comme moins performante que les méthodes utilisant les filtres de silice.

L'ultrafiltration sur microconcentrateur permet, via l'utilisation de membranes semi-perméables, d'éliminer les molécules de tailles inférieures au seuil d'arrêt (*cut off*) dont la valeur varie suivant le modèle de microconcentrateur. Cette technique permet d'éliminer les courts fragments d'ADN qui pourraient se révéler inhibant lors de l'amplification. Outre cette propriété, l'ultrafiltration vise à concentrer les extraits en évitant toute dégradation mécanique des acides nucléiques.

La précipitation à l'isopropanol est parfois utilisée (Hänni *et al.*, 1995) cependant ce type de technique est généralement plus adapté à de l'ADN non dégradé et relativement concentré.

II.3.2.4.2 L'amplification

Malgré les précautions décrites ci-dessus il arrive que ces traitements de soient pas suffisants et que des substances inhibitrices soient toujours présentes dans l'extrait d'ADN purifié. Plusieurs solutions peuvent alors être envisagées :

Le polyvinylpolypirrolidone (PVPP) ainsi que le polyvinylpyrrolidone (PVP) permettent de séparer les composés humiques de l'ADN par électrophorèse sur gel d'agarose (Young *et al.*, 1993). Cependant une concentration de PVP supérieure à 0,5% a tendance à réduire le rendement de la réaction d'amplification.

Une méthode simple et efficace pour lever les inhibitions consiste à diluer l'extrait afin de réduire l'action des inhibiteurs jusqu'à un seuil qui permette l'amplification. Il suffit de réaliser une gamme de dilutions en cascade de 1/2 à 1/10 (qui permet généralement d'atteindre un taux de dilution correct). Le point faible de cette méthode réside évidemment dans le fait que l'ADN sera dilué dans la même proportion que les inhibiteurs ce qui peut s'avérer critique lors de l'analyse d'extraits de départ très faiblement concentrés.

Certains réactifs peuvent être ajoutés directement dans le milieu réactionnel lors de l'amplification. La BSA (*Bovine Serum Albumine*) permet de s'affranchir de l'action inhibitrice des porphyrines, des composés phénoliques présents dans les acides humiques et fulviques ainsi que des acides tanniques en empêchant la liaison de ces composés à la Taq polymérase (Kreader, 1996). Il est cependant nécessaire de limiter la concentration de BSA dans le mélange réactionnel car ce réactif peut avoir une action inhibitrice au delà de 1 μ g/ μ l (Lin *et al.*, 1995). La protéine gène 32 (gp32) permet également de diminuer l'inhibition (Panaccio et Lew, 1991 ; Widjojoatmodjo *et al.*, 1992) ; cependant ce réactif est peu utilisé car son efficacité est similaire à celle de la BSA et son coût plus élevé (Kreader *et al.*, 1996).

II.3.3 Les contaminations

II.3.3.1 Les sources de contaminations

Les contaminations représentent l'obstacle principal de toute étude génétique ayant pour sujet des échantillons anciens. Ce problème est principalement lié à l'état dégradé des molécules que l'on désire étudier ainsi qu'à la méthode d'amplification utilisée. En effet, l'extrême sensibilité de la PCR permet, en théorie, l'amplification d'une seule copie d'ADN. Cependant une molécule intacte sera amplifiée préférentiellement à une molécule présentant des lésions structurales ou chimiques. L'exemple des études portant sur des restes de néandertaliens permet de bien comprendre l'importance des contaminations dans les analyses paléogénétiques. Dans une étude récente, l'équipe de Serre *et al.* (2004) a montré que sur 24 échantillons néandertaliens provenant de divers pays d'Europe, seuls 4 contenaient de l'ADN amplifiable à l'aide d'amorces spécifiques des séquences mitochondrielles néandertaliennes. Toutefois, l'utilisation dans cette étude d'amorces spécifiques aux séquences humaines a révélé la présence d'ADN contaminant humain dans des échantillons osseux provenant des sujets Néandertaliens et également de restes animaux (ours des cavernes). Lors des études en anthropologie moléculaire, les contaminations par de l'ADN exogène humain sont encore plus problématiques puisque l'ADN contaminant devient difficilement discernable de l'ADN endogène. L'utilisation de locaux et de matériel adaptés ainsi que la mise en place de protocoles rigoureux est par conséquent impérative.

La difficulté majeure dans la détection et la prévention des contaminations provient de l'importance des sources possibles de contaminants que ce soit avant la découverte du sujet, lors de la fouille ou au sein du laboratoire.

Les contaminations les plus difficiles à mettre en évidence sont celles qui peuvent survenir avant même la découverte des restes. Les sépultures secondaires qui impliquent un déplacement des restes d'un sujet peuvent entraîner une contamination par les personnes participant à ces rites funéraires. Les pillages peuvent également représenter une source potentielle de contaminations. Ces deux cas de figure peuvent être mis en évidence lors de la fouille, il est donc impératif d'évaluer ces risques par une analyse anthropologique avant de débuter toute étude moléculaire. L'importance de ces contaminations survenant en amont des

analyses en laboratoire a été soulignée dans plusieurs études récentes (Pääbo *et al.*, 2004 ; Gilbert *et al.*, 2005b ; Willerslev et Cooper 2005, Lalueza-Fox *et al.*, 2006).

Lors de la fouille, si certaines précautions ne sont pas prises (cf chapitre II.4.3.2.), le nombre généralement important de personnes ayant accès aux échantillons entre la découverte des restes et l'analyse en laboratoire augmente considérablement les risques de contamination. Les sujets conservés dans les musées représentent une difficulté supplémentaire considérant l'impossibilité de retrouver l'ensemble des personnes ayant participé à la découverte, impliquées dans le stockage, ayant eu accès ou simplement ayant approché le matériel (e.g. les visiteurs).

Enfin, des contaminations peuvent survenir dans le laboratoire. Les expérimentateurs représentent une source non négligeable de contamination cependant ce type de contamination reste aisément identifiable grâce au typage des manipulateurs. L'ADN amplifié représente par contre un type de contamination beaucoup plus pernicieux. En effet, lors d'une PCR le nombre de copies du fragment amplifié peut atteindre 10^{12} à 10^{15} . Les mouvements d'air créés lors de l'ouverture des tubes ou des mouvements de liquide vont générer et disperser des gouttes microscopiques pouvant contenir jusqu'à un million de copies d'un fragment d'ADN dans 0,005µl (Willerslev *et al.*, 2005). Les produits d'amplification peuvent donc se disperser rapidement dans les laboratoires et même dans des immeubles entiers à la faveur des systèmes d'air conditionné. Considérant qu'une seule de ces microgouttes contient mille fois plus d'ADN que ce qu'il est possible d'extraire d'un gramme de matériel osseux (10^5 à 10^6 copies ; Handt *et al.*, 1996, Cooper *et al.*, 2001) le risque de contaminations par ces aérosols est particulièrement important.

L'ADN humain et microbien est présent dans tous les laboratoires. Il est par conséquent nécessaire de considérer tout réactif ou matériel de laboratoire provenant du fournisseur comme potentiellement contaminé (une circulaire provenant d'Applied Biosystems concernant la contamination de tous les emballages de kit par l'échelle allélique a été envoyée aux utilisateurs de certains kit durant l'année 2005). Il est important de prendre en compte que les réactifs dits stériles ne sont pas toujours exempts de tout acide nucléique.

Enfin, les extraits d'ADN provenant d'échantillons anciens peuvent contenir certaines molécules comme des sucres ou de l'ADN de microorganismes qui vont jouer le rôle de transporteur durant la PCR (Pääbo, 1990). La présence de ces molécules va entraîner l'amplification de contaminants présents en faible quantité dans les extraits mais qui ne pourront pas être mises en évidence par les témoins négatifs.

II.3.3.2 Précautions nécessaires à l'étude de l'ADN ancien

La grande sensibilité de la PCR peut conduire à l'amplification préférentielle de toute molécule d'ADN contaminante si l'échantillon présente un état de dégradation trop important ou si le nombre de copies est trop faible. Il est donc nécessaire de respecter un certain nombre de précautions de la fouille jusqu'à l'analyse en laboratoire afin de s'assurer de la validité des résultats.

II.3.3.2.1 La fouille et le prélèvement des échantillons

En premier lieu, il est important que les équipes responsables de la fouille soient sensibilisées à la problématique de recherche ainsi qu'aux contraintes inhérentes aux études en anthropologie moléculaire. Cette discipline est en effet assez récente à l'échelle de l'archéologie et bien que de plus en plus d'archéologues s'intéressent à l'outil moléculaire en raison de ses potentialités, tous ne sont pas conscients des précautions nécessaires à cette discipline.

Ensuite, lors de la fouille il est impératif que tout échantillon soit manipulé à l'aide de gants en latex et lors de la description le port de masques faciaux est fortement recommandé. De nombreuses études ont en effet mis en évidence des contaminations ayant eu lieu en amont des analyses en laboratoire (Hofreiter *et al.*, 2001 ; Wandeler *et al.*, 2003 ; Serre *et al.*, 2004 ; Malmström *et al.*, 2005 ; Salamon *et al.*, 2005 ; Lazuela-Fox *et al.*, 2005 ; Sampietro *et al.*, 2006). Une fois l'échantillon prélevé il doit être placé dans un contenant hermétique ayant été de préférence décontaminé (les sachets minigrips se prêtent bien à cet usage car ils peuvent être aisément décontaminés et transportés) (Yang et Watt, 2005).

II.3.3.2.2 Typage des personnes impliquées

L'ADN de toutes les personnes ayant été en contact avec les échantillons (archéologues, anthropologues, fouilleurs, personnel du laboratoire) doit être analysé pour les mêmes marqueurs que ceux étudiés sur les échantillons anciens. Les résultats obtenus doivent toujours être comparés avec ceux de l'équipe afin d'écartier tout doute sur une éventuelle contamination.

Les échantillons conservés dans les musées représentent à ce titre un problème puisqu'il est impossible de collecter l'ADN de toutes les personnes ayant été en contact avec ces

échantillons. Le problème est encore plus marqué pour les échantillons mis au jour avant l'apparition de l'anthropologie moléculaire car aucune précaution n'était prise, ou pour des échantillons stockés durant de longues périodes car le nombre de personnes ayant eu accès aux échantillons n'est pas contrôlable. Les processus de décontamination sont donc extrêmement importants lors de l'analyse de ce type de prélèvements. En plus des problèmes de contamination, les conditions de stockage dans les musées, principalement la température, semblent favoriser les dégradations de l'ADN endogène (Gilbert *et al.*, 2006).

Afin de limiter le doute sur la validité des résultats il est préférable que les personnes impliquées dans l'analyse n'appartiennent pas à la même population que le sujet étudié (Richards *et al.*, 1995). Il sera en effet plus aisément de suspecter une contamination si un haplotype européen est retrouvé à partir d'extrait d'un sujet appartenant à une population asiatique par exemple.

II.3.3.2.3 Stockage

Une fois les échantillons arrivés au laboratoire, ils doivent être entreposés dans une pièce isolée afin de limiter les possibilités de contaminations croisées. La température de stockage est également importante. Les possibilités de conserver les échantillons anciens avant et après l'analyse à -20°C semble limiter les dégradations. Nous avons pu observer que la conservation à température ambiante durant quelques mois d'échantillons fragiles peut influer grandement sur la qualité des résultats obtenus.

II.3.3.2.4 Bonnes pratiques de laboratoire

En plus des précautions à appliquer habituellement dans un laboratoire de biologie moléculaire, l'étude de l'ADN ancien requiert que certaines règles soient respectées afin d'éviter la diffusion d'ADN aérosols, la contamination des réactifs, les contaminations croisées entre échantillons analysés simultanément ou la simple contamination par l'expérimentateur. Le stockage et le traitement pré-extraction (*i.e.* abrasion, trépanation et cryobroyage) des échantillons anciens doivent être effectués dans une pièce séparée du laboratoire où sont réalisées les analyses moléculaires. Pour cette même raison les pièces pré- et post-PCR doivent également être séparées et le mouvement des personnes, du matériel et des réactifs doit se faire uniquement dans le sens pré-PCR vers post-PCR. L'ensemble des

réactifs doit être aliquoté afin de limiter le nombre d'utilisation et par conséquent le risque de contamination.

II.3.3.3 Méthodes de décontamination

Malgré l'ensemble des précautions prises lors du prélèvement des échantillons sur le terrain, ou dans le cas d'un échantillon provenant d'un musée, il est nécessaire d'appliquer une procédure de décontamination. Pour les substrats osseux, trois méthodes sont généralement employées ; toutes basées sur le fait que la contamination est liée à des molécules présentes en surface.

II.3.3.3.1 Décontamination physique

La décontamination physique des échantillons a pour but d'éliminer les contaminants de surface pouvant être présents. Différents moyens peuvent être employés tels l'utilisation de lames de scalpels stériles, de papier de verre ou encore d'une perceuse de type Dremel. La principale difficulté de cette technique est de déterminer combien de millimètres doivent être abrasés. Le problème ne se pose pas pour les échantillons possédant une corticale épaisse comme les diaphyses d'os longs, cependant dans le cas de fragments osseux ou pour les os de sujets immatures il n'est possible d'abréger la surface que sur un ou deux millimètres. La décision doit par conséquent être prise au cas par cas. L'utilisation de la perceuse Dremel (outil utilisé dans notre étude) est avantageuse car la décontamination des fraises ainsi que de l'appareil à l'aide d'un four à UV est aisée. Néanmoins, pour les fragments osseux de petite taille, le risque que le gant soit pris accidentellement dans la fraise peut ajouter un facteur de contamination. Il est également important de limiter la vitesse de rotation de la fraise afin d'éviter tout échauffement de l'os, ce qui pourrait avoir un effet délétère pour l'ADN endogène.

II.3.3.3.2 Décontamination chimique

La décontamination chimique utilise les propriétés d'altération de la molécule d'ADN de différentes solutions comme l'hypochlorite de sodium (NaOCl) qui provoque des lésions oxydatives et la dénaturation de l'ADN double brin par destruction des liaisons hydrogènes (Hawkins *et al.*, 2002 et 2003, Whiteman *et al.*, 2002), l'acide chlorhydrique qui entraîne des dépurations et hydrolyses (Savage et Plaut, 1958, Whiteman *et al.*, 1997) ou encore l'EDTA qui va permettre de décalcifier superficiellement l'échantillon (Kolman et Tuross, 2000). L'efficacité de cette méthode a été testée dans différentes études (Richard *et al.*, 1995 ; Kemp et Smith, 2005 ; Watt, 2005) suivant plusieurs protocoles. Cependant aucun consensus ne se dégage de ces publications que ce soit sur le mode de traitement, par immersion ou lavage, la concentration de la solution à utiliser ou le temps de contact. Deux études récentes (Watt, 2005 ; Salamon *et al.*, 2005) apportent un élément intéressant puisqu'il semblerait que l'utilisation de NaOCl permette d'éliminer ou du moins de diminuer l'effet des inhibiteurs. L'utilisation de solutions permettant de dégrader l'ADN présente tout de même un risque pour l'ADN endogène de l'échantillon. En effet, dans le cas d'ossements poreux une infiltration du produit en profondeur est possible et aura pour conséquence une altération de l'ADN du sujet.

II.3.3.3.3 Décontamination des locaux et du matériel de laboratoire

La décontamination des locaux, du matériel et des réactifs représente une étape impérative entre l'analyse de chaque échantillon ancien. Les techniques les plus fréquemment utilisées sont l'irradiation UV (254nm) et l'autoclavage, méthodes dérivées de celles employées en microbiologie, qui permettent l'élimination des microorganismes et la fragmentation de l'ADN de haut poids moléculaire. Cependant, elles ne garantissent pas l'élimination de toute trace d'ADN et il peut toujours subsister des fragments de petite taille (inférieure à 150 pb). De plus, l'utilisation d'autoclave peut entraîner la diffusion d'ADN bactérien (Willerslev et Cooper, 2005). Il est donc nécessaire de compléter ces deux traitements par l'emploi d'acide chlorhydrique (2,5 M), d'hypochlorite de sodium (concentré à 50%) ou de solutions contenant des DNases du type DNA Away (Molecular Bioproducts).

L'ensemble des processus de décontamination décrits ci-dessus doit garantir l'authenticité des résultats. Il est toutefois nécessaire de mettre en place des lignes directrices afin de valider les données obtenues.

II.4 Critères d'authentification des résultats

Les premiers critères d'authentification furent proposés par Svante Pääbo en 1989. Ils étaient limités à trois points : (i) des témoins négatifs doivent être inclus lors de chaque extraction et amplification afin de détecter d'éventuels contaminants dans les réactifs utilisés lors de l'extraction ; (ii) les extractions et les amplifications doivent être dupliquées pour un même échantillon et les résultats obtenus doivent être identiques ; (iii) une corrélation négative doit être observée entre l'efficacité de la PCR et la taille des fragments amplifiés en raison de la fragmentation de la molécule d'ADN.

Le risque important de contamination par de l'ADN exogène, une meilleure compréhension des processus de dégradation des acides nucléiques ainsi que le nombre important d'études sur l'ADN ancien à la fin des années 1990 ont conduit à la publication en 2000 par Cooper et Poinar d'une liste plus aboutie de critères d'authentification. Ces critères, majoritairement adoptés aujourd'hui, représentent les lignes directrices à prendre en compte lors des analyses sur l'ADN ancien, en plus du cahier des charges déjà établi concernant les bonnes pratiques de laboratoire.

La liste a été étendue au fur et à mesure de la découverte des nouvelles altérations survenant au cours du temps. Les critères peuvent donc varier en fonction des auteurs mais ceux présentés ci-après sont retrouvés dans la plupart des études.

- L'état de préservation biochimique d'un échantillon est une preuve indirecte de la conservation de l'ADN endogène. L'analyse biochimique des échantillons peut servir à sélectionner les échantillons afin d'étudier uniquement ceux dont l'ADN semble préservé. Différentes méthodes ont été développées pour ce faire. La plus couramment utilisée repose sur l'étude de la racémisation des acides aminés présents dans l'échantillon (Poinar *et al.*, 1996). Les facteurs environnementaux, comme la température et l'humidité, responsables de la dégradation de l'ADN influent sur le changement progressif de configuration des énantiomères Lévogyre en Dextrogyre. Il a ainsi été établi que le taux de racémisation de l'acide aspartique était proche de la vitesse de dépurination de l'ADN (Poinar *et al.*, 1996). La proportion de chaque énantiomère de cet acide aminé permet de déterminer dans quelle mesure l'ADN d'un échantillon donné sera étudiable. Bien que cette méthode soit largement utilisée, il faut toutefois nuancer cette hypothèse. Les vitesses de racémisation *in vitro*

utilisées par Poinar seraient différentes de celles observées dans les tissus. La vitesse de racémisation serait, par ailleurs, différente suivant le tissu considéré (Collins *et al.*, 2002). De plus, certains facteurs (i.e. présence d'aspartyl et d'asparagine, séquence et conformation des protéines) influeraient sur la racémisation sans affecter la vitesse de dépurination de l'ADN (Collins *et al.*, 1999). On constate également que le seuil de racémisation associé à la conservation de l'ADN peut varier (Poinar *et al.*, 1996 ; Schmitz *et al.*, 2002 ; Serre *et al.*, 2004) et qu'il n'est pas toujours fiable. Dans l'étude sur les restes de Néandertaliens réalisée par Krings *et al.* (1997), le fragment osseux ayant permis l'amplification de l'ADN mitochondrial ne rentrait pas dans les critères de racémisation.

Il existe des méthodes alternatives permettant de déterminer l'état de préservation des macromolécules présentes dans un échantillon : la pyrolyse suivie de l'analyse par chromatographie en phase gazeuse ou par spectrométrie de masse (Poinar *et al.*, 1999), l'histologie (Hagelberg *et al.*, 1991 ; Colson *et al.*, 1997), les mesures de porosité et de densité osseuse (Nielsen-Marsh et Hedges, 2000), le degré de préservation du collagène ou de l'ostéocalcine (Collins *et al.*, 2000 ; Götherstörn *et al.*, 2002). Bien qu'un outil permettant d'estimer la préservation de l'ADN avant toute analyse soit nécessaire, il est cependant difficile de fixer des normes attendu le grand nombre de facteurs pouvant influer sur la conservation de l'ADN.

- Il est indispensable de **réaliser des témoins négatifs en parallèle de chaque extraction et amplification**. Il est parfois conseillé de réaliser plusieurs témoins négatifs pour une même PCR (Pääbo *et al.*, 2004), les contaminants pouvant être présents dans les réactifs en concentrations très faibles et donc être difficiles à détecter lors du pipetage de faibles volumes. Toutefois, ces témoins négatifs ne permettent pas de vérifier les contaminations liées à la présence de molécules (sucres ou ADN microbien) agissant comme transporteurs puisque ces substances ne sont présentes que dans les extraits d'ADN.

- **La quantification du nombre de molécules présentes dans l'extrait** permet de déterminer dans quelle mesure les résultats sont fiables. Il a été déterminé qu'un nombre de molécules inférieur à 1000 ne permettait pas d'obtenir des résultats valides (Handt *et al.*, 1996).

Il n'est cependant pas toujours possible d'obtenir de telles quantités d'ADN à partir d'échantillons anciens, en particulier concernant l'ADN nucléaire (cf V.1. Quantification de l'ADN nucléaire). Ce nombre de copies, correspondant à peu près à une concentration de 3 ng/ μ l, peut être retrouvé dans des échantillons extrêmement bien conservés mais cette valeur

est loin de représenter la norme. La validité des analyses sur des échantillons contenant un faible nombre de copies d'ADN (*Low Copy Number* : LCN) a été testée dans plusieurs études (Taberlet *et al.*, 1996 ; Gill *et al.*, 2000 ; Whitaker *et al.*, 2001 ; Kloosterman et Kesbergen, 2003). Pour l'analyse de STR (Short Tandem Repeats) nucléaires dans le cadre des identifications génétiques en Médecine Légale, l'utilisation d'échantillons contenant 100 pg/μl d'ADN (18 cellules diploïdes (Kloosterman et Kesbergen, 2003) permet l'obtention de profils complets à condition que le nombre de cycles d'amplification soit augmenté de 28 à 34 cycles (Gill *et al.*, 2000). Ce nombre de cycles de PCR permettrait de détecter une seule copie d'ADN (Gill *et al.*, 2000). Il faut néanmoins noter que ces conditions de PCR peuvent générer certains artefacts (perte d'hétérozygotie, déséquilibre de hauteur des pics, faux allèles et *stutters*) et augmentent la sensibilité à la contamination puisque des fragments de chromosomes ou des contaminants en très faible quantité pourront être amplifiés (Kloosterman et Kesbergen, 2003).

- **Pour chaque extrait d'un même sujet, plusieurs amplifications doivent être réalisées.** La multiplication des amplifications permet, tout d'abord, de mettre en évidence la contamination d'une extraction ou d'une amplification donnée. Dans le cas d'extraits contenant un très faible nombre de copies, la répétition des amplifications peut mettre en évidence si une absence d'amplification est liée au pipetage ou si l'extrait ne contient pas d'ADN utilisable (Pääbo *et al.*, 2004). Enfin, seule la répétition permet de détecter les éventuelles erreurs de réplication survenues au cours de l'amplification en raison de la qualité des acides nucléiques ou d'une erreur de la Taq polymérase.
- **Il doit exister une corrélation inverse entre l'efficacité de PCR et la longueur des amplicons.** Il est, en effet, peu probable d'arriver à amplifier des fragments de 500 à 1000pb et si tel est le cas une contamination peut être supposée. Cependant, certaines études ont montré que des fragments de grandes tailles pouvaient être amplifiés grâce à des conditions de préservations particulières (Lambert *et al.*, 2002).
- **L'ensemble des produits d'amplification doit être cloné et de multiples clones doivent être séquencés.** Cette précaution permet de détecter les hétérogénéités dans les produits de PCR pouvant provenir de contamination, de la dégradation des molécules d'ADN ou d'événements de *jumping PCR* (Pääbo *et al.*, 2004). Un logiciel permet d'estimer le nombre de clones qui doivent être séquencés afin de déterminer si la séquence consensus est valable.

Cette validation va dépendre de la fréquence de chaque nucléotide incorrect à une position donnée (Bower *et al.*, 2005). Douze semble être le nombre minimum de clones à séquencer afin de s'assurer de la validité des résultats. Il est ensuite possible de tester les séquences obtenues grâce à un logiciel permettant de vérifier la qualité des résultats provenant d'une PCR (Consensus Confidence Program, disponible à l'adresse suivante : <http://www.mcdonald.cam.ac.uk>).

Ce critère est toutefois discutable. En effet, les produits d'amplification correspondent bien souvent à un mélange entre les fragments correspondant à l'ADN de l'échantillon, à l'amplification de molécules dégradées et à de l'ADN contaminant. Dans le cas où seuls deux types de molécules vont être majoritaires il est possible d'invalider l'expérience car la contamination est patente. Cependant, comment exclure que l'espèce majoritaire lors du clonage et des amplifications précédentes ne correspond pas à une contamination ? Le clonage permet donc de mettre en évidence la présence de plusieurs acides nucléiques mais n'apporte pas de réelle solution.

- **La répétition des analyses par un laboratoire indépendant** peut mettre en évidence une contamination survenue au cours des manipulations et qui n'aurait pas été détectée dans les témoins négatifs. Ce contrôle n'a pas lieu d'être systématique mais se justifie dans le cas d'échantillons particulièrement intéressants ou si des résultats inattendus sont obtenus. Les échantillons devraient donc être expédiés directement depuis le site de fouille ou le musée afin d'éviter tout risque de transmission des contaminants entre les laboratoires (Pääbo *et al.*, 2004).

Dans le cadre des projets Egide de collaboration entre les pays de la communauté européenne, un dossier de partenariat avec le Laboratoire d'Identification Génétique de l'université du Pays Basque a été soumis. Ce projet s'intègre dans le programme PICASSO (collaboration franco-espagnole) et devrait permettre une duplication des analyses dans les meilleures conditions.

En complément des critères précédemment décrits, d'autres éléments peuvent être pris en considération pour confirmer l'authenticité des résultats :

- **La concordance entre la détermination morphologique et génétique du sexe du sujet** peut constituer une preuve de l'authenticité des résultats (Meyer *et al.*, 2000). Ce critère n'est pas toujours applicable puisque la diagnose sexuelle est dépendante des restes osseux

retrouvés ainsi que de l'âge du sujet. De plus, la probabilité que la personne ayant potentiellement contaminé l'échantillon soit du même sexe que le sujet étudié est de $\frac{1}{2}$. Il faut donc compléter, dans la mesure du possible, cette comparaison par l'analyse des STR autosomaux.

- Dans les études portant sur des restes humains, **l'analyse des restes animaux associés peut représenter une étape importante pour la validation des résultats**. L'utilisation d'amorces spécifiques d'un fragment d'ADN humain lors de l'amplification d'extraits provenant de restes animaux permet de mettre en évidence une contamination par effet *carrier* (Pääbo *et al.*, 2004). Dans le cas d'échantillons particulièrement précieux il est également possible de vérifier la présence d'ADN amplifiable dans les restes associés avant de tester les prélèvements d'intérêt afin d'éviter une dégradation inutile des fossiles rares.
- Les **résultats des études doivent avoir un sens du point de vue phylogénétique** (Cooper et Poinar, 2000). Des fragments d'ADN dérivant du génome des mitochondries sont parfois présents dans le génome nucléaire (Timmis *et al.*, 2004). L'ADN mitochondrial représentant fréquemment l'objet des études en paléogénétique l'amplification et l'analyse de ces insertions soumises à des pressions de mutation différentes peut mener à des erreurs d'interprétation concernant la position phylogénétique d'une espèce (van der Kuyl *et al.*, 1995) ou des variations intraspécifiques (Thalmann *et al.*, 2004). L'exemple le plus connu est celui d'une séquence supposée appartenir à un dinosaure et qui s'est révélée être une insertion nucléaire (numt) (Zichler *et al.*, 1995). La présence de numts a également été décrite chez le mammouth (Greenwood *et al.*, 1999), le moa (Cooper *et al.*, 2001) et le rhinocéros (Orlando *et al.*, 2003). L'utilisation de plusieurs couples d'amorces peut permettre de contourner ce problème puisqu'il est peu probable que deux paires d'amorces chevauchantes amplifient une même insertion nucléaire (Krings *et al.*, 1997).

Dans le cas des études portant sur des marqueurs haplotypiques dont les résultats peuvent être interprétés en considérant la répartition préférentielle de certaines mutations il est important de vérifier la vraisemblance des données obtenues. La découverte de données rattachant les échantillons à la population des personnes étant intervenues au cours de l'étude plutôt qu'à la population d'origine du sujet peut permettre d'identifier une contamination (Stone et Stoneking, 1998). Cette attitude, nécessaire afin d'éviter la validation de résultats erronés, pourrait néanmoins conduire à un biais. Comme nous l'avons vu précédemment, dans les

études sur les Néandertaliens toute séquence rappelant les séquences d'*Homo sapiens sapiens* est écartée.

La vérification phylogénétique des séquences par l'utilisation de programmes comme NETWORK (<http://www.fluxus-engineering.com/sharenet.htm>) permet la mise en évidence de certaines erreurs pouvant provenir de mutations fantômes, d'erreurs survenues lors du séquençage ou de la retranscription des résultats (Bandelt *et al.*, 2001). Ces analyses *a posteriori* permettent de révéler des paradoxes phylogénétiques ou des structures mosaïques qui donnent des indications précises sur la validité des séquences obtenues (Bandelt, 2005).

- **La reproductibilité des résultats obtenus à partir de différents substrats** peut être considérée comme une preuve de l'authenticité des résultats (Willerslev et Cooper, 2005). La comparaison des analyses d'un même sujet à partir d'échantillons osseux, dentaires et parfois pileux permet d'établir un consensus fiable puisqu'il apparaît peu probable que les différents substrats aient été contaminés par la même source (Amory *et al.*, 2006).

III APPROCHE ADOPTEE ET MARQUEURS ANALYSES

III.1 Méthodes de décontamination et de préparation des substrats

III.1.1 Prélèvement des échantillons archéologiques

Comme nous l'avons souligné dans le chapitre II.3.3, les précautions lors du prélèvement des échantillons sont une des étapes cruciales afin d'éviter les contaminations par de l'ADN exogène. Il a en effet été rapporté que ces contaminations antérieures aux processus d'analyse en laboratoire représentent une des sources majeures d'ADN exogène et sont particulièrement difficiles à mettre en évidence (Sampietro *et al.*, 2006).

Lors de ce travail de thèse, nous avons eu l'opportunité de travailler sur des échantillons dont la plupart ont été prélevés à l'occasion de campagnes de fouille effectuées dans une optique pluridisciplinaire incluant l'analyse génétique des sujets mis au jour. Lors de la description des corps, les membres de l'équipe portaient de manière systématique des gants et dans la mesure du possible des masques considérant les conditions de fouilles parfois difficiles. De plus, l'ensemble des chercheurs ayant participé aux fouilles a été typé afin de procéder à des comparaisons pour les différents marqueurs étudiés. Enfin, aucun lavage ou autre traitement n'a été effectué sur le terrain afin de minimiser les risques de contaminations (Gilbert *et al.*, 2005b, 2006c). Les échantillons osseux, dentaires et pileux étaient prélevés sur le sujet puis immédiatement placés dans des sachets plastiques afin d'éviter les contaminations croisées entre les différents prélèvements ou lors du transport de ceux-ci.

Ces trois dernières années, il a donc été possible de procéder à des prélèvements pour 61 sujets, dont 5 lors de fouille de sauvetage, qui ont été complétés par l'analyse de 9 sujets provenant de musées ou de fouilles antérieures (cf partie III.1.2. Les échantillons anciens). Pour les différents sujets étudiés, les prélèvements d'échantillons osseux ont été systématiques car ce substrat, qui est le plus fréquemment rencontré, permet de disposer de suffisamment de matériel pour renouveler les extractions afin de valider les résultats obtenus. Pour les sujets adultes, le choix des prélèvements c'est généralement porté sur les diaphyses d'os long qui représentent plus de la moitié de nos échantillons. Les os longs, et en particulier le fémur et le tibia, ont été privilégiés en raison de leur corticale dense et épaisse qui permet de générer une quantité importante de poudre d'os nécessaire à la répétition des extractions. Des dents ont été extraites lorsque l'état de conservation macroscopique satisfaisait les critères de sélection pour les études de paléogénétique.

Enfin, les conditions taphonomiques et environnementales permettant la pérennité des éléments pileux étant rares nous n'avons été en mesure d'effectuer des prélèvements de cheveux que pour 12 des sujets mis au jour.

Pour les sujets immatures, les prélèvements de plusieurs os longs ont été effectués afin de disposer d'une quantité de matériel suffisante à la duplication des résultats.

L'ensemble des prélèvements a été stocké dès le retour au laboratoire dans des sachets en plastique décontaminés sous UV, à une température de -20°C.

III.1.2 Les échantillons anciens

Le tableau 1 ci-dessous présente l'ensemble des échantillons collectés durant les années 2003 à 2005 par l'équipe de la MASFO. Nous avons été en mesure de prélever 72 échantillons provenant de tombes fouillées par nos soins ou d'échantillons fournis par les musées de Yakoutsk. Ainsi huit échantillons analysés proviennent d'échantillons collectés dans les musées (YAKa 25, 27, 28, 43, 44, 47, 48, 49). Lors des missions 2004 et 2005 des fouilles de sauvetage ont été effectuées pour 8 tombes (site de *Bekh Alaas*, site nommé *Bulldozer* et pour un site localisé à l'extérieur du village de Baïagah). La tombe de Ken Ebe 1 (YAKa 77) qui renfermait un nouveau-né n'a pas donné lieu à un prélèvement génétique car l'état de conservation des restes osseux ne le permettait pas.

Tous les échantillons prélevés ont fait l'objet d'une analyse génétique à l'exception des sujets du site de *Bulldozer* car les tombes ayant été éventrées et les ossements dispersés il n'était pas possible d'estimer dans quelle mesure il avaient pu faire l'objet de manipulations. De plus, aucune information archéologique n'était disponible pour ces trois sujets ce qui limitait grandement l'intérêt de l'analyse. Les restes du sujet de la tombe découverte par les archéologues yakoutes à l'extérieur du village de Baïagha n'ont pas été analysés pour les mêmes raisons.

III.1.2.1 Méthodes de fouilles

La fosse de creusement est repérée après un décapage superficiel de l'humus de surface. Un surface plus large que la fosse est ensuite creusée à la pelle afin de mieux révéler les structures mais également de faciliter les déplacements autour du coffre ou du cercueil.

N° LABO	NOM	SUBSTRAT disponible		
		OS	DENT	CHEVEUX
2003				
YAKa 15	ouhoraï	partie moyenne fémur		
YAKa 16	TD1	partie prox fémur		
YAKa 17	n°10	partie prox fémur		
YAKa 18	KM1	partie prox fémur		
YAKa 19	n°9	partie prox humérus		
YAKa 20	N°11	fémur G		
YAKa 21	TA1	partie prox fémur droit		
YAKa 22	12	tibia, fémur		
YAKa 23	n°5	ext. Prox tibia droit		
YAKa 24	n°6	partie prox fémur		
YAKa 25	n°4 (Pokrovsk)	partie prox fémur		
YAKa 26	BKH	partie moyenne fémur		
YAKa 27	squelette néolithique, n°3	phalange prox hallux droit		
YAKa 28	squelette néolithique, n°2	fragment diaphyse fémorale		
YAKa 29	STR1a	fémur, tibia		
YAKa 30	SRT1b	fémur, tibias et mandibule		
YAKa 31	AE2	rotule gelée		
YAKa 32	SRT2	humérus et rotule		
YAKa 33	KM2	2 rotules		
2004				
YAKa 34	Arbre 1 sujet 1	hemi-diaphyse d'humerus	2 PM, 8 I+C	X
YAKa 36	Arbre 1 sujet 2	rotule		
YAKa 37	Arbre 1 sujet 3	rotule	2 M très usées, 4 I	X
YAKa 38	Arbre 1 sujet 4	femur		
YAKa 35	Arbre 1 sujet 5	femur		
YAKa 39	Mounour Urekh 1	os du tarse + demi tibia	1 M	X
YAKa 40	Djoussou Len 1	metatarsiens + demi humerus	1M, 1PM	
YAKa 41	Arbre 2	hemi-diaphyse de femur	1 M mdb	
YAKa 42	Koulousoun Nakh 2	rotule + fragment humerus	3 M	
YAKa 43	Fedoseva Allalaika	fragment diaphyse femur		
YAKa 44	Fedosseva Tchersinsky	fragment diaphyse tibia		
YAKa 45	Arbre 3	fragment diaphyse femur		X
YAKa 46	Mounour Urekh 1bis	rotule + demi femur		X
YAKa 47	Musee ethno	fragment diaphyse femur		
YAKa 48	Fedoseva Diring Urekh	demi femur + maxillaire		
YAKa 49	Jardin Botanique	os du tarse		
YAKa 50	Jarama 1	1 femur + 2 rotules		
YAKa 51	Jarama 2	2 femurs + 2 tibias		
YAKa 52	Jarama 3	2 femurs + 2 tibias		
YAKa 53	Jarama 4	2 femurs + 2 tibias		
YAKa 54	Toumousaktakh 1	2 femurs + 2 tibias + 2 ulnas		
YAKa 55	Koulousoun Nakh 1	fragment femur	1 PM	X
YAKa 56	Dirilaa Sayliga 1	2 femurs + 1 mandibule		
YAKa 57	Bekh Alaas 1	demi femur		
YAKa 58	Bekh Alaas 2	demi femur		
YAKa 59	Bekh Alaas 3	demi humerus		
YAKa 60	Bekh Alaas 4	demi femur		

Tableau 1 : Liste des échantillons anciens prélevés. Lignes grises : échantillons écartés.

N° LABO	NOM	SUBSTRAT disponible		
		OS	DENT	CHEVEUX
2004				
YAKa 61	Bul 1	femur		
YAKa 62	Bul 2	humerus		
YAKa 63	Bul 3	demi femur		
2005				
YAKa 64	Sette Toumoul	1/2 fémur	2M	
YAKa 65	Oulakhans Alas 1	1/2 fémur	1M+2I	
YAKa 66	Batta Tcharana	3 os du tarse	3M	X
YAKa 67	Ken Ebe 3	humérus	1 M	X
YAKa 68	Ken Ebe 2	fémur+humérus		X
YAKa 69	Kous Tcharbyt	rotule	2M	X (peu, sales)
YAKa 70	Seden	rotule		X (très peu)
YAKa 71	Ken Ebe 4	1/2 tibia+2 rotules		
YAKa 72	Ken Ebe 5	fémur+humérus		
YAKa 73	Ken Ebe 6	fémur+tibia		
YAKa 74	Ken Ebe 7	fémur+humérus		
YAKa 75	Ken Ebe 8	1/2 fémur		
YAKa 76	Ken Ebe 9	fémur		
YAKa 77	Ken Ebe 1	pas de prélèvement		
YAKa 78	Bouogaryma 1	1/2 fémur	2M+1I	
YAKa 79	Bouogaryma 2	rotule	2M	X
YAKa 80	Nelegher	1/2 fémur	2M	
YAKa 81	Orto Aryy	humérus	2M+1C	
YAKa 82	Okhtobout 2	1/2 fémur	1I	
YAKa 83	Kyys Ounouoga		3I	
YAKa 84	Kyys Ounouoga	(talus trouvé en surface)		
YAKa 85	bord de route Bailigah	humérus+fémur		
YAKa 86	Tyyt Bappyt	1/2 fémur	3 M (petites)	
TOTAL		70	19	12

Tableau 1 (suite).

Une fois le coffre repéré, la surface ainsi que les côtés sont dégagés à la truelle et au pinceau pour permettre les mesures ainsi que les photos. Après l'ouverture du coffre, deux cas de figure se présentent. Soit le sujet est totalement squelettisé, la tombe est alors fouillée de manière classique (outils de dentiste, pinceaux, etc...) puis une couverture photographique est réalisée ; soit la tombe contient un sujet dont les tissus mous sont encore conservés auquel cas les photographies sont rapidement effectuées avant l'apparition de moisissures.

La description du mobilier et des vêtements bénéficie grandement des connaissances des archéologues yakoutes (Edouard Jyrkov, Vassili Popov et Sergeï Kolodeznikov) qui sont capables de reconnaître la plupart des fourrures et des peaux ainsi que le mode de traitement utilisé. Une fois la description du mobilier, des vêtements et du corps effectuée, les prélèvements parasitologiques et génétiques sont réalisés.

L'ensemble du processus de description et prélèvement peut s'étaler sur une heure pour une tombe simple à plus de sept heures pour les tombes les plus riches.

III.1.2.2 Répartition géographique

Les tombes que nous avons étudiées proviennent principalement de 3 *ulus* de Yakoutie Centrale : Khangalassky (Khangalass), Churapchinsky (Churapcha) et Tattinsky (Tattaa) (figure 19). Deux sujets ont cependant été mis au jour dans la région de la Kolyma et deux échantillons osseux prélevés dans des musées proviennent de sujets sujets découverts dans la région de la Vilyuy. La localisation de chaque échantillon est détaillée dans le tableau 2.

Les prélèvements génétiques réalisés sur la population actuelle ont été réalisés dans les trois *ulus* au sein desquels les sujets anciens ont été mis au jour.

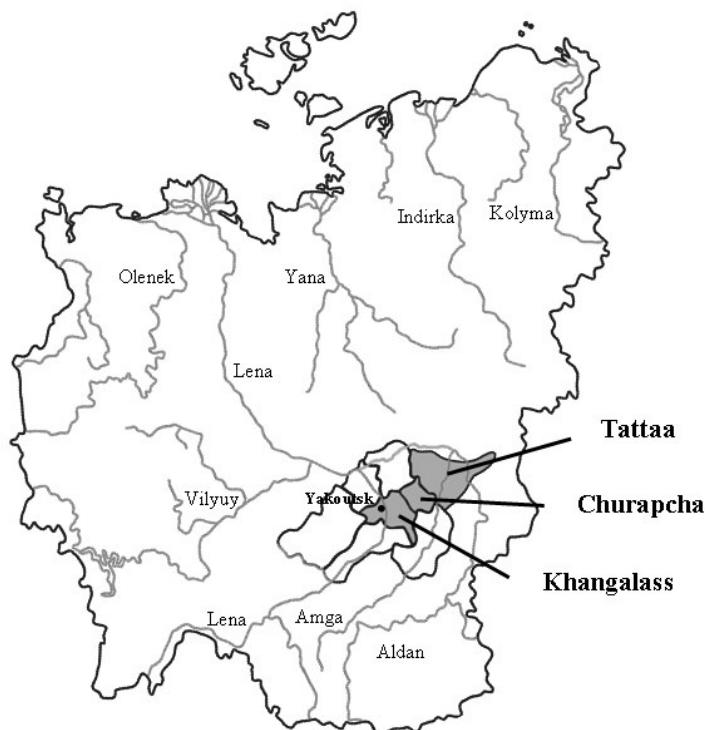


Figure 19 : Localisation des trois *ulus* de Yakoutie Centrale au sein desquelles les fouilles ont été réalisées.

La répartition des sujets anciens analysés découverts en Yakoutie Centrale est la suivante : 25 sujets proviennent de l'*ulus* de Churapcha, 20 sujets ont été mis au jour dans l'*ulus* de Khangalass et 18 sujets étaient inhumés dans l'*ulus* de Tattaa.

n° Labo	Nom	Localisation	n° Labo	Nom	Localisation	n° Labo	Nom	Localisation
YAKa34	Arbre 1 sujet 1		YAKa15	ouhoraï		YAKa64	Sette Toumoul	
YAKa35	Arbre 1 sujet 5		YAKa16	TD1		YAKa65	Oulakhans Alas 1	
YAKa36	Arbre 1 sujet 2		YAKa17	n°10		YAKa66	Batta Tcharana	
YAKa37	Arbre 1 sujet 3		YAKa18	KM1		YAKa67	Ken Ebe 3	
YAKa38	Arbre 1 sujet 4		YAKa19	n°9		YAKa68	Ken Ebe 2	
YAKa39	Mounour Urekh 1		YAKa20	N°11		YAKa69	Kous Tcharbyt	
YAKa40	Djoussou Len 1		YAKa21	TA1		YAKa70	Seden	
YAKa41	Arbre 2		YAKa22	12		YAKa71	Ken Ebe 4	
YAKa42	Koulousoun Nakh 2		YAKa23	n°5		YAKa72	Ken Ebe 5	
YAKa45	Arbre 3		YAKa24	n°6		YAKa73	Ken Ebe 6	
YAKa46	Mounour Urekh 1bis		YAKa25	n°4 (Pokrovsk)	Khangalass	YAKa74	Ken Ebe 7	Tattaa
YAKa50	Jarama 1		YAKa26	BKH		YAKa75	Ken Ebe 8	
YAKa51	Jarama 2	Churapcha	YAKa29	STR1a		YAKa76	Ken Ebe 9	
YAKa52	Jarama 3		YAKa30	SRT1b		YAKa77	Ken Ebe 1	
YAKa53	Jarama 4		YAKa31	AE2		YAKa78	Bouogaryma 1	
YAKa54	Toumousaktakh 1		YAKa32	SRT2		YAKa79	Bouogaryma 2	
YAKa55	Koulousoun Nakh 1		YAKa33	KM2		YAKa80	Nelegher	
YAKa56	Dirilaa Sayliga 1		YAKa47	Musee ethno		YAKa81	Orto Aryy	
YAKa57	Bekh Alaas 1		YAKa48	Fedoseva Diring Urekh				
YAKa58	Bekh Alaas 2		YAKa49	Jardin Botanique				
YAKa59	Bekh Alaas 3		YAKa27	3, squel néolithique	Kolyma			
YAKa60	Bekh Alaas 4		YAKa28	squel néolithique, n°2	Kolyma			
YAKa82	Okhtobout 2		YAKa43	Fedoseva Allalaika	Vilyuisky			
YAKa83	Kyys Ounouoga		YAKa44	Fedosseva Tchersinsky	Vilyuisky			
YAKa86	Tyyt Bappyt							

Tableau 2 : Répartition géographique des sujets anciens étudiés suivant la région de leur mise au jour.

III.1.2.3 Datations

Considérant le nombre très important de tombes, il n'a pas été possible de réaliser des datations absolues au carbone 14 sur l'ensemble d'entre elles. De plus, la méthode AMS (Accelerator Mass Spectrometry) utilisée nécessite des quantités relativement importantes de matériel osseux, environ 30gr (<http://www.radiocarbon.com/sending.htm>) qui pour certains échantillons n'étaient pas disponibles. Cependant, les sujets présentant un intérêt particulier ou des particularités dans les modes d'inhumation et pour lesquels suffisamment d'os était disponible ont été datés par cette méthode. Nous disposons ainsi de datations absolues pour 4 sujets (cf Annexe 1).

Des prélèvements ont été réalisés sur les troncs constituant les cercueils ce qui permettra d'obtenir une datation dendrochronologique et par conséquent d'obtenir des datations absolues pour un nombre plus important de sujets.

Pour l'ensemble des autres tombes, il a été possible de proposer des datations grâce à la comparaison des modes de construction, du mobilier associé au corps ou des vêtements des sujets entre les tombes pour lesquelles nous disposons de datations radiocarbone et les sépultures présentant les mêmes particularités. Du point de vue archéologique, les recherches menées par les membres de la MAFSO en collaboration avec les chercheurs Yakoutes ont permis de réaliser certains recoupements entre le mobilier retrouvé dans les tombes et les données de la littérature. Le classement des différentes tombes dans les périodes détaillées ci-dessous est présenté en Annexe 1.

Tout d'abord, nous avons regroupé sept tombes, dites anciennes, c'est à dire qui précèdent l'émergence de la culture yakoute traditionnelle. La tombe de Pokrovsk datée de 2140 à 2340 ans BP (Beta – 198197) a été placée dans cet ensemble. Il s'agit de la seule tombe pour laquelle nous disposons d'une datation précise. Les autres échantillons correspondent à des sujets anciens retrouvés par les archéologues yakoutes mais pour lesquels nous ne disposons pas d'informations archéologiques si ce n'est des datations très larges. Les sujets des sites d'Allalaïka et Tchersinsky (peut être Tchirkouo) auraient été mis au jour dans la Vilyuy et seraient datés du Néolithique Moyen (6000/5000 ans) ; le sujet provenant du site de Diring Urekh, dans l'*ulus* Khangalassky, pourrait être âgé de 3800 ans. Le sujet échantillonné dans le musée d'ethnologie de Yakoutsk (YAKa47) pourrait dater de l'Age du Bronze. Cependant il est malheureusement impossible de vérifier ces datations.

Pour les tombes de la période médiévale, du XV^{ième} au XVIII^{ième} siècle environ, il est possible de distinguer clairement deux périodes :

La première période regroupe les neuf tombes les plus anciennes pour l'époque médiévale en Yakoutie comprises entre le XV^{ième} et le XVII^{ième} siècle. La tombe de Djoussoulen, caractéristique de cette période a été datée au C14 entre 1440 et 1660 AD (Beta-198198). Les tombes de cette période sont caractérisées par la présence d'une selle parfois placée sous la tête du sujet, d'étriers et assez fréquemment d'un arc (cf Figure 20). En revanche les éléments d'importation telles que les perles sont absents des tombes de cette période. Le mode de construction des cercueils ou des coffres correspond généralement à plusieurs rondins imbriqués.



Figure 20 : Tombe de *Oulakhan Alas*, correspondant à la première période médiévale. On note la présence de la selle sous la tête, d'étrier en bois, d'un arc déposé à l'extérieur du coffre ainsi que d'un pot en bois traditionnel (*tchoron*).

Dans le second ensemble de la période médiévale composé de 20 tombes, le mobilier présente, en opposition à celui du premier groupe, des éléments importés sans doute de Chine comme des perles en verre ou des vêtements en soie (cf Figure 21). Aucun des sujets de cette période n'a été inhumé avec une selle sous la tête et le mode de construction des coffres diffère puisqu'il s'agit fréquemment de troncs évidés pour les tombes riches ou d'hémi-troncs assemblés pour former les parois supérieures, inférieures et latérales. Ces tombes peuvent être datées entre le XVII^{ème} et la première partie du XVIII^{ème} siècle.



Figure 21 : Tombe de *Kyys Ounouoga* dont le sujet féminin porte un manteau décoré par les perles caractéristiques des tombes de cette période. Deux pots en bois ont également été déposés, dont un contenant des produits laitiers ainsi que des côtes de chevaux.

Une période intermédiaire aux ensembles précédents a été distinguée. Elle ne comporte que deux tombes provenant de deux sites distincts : Nelegher et Kous Tcharbyt. Le mobilier de ces tombes est caractérisé par l'absence à la fois de selle ou d'étrier et de mobilier d'importation. Par ailleurs l'armement déposé avec le corps rappelle celui du premier ensemble mais le mode de construction évoque les troncs évidés de la deuxième période.

Enfin, il est possible de faire une distinction nette entre les tombes caractéristiques de la période médiévale (du XV^{ème} au XVIII^{ème} siècle) et les 29 tombes plus tardives où la christianisation des pratiques funéraires est clairement visible par la présence d'une croix, d'un cierge ou d'une couronne en papier correspondant à la couronne céleste orthodoxe (cf figure 22). Des croix gravées sur le coffre des cercueils sont bien souvent présentes et de manière générale le mobilier associé au corps est beaucoup moins important que dans les tombes de la période médiévale.



Figure 22 : Tombe de *Seden*, caractéristique de la période post-colonisation par la présence d'une croix ainsi que d'un cierge.

III.1.3 Les substrats osseux

Parmi les différentes méthodes de décontamination décrites précédemment, celle qui a été adoptée dans notre étude consiste à éliminer mécaniquement une épaisseur d'environ 2mm de la surface de l'os. Les échantillons osseux sont, tout d'abord, débarrassés de toutes traces de terres ou de fragments de tissu à l'aide d'une lame de scalpel stérile. L'étape d'abrasion a été réalisée à l'aide d'une perceuse de type Dremel qui autorise une liberté de mouvements importante et le travail sous une hotte dont l'usage est exclusivement dédié à cette étape. Il est également possible de décontaminer l'ensemble du matériel servant à l'abrasion (outil Dremel, fraises, pince, etc.) par les méthodes classiques et une étape d'irradiation dans un four à UV crosslinker Biolink 254nm (Fisher Bioblock). L'ensemble des échantillons osseux a été soumis à ce traitement, les prélèvements destinés à la méthode de trépanation (diaphyses d'os longs) comme ceux ayant nécessité une étape de cryobroyage (rotules, os du carpe/tarse, fragments de diaphyses ou os de sujets immatures). Cette méthode a été préférée aux techniques utilisant des lavages à l'hypochlorite de sodium ou une décalcification à l'EDTA en raison de la qualité des échantillons et des précautions prises lors des fouilles. En effet, la conservation, au niveau macroscopique, des ossements mis au jour fait état d'une faible porosité. Il ne nous a donc pas semblé nécessaire de procéder à ce type de décontamination chimique. Ce choix a également été motivé par le fait que cette méthode s'est révélée efficace dans plusieurs études (Stoneking *et al.*, 1998 ; Fily *et al.*, 1998 ; Bouwman *et al.*, 2006) et notamment les études menées au sein du laboratoire d'anthropologie moléculaire (Ricaut *et al.* 2005, 2006 ; Keyser-Tracqui *et al.*, 2003).

III.1.4 Les dents

Comme nous l'avons décrit dans le chapitre II.2.2, la structure de la dent accorde une certaine protection à l'ADN endogène rendant ce substrat particulièrement intéressant pour les études de paléogénétique. La sélection des dents utilisées lors de notre étude a suivi des critères stricts toujours dans l'optique de s'affranchir des contaminations. Ainsi les dents présentant une usure apicale trop importante, des caries, des fissures ou tout autre trace de traumatisme n'ont pas été utilisées pour notre étude. Après sélection nous avons pu extraire de l'ADN à partir de dents pour 19 des 67 sujets étudiés. Il existe une grande variété de protocoles de

décontamination des dents : depuis la simple immersion dans l'hypochlorite de sodium immédiatement suivie par le broyage (Richards, Sykes and Hedges, 1995 ; Sampietro 2006), le nettoyage de la surface par différents réactifs (peroxyde d'hydrogène, DNA Away, éthanol absolu, etc), l'irradiation UV est également employée ou encore une combinaison de ces différentes techniques (Malmström *et al.*, 2005).

Afin de s'affranchir de tout risque de contamination nous avons choisi d'employer une combinaison de ces différentes méthodes. Le protocole de décontamination suivi consiste en plusieurs lavages, tout d'abord, en utilisant de l'eau milliQ décontaminée afin de retirer toutes traces de terre ou restes de tissus de la surface de la dent. Les dents sont ensuite brièvement nettoyées avec un décontaminant de surface (DNA Away™, Molecular Bioproducts) puis rapidement rincées dans de l'eau milliQ décontaminée. Enfin un étape de rinçage et de séchage dans de l'éthanol absolu est pratiquée avant d'exposer les dents aux UV durant 10 min sur les différentes faces. Toutes les dents sont, par la suite, entièrement cryobroyées dans l'azote liquide. Cette technique, bien qu'entraînant la destruction complète de l'échantillon, a été préférée aux méthodes de trépanation visant à récupérer la pulpe dentaire (Gilbert *et al.*, 2003b). La dentine, en tant que tissu fortement minéralisé, contient des cristaux d'hydroxyapatite tout comme dans le tissu osseux. Or, il semble que ces cristaux favorisent l'adsorption et la protection de l'ADN au cours du temps (Salamon *et al.*, 2005). Il serait par conséquent possible d'obtenir une quantité plus grande d'ADN à partir de la méthode de cryobroyage (Smith *et al.*, 1993). De plus, la nécessité de répéter les extractions pour la validation des résultats implique une quantité de matériel relativement importante qu'il n'est possible d'obtenir que par la méthode de cryobroyage. Cette technique diminue également les risques de contaminations en réduisant le nombre des manipulations nécessaires à la découpe de la dent et permet enfin d'utiliser les incisives et les canines qui sont difficilement étudiables par les autres méthodes.

III.1.5 Les cheveux

L'état de conservation exceptionnel de certains des sujets mis au jour durant ces trois années nous a amené à envisager l'extraction d'ADN à partir de cheveux pour 12 individus. En effet, ce substrat avait été utilisé de longue date en médecine légale et plus récemment dans les recherches paléogénétiques. Cependant aucune étude n'avait mis en évidence de résultats sur

de l'ADN nucléaire extrait à partir des tiges de cheveux. Considérant l'état de préservation de l'ADN des sujets inhumés dans des tombes proches du permafrost et les résultats d'études récentes sur l'ADN mitochondrial extrait de cheveux anciens (Gilbert *et al.*, 2004 et 2005 ; McNevin *et al.*, 2005a et b), il semblait pertinent d'adapter et d'optimiser ces protocoles pour nos échantillons. La publication n°1 (Annexe 3) présente en détail les différentes étapes du protocole expérimental que nous avons suivi aussi nous ne le détaillerons pas ici.

III.1.6 Trépanation et cryobroyage des tissus durs

La méthode de trépanation a été utilisée de manière préférentielle pour toutes les diaphyses d'os long. Cette méthode emploie une perceuse à colonne utilisée à vitesse de rotation réduite afin de limiter tout échauffement qui pourrait être néfaste pour l'ADN. Des trépans de différents diamètres ont été utilisés suivant la taille de la diaphyse dont nous disposions. La trépanation était réalisée dans une zone ne s'approchant pas de plus d'un centimètre de la zone non abrasée afin d'éviter que la poudre collectée ne soit en contact avec la partie non décontaminée de l'os (cf Figure 23).



Figure 23 : Hémi-diaphyse de fémur abrasée et trépanée.

La méthode de broyage utilisant un cryobroyeur (Freezer Mill 6800, Spex) est utilisée pour réduire en poudre les fragments osseux de petite taille qui ne permettent pas d'employer la méthode de trépanation. L'azote liquide permet d'éviter les augmentations de température qui pourrait entraîner une dégradation de l'ADN. Les échantillons ainsi qu'un aimant cylindrique sont placés dans des cylindres en polycarbonate. L'ensemble est ensuite immergé dans l'azote

liquide et soumis à un champ magnétique qui va déplacer l'aimant et ainsi réduire le fragment osseux en poudre.

Afin d'éviter les chocs thermiques liés aux congélations/décongélations, les poudres d'os ainsi que les cheveux sont stockées à température ambiante durant la phase pendant laquelle les extractions sont réalisées. Une fois l'ensemble des marqueurs analysés les poudres sont transférées dans un congélateur, à -20°C, dédié à la conservation. En effet, nous avons constaté qu'un stockage trop prolongé à température ambiante pouvait être néfaste et entraîner une réduction de la quantité et de la qualité de l'ADN extrait à partir d'une même poudre.

III.2 Extraction et purification de l'ADN

III.2.1 L'extraction

L'étape d'extraction est primordiale en anthropologie moléculaire car elle doit permettre de purifier des quantités très faible d'ADN à partir des substrats anciens tout en limitant la présence d'inhibiteurs et en minimisant les risques de contaminations. Quelque soit le type de substrat de départ la méthode d'extraction d'ADN comprend les même étapes : une lyse cellulaire suivie de la dégradation des protéines et enfin la séparation des acides nucléiques des autres substances libérées.

Les méthodes d'extraction phénoliques représentent à l'heure actuelle un consensus dans les études récentes (Gilbert *et al.*, 2005 ; Imaizumi *et al.*, 2005 ; Orlando *et al.*, 2006 ; Sampietro *et al.* 2006 ; etc.) seules les compositions des tampons de lyse et les méthodes de purification diffèrent. La technique utilisée dans cette étude a été optimisée au sein du laboratoire d'Anthropologie Moléculaire de Strasbourg et validée par plusieurs publications (Keyser-Tracqui *et al.*, 2003 ; Keyser et Ludes, 2004 ; Ricaut *et al.*, 2004, 2005, 2006).

III.2.2 Lyse et décalcification

L'étape de décalcification/lyse est réalisée dans un tampon contenant 0,5 mM d'EDTA (décalcification des tissus osseux par chélation des ions Ca^{2+}), du Sodium Dodécyl Sulfate 2% (solubilisation des lipides), 0,01 M Tris HCl (stabilisation du pH), 0,3 M d'acétate de Sodium (stabilisation de l'ADN sous forme double brin) et 1 mg/ml de protéinase K (lyse des protéines). Deux grammes de poudre, obtenue à partir d'os ou de dents, sont placés dans 5 à 10 ml de la solution de lyse suivant la taille des copeaux qui peut varier en fonction de la méthode de pulvérisation utilisée et de la qualité des os. Les tubes contenant la poudre dans le tampon de lyse sont ensuite incubés sous agitation durant une nuit, à 55°C.

Utilisation de Bromure de N-Phényl Thiazolium (PTB)

Comme nous l'avons mentionné dans le chapitre II.3.1.1.4, le PTB permet de dissocier les complexes sucre-protéine pouvant se former au cours du temps. Nous avons mis en place un protocole de test durant l'année 2006 afin d'évaluer dans quelle mesure nous obtenions une

amélioration de la qualité et de la quantité d'ADN extrait de nos échantillons anciens en présence de PTB. Nous avons réalisé des extractions parallèles pour 13 échantillons en ajoutant au tampon de lyse initial 0,001M de PTB comme recommandé dans l'étude de Poinar et collaborateurs (1998).

III.2.3 Extraction organique

Après cette étape de lyse survient l'étape d'extraction organique de l'ADN qui permet de séparer l'ADN des restes cellulaires, des débris membranaires lipidiques mais également des résidus protéiques et autres composés phénoliques (O'Rourke *et al.*, 2000). Cette étape est réalisée en ajoutant un volume de phénol chloroforme isoamyl alcool (25/24/1, v/v) équivalent au volume de tampon de lyse utilisé. Après centrifugation (15min à 4000rpm), la phase aqueuse contenant les acides nucléiques est prélevée. Nous avons constaté que pour certains échantillons il était préférable de renouveler cette étape une seconde fois. En effet, la présence d'une coloration brune dans le surnageant a été décrit comme caractéristique de la présence d'inhibiteurs (Cooper, 1992). Le fait de réitérer cette étape permet de diminuer la coloration du surnageant et par conséquent de diminuer la concentration en inhibiteurs.

III.2.4 Purification et concentration

L'étape de purification vise à séparer l'ADN des inhibiteurs qui auraient pu être co-purifiés lors de l'extraction organique. Plusieurs méthodes sont disponibles (Sephadex, colonne à membrane anisotrope, colonne à filtre de silice).

La méthode de purification choisie dans notre étude utilise la forte affinité de l'ADN pour la silice (Höss et Pääbo, 1993) qui permet une meilleure purification vis à vis des inhibiteurs que l'utilisation de filtre de type Centricon (Yang *et al.*, 1998). Pour cette étape nous avons utilisé le kit CleanMix (Talent) qui permet la purification de fragments d'ADN de grande taille, entre 100pb et 50kb, à partir du surnageant de l'extraction organique. La purification comprend trois phases : (i) La première étape consiste à fixer l'ADN sur une résine à base de silice. La fixation de l'ADN à la silice est facilitée et accélérée par l'utilisation d'un agent chaotropique qui va également permettre d'éviter que des protéines résiduelles ne se fixent sur

la résine entraînant la co-purification d'inhibiteurs éventuels. Un volume de solution de liaison (Binding solution) correspondant à la moitié du surnageant récupéré est ajouté au surnageant puis l'ensemble du volume est transféré sur des colonnes munies d'un filtre qui va retenir le complexe ADN-résine. (ii) Après une centrifugation rapide permettant de récupérer la résine sur le filtre de la colonne, l'ADN adsorbé à la résine est rincé deux fois à l'aide d'une solution de lavage (Washing solution). (iii) Puis l'ADN est éluté dans 400µl d'eau ultrapure stérile (Biosolve Ltd) à 72°C permettant ainsi la re-suspension de l'ADN. Il faut toutefois être vigilant lors de cette étape car la silice est un inhibiteur de PCR, il faut par conséquent s'assurer que l'extrait purifié sera exempt de toute trace de silice (Yang *et al.*, 1997).

L'étape de concentration va permettre d'éluer l'extrait d'ADN dans le volume le mieux adapté qui correspond à un compromis entre le nombre d'amplifications des différents marqueurs étudiés et la concentration minimale permettant de réaliser ces amplifications. Les 400µl élus lors de l'étape précédente sont transférés sur la membrane du filtre. Les Microcons sont par la suite centrifugés à 9000rpm pendant 9 à 11 minutes afin de réduire le volume. Ainsi pour les extraits obtenus à partir d'ossements de sujets immatures ou de cheveux le volume final sera réduit à 30/20µl final alors que pour les extraits provenant de diaphyse d'os long ou de dents le volume d'élution est de 40µl.

Cette étape permet également d'éliminer les fragments d'ADN de petite taille, le seuil d'arrêt de la membrane des Microcons YM30 (Amicon) étant de 50pb pour de l'ADN double brin. L'élimination des fragments de petite taille, présents dans les extraits d'ADN dégradés, permet de réduire les risques d'inhibition de PCR (Pusch *et al.*, 1998 ; Pääbo *et al.*, 2004).

Après cette étape les échantillons sont stockés à -20°C et l'ADN est utilisable pour les amplifications des différents marqueurs moléculaires.

III.3 Les marqueurs moléculaires

III.3.1 Quantification de l'ADN par PCR en temps réel

III.3.1.1 Intérêt de la quantification

La quantité d'ADN présent dans chaque extrait peut varier de manière importante suivant les échantillons et les substrats utilisés mais également d'une extraction à une autre à partir d'un même échantillon. Par conséquent, comme étape préalable à toute amplification nous avons réalisé la quantification de l'ADN présent dans nos extraits à l'aide de l'ABI Prism® 7000 *Sequence Detection System* (Applied Biosystems). L'utilisation du kit Quantifiler™ *Human Quantification kit* (Applied Biosystems) apparaît comme indispensable car la quantité de matériel disponible pour l'extraction est parfois limitante en terme de nombre d'extractions (e.g. fragments d'os de petite taille sans possibilité de renouveler l'échantillon) et les extraits d'ADN sont toujours précieux. La quantification permet de déterminer le volume exact d'extrait à ajouter dans une réaction en fonction des recommandations du fabricant puisque la quantité d'ADN varie suivant les kits et les marqueurs utilisés.

La quantification par PCR en temps réel permet, outre l'estimation de la concentration de l'ADN présent dans l'extrait, d'évaluer la présence d'inhibiteurs. Il est par conséquent possible de réaliser une gamme de dilution à partir des extraits contenant des inhibiteurs tout en conservant une concentration en ADN suffisante pour réaliser l'amplification des fragments d'intérêt.

III.3.1.2 Principe de la quantification

Le principe du kit Quantifiler™ est basé sur l'utilisation de sondes TaqMan (Figure 24) dont la cible correspond à un fragment intronique de 62pb du gène de la télomérase humaine (TERT : *Telomerase Reverse Transcriptase*).

Ces sondes sont marquées à leurs extrémités 5' avec un fluorophore émetteur et à leur extrémité 3' par un fluorophore récepteur. Du fait de sa proximité, le récepteur absorbe la fluorescence de l'émetteur. Par conséquent lorsque la sonde est intacte aucune fluorescence

n'est détectée. Lors de la phase d'elongation de la réaction de PCR, la polymérase grâce à son activité 5' nucléase va cliver la sonde si celle-ci se trouve en aval de la zone d'hybridation de l'amorce sens. Le fluorophore émetteur n'étant plus à proximité du fluorophore récepteur, il sera possible de visualiser un signal fluorescent.

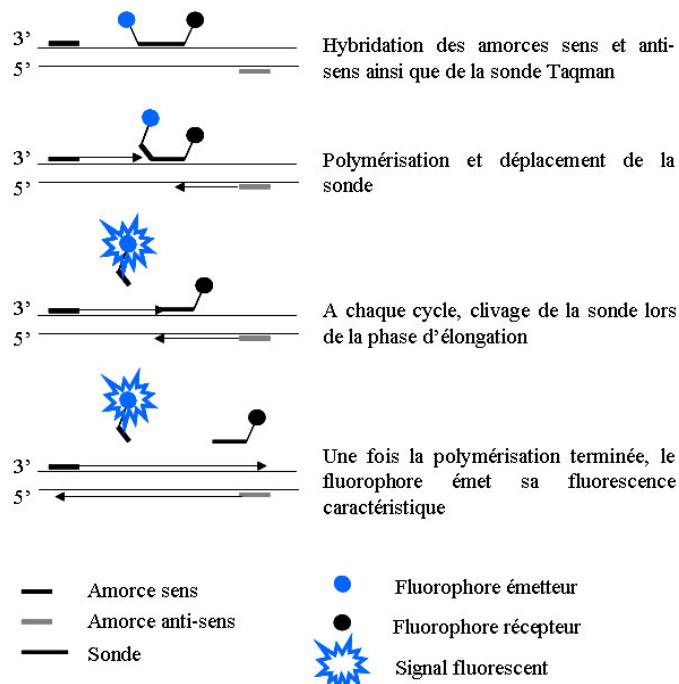


Figure 24 : Principe de fonctionnement des sondes Taqman.

Le nombre de cycles de PCR permettant d'atteindre le seuil de détection de fluorescence est défini comme le cycle seuil (C_t : Cycle threshold). Le C_t se situe durant la phase exponentielle de la PCR durant laquelle il existe une corrélation directe entre le nombre de cycles et le nombre de molécules synthétisées. L'augmentation du signal fluorescent durant cette phase sera proportionnelle à la quantité d'amplissons générée. Il est, par conséquent, possible de déterminer le nombre de copies de départ de la cible en fonction du nombre de cycles nécessaires pour atteindre ce seuil. Un extrait d'ADN de concentration connue sert à réaliser une gamme de dilution qui va permettre d'obtenir une courbe standard de laquelle seront déduite les concentrations d'ADN présentes dans les échantillons. Le kit Quantifiler™ nécessite 2 μ l d'extrait d'ADN ajouté à un mix réactionnel comprenant 12,5 μ l de mix PCR (contenant la polymérase, les dNTP et le tampon d'amplification) et 10,5 μ l de mix contenant l'*Internal PCR Control* (IPC) ainsi que les amorces et les sondes pour la TERT et l'IPC. La gamme de détection est comprise entre 23pg et 50ng en utilisant 2 μ l d'extrait d'ADN. Il est

important de noter que dans la gamme de concentration inférieure à $100\text{pg}.\mu\text{l}^{-1}$, l'exactitude des mesures réalisées avec le kit Quantifiler est réduite du fait des variations qui peuvent être importante lors du pipetage et de la sensibilité intrinsèque de la méthode.

L'IPC inclus dans le kit QuantifilerTM est utilisé pour la détermination de la présence d'inhibiteur dans l'extrait d'ADN des échantillons. La cinétique d'amplification de l'IPC est connue dans un milieu exempt d'inhibiteurs. Lors de l'amplification simultanée avec l'extrait d'ADN de l'échantillon, les délais observés dans le Ct de l'IPC sont par conséquent imputables à une inhibition de l'amplification. Plus le délai observé sera important, plus la concentration en inhibiteurs sera importante sachant que l'inhibition peut aller jusqu'à l'absence complète d'amplification.

III.3.2 Les marqueurs à transmission biparentale

III.3.2.1 Détermination du sexe des individus : le gène de l'amélogénine

La détermination génétique du sexe des individus complète la diagnose morphologique effectuée par les méthodes classiques pour les sujets matures et permet d'accéder à cette information pour les individus immatures.

La méthode employée dans notre étude est la plus communément utilisée à la fois dans le domaine de la paléogénétique mais également dans delui de l'identification des individus en Médecine Légale. Cette méthode est basée sur l'amplification d'une région du premier intron du gène de l'amélogénine qui comporte une délétion de 6pb sur le chromosome X. Cette technique, décrite par Sullivan *et al.* en 1993, utilisait à l'origine un couple d'amorces entraînant l'amplification de fragments de 212 et 218pb respectivement pour les chromosomes X et Y. Les kits d'amplification utilisés dans notre étude comprennent deux paires d'amorces plus proches de la cible permettant l'amplification de fragments de 106pb pour le chromosome X et 112pb pour le chromosome Y.

Il existe de rares cas qui peuvent conduire à une interprétation erronée des données. Des délétions du gène de l'amélogénine peuvent survenir sur le chromosome Y (Santos *et al.*, 1998) ou X (Shewale *et al.*, 2000) cependant ces anomalies sont extrêmement rares. Toutefois, dans certaines populations, notamment indienne (Cadenas *et al.*, 2006), il semble que les délétions sur le chromosome soient plus fréquentes.

III.3.2.2 Les STR autosomaux

III.3.2.2.1 Généralités et propriétés

Le génome humain est composé de séquences codantes communément appelées gènes mais il est principalement formé par des parties dites non codantes qui ne sont pas directement impliquées dans la synthèse de protéines. Environ 30% de l'ADN non codant est composé par des séquences répétées dont la longueur et le nombre de répétition des motifs est variable. Les régions comportant des motifs répétés allant de 2 à 6pb sont appelées microsatellites ou Short Tandem Repeats (STR) et représenteraient 3% du génome total (d'après l'International Human Genome Consortium, 2001). Les STR sont présents sur l'ensemble des autosomes et également sur les chromosomes X et Y et sont répartis en moyenne toutes les 10kb. La position chromosomique permet de définir un locus, ou marqueur, pour lequel le nombre de répétition définit les allèles. Le nombre de répétition de ces marqueurs est très variable suivant les individus ce qui en fait des marqueurs de choix pour l'identification des personnes. Pour l'ensemble des chromosomes, excepté le chromosome Y, l'individu reçoit un allèle de son père et un de sa mère suivant le modèle mendélien ce qui permet d'envisager l'analyse des relations de parenté existant entre les individus.

Le taux de mutation des STR autosomaux est faible et varie suivant les loci de 0 à 7×10^{-3} . Il apparaît également des différences entre les sexes sachant que le taux de mutation est plus important (environ 5 fois plus élevé) chez les hommes que chez les femmes ; l'âge est également un facteur qui peut entraîner une augmentation du taux de mutation (Brinkmann *et al.*, 1998).

III.3.2.2.2 Artéfacts d'amplification

Le typage de STR à partir d'échantillons anciens peut présenter des difficultés du fait des altérations subies par les molécules d'ADN ainsi que des concentrations faibles ce qui a tendance à favoriser l'apparition d'artéfacts d'amplification.

L'absence d'amplification est le problème auquel nous avons été le plus fréquemment confronté (cf chapitre IV.3.2 Efficacité de typage). La taille des marqueurs pouvant être relativement importante, jusqu'à 400pb, il n'est pas rare que l'ADN présent dans le tissu ancien soit trop fragmenté pour permettre l'amplification des fragments de grande taille.

Outre la fragmentation du brin d'ADN, les variations stochastiques liées à l'amplification d'ADN en faible quantité vont également être responsables d'absence d'amplification (Whitaker *et al.*, 2001).

La perte d'hétérozygotie (ou *allelic dropout*) est également un des artefacts fréquemment rencontré lors de l'analyse de STR sur de l'ADN dégradé et/ou en faible nombre de copies. Cette absence d'amplification pour un seul des deux allèles affecte généralement l'allèle de plus haut poids moléculaire. Cependant dans le cas de l'utilisation de conditions de PCR comprenant un nombre de cycles plus important (*i.e.* 34 cycles), l'allèle de plus petite taille peut également être affecté (Schneider *et al.*, 2004).

Ces deux types d'absence d'amplification peuvent être en rapport avec la fragmentation du brin d'ADN mais également avec la présence d'une mutation dans la zone d'hybridation d'une des amorces (Butler, 2005). Les insertions, délétions ou les changements de base à proximité de l'extrémité 3' de la zone d'hybridation des amorces sont les mutations les plus fréquemment décrites comme responsables des absences d'amplifications (Budowle *et al.*, 2001 ; Budowle et Sprecher, 2001). Du fait des dégradations subies par les molécules d'ADN issus de tissus anciens ces modifications peuvent survenir et par conséquent être responsable de ce type d'erreurs lors du typage de STR.

La présence de pics de moindre intensité localisés en amont du pic principal peut être observée lors de l'analyse des électrophorégrammes. Ces pics surnuméraires, ou *stutters*, peuvent compliquer l'interprétation des profils lorsqu'on se trouve en présence de pics de faible intensité, fréquents lors de l'analyse d'échantillons d'ADN dégradé. Ces *stutters* correspondent généralement à l'allèle du sujet moins une répétition et proviennent d'un mésappariement lors de la réPLICATION qui va entraîner une insertion (rare) ou une délétion (cf Figure 25). Ces artefacts sont plus fréquents sur les loci présentant un faible nombre de répétition et dont le motif de base est court (Butler *et al.*, 2005).

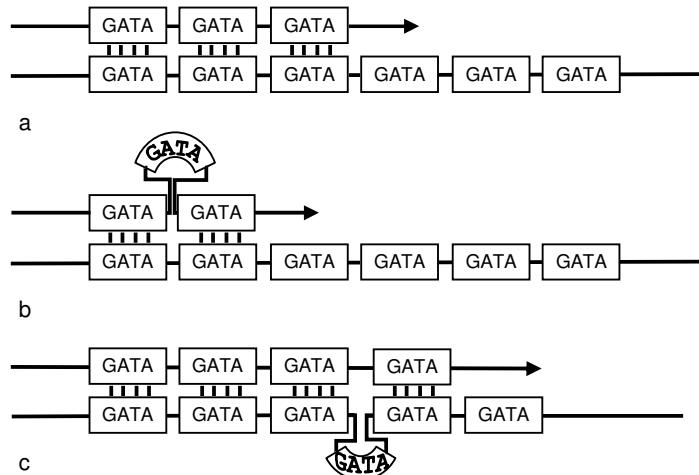


Figure 25 : Modes de formation des *stutters* : a : réPLICATION normale, b : apparition d'un stutter en raison d'un glissement en arrière, c : mode le plus fréquent de formation des stutters par glissement en avant de la polymérase.

La fréquence d'apparition de ces artéfacts peut être relativement élevée lors de l'analyse d'échantillons anciens. Afin de s'assurer de la validité des résultats, la réPLICATION de chaque amplification, pour chaque extrait mais également à partir d'ADN provenant de différentes extractions, est particulièrement importante.

III.3.2.2.3 Amplification et analyse

Pour le typage des STR autosomaux nous avons utilisé les kits Ampf/STR® Profiler Plus™ lors de la première année puis le kit Ampf/STR® Identifiler™. Le kit Profiler Plus™ permet l'amplification simultanée de 9 loci (FGA, vWA, D3S1538, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11) plus le locus de l'amélogénine. Le kit Identifiler™ comprend 6 loci supplémentaires (CSF1PO, TH01, TPOX, D16S539, D2S1338, D19S433). La taille des différents loci amplifiés ainsi que les fluorophores utilisés pour les marqueurs sont résumés dans les schémas des figures 26 et 27.

Outre un pouvoir de discrimination plus important (en raison du nombre de marqueurs accru), le kit Identifiler™ présente deux innovations techniques par rapport au kit Profiler Plus™ permettant la réDUCTION de la taille des amplicons par l'utilisation de fluorophores supplémentaires et par l'utilisation de courtes séquences hexanucléotidiques liées aux amorces de PCR permettant de séparer les fragments de tailles similaires provenant de loci différents. Cependant, l'augmentation du degré de multiplexage est souvent décrite comme

responsable d'une diminution de l'efficacité d'amplification pour des extraits dont l'ADN est dégradé et/ou en faible nombre de copies. Nous avons par conséquent évalué l'efficacité de typage de chacun des multiplexes (cf chapitre IV.3.2 Efficacité de typage).

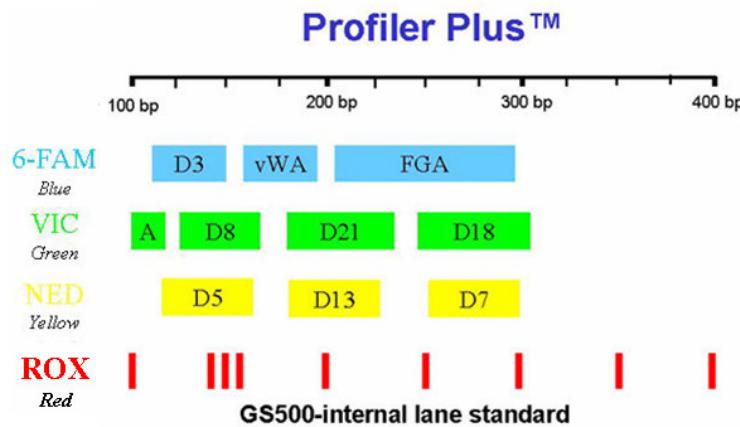


Figure 26 : Fluorophores et échelle de taille des produits de PCR suivant les loci inclus dans le kit Ampf/STR® Profiler Plus™ (modifié d'après STR DNA Internet database : <http://www.cstl.nist.gov/div831/strbase/multiplx.htm>).

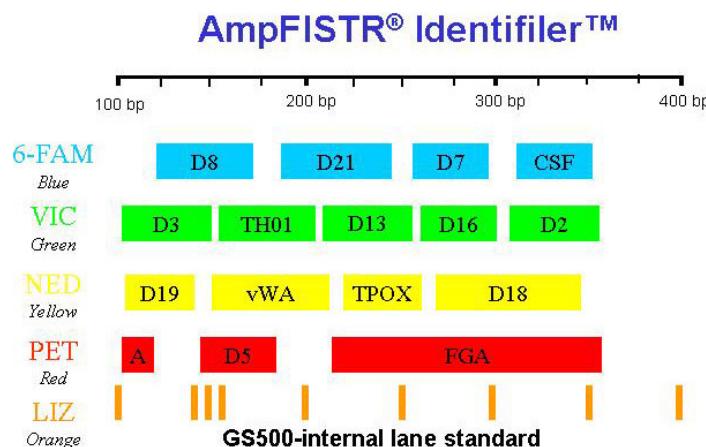


Figure 27 : Fluorophores et échelle de taille des produits de PCR suivant les loci inclus dans le kit Ampf/STR® Identifier™ (modifié d'après STR DNA Internet database : <http://www.cstl.nist.gov/div831/strbase/multiplx.htm>).

L'amplification des STR autosomaux a été réalisée conformément aux recommandations du fabricant, à l'aide d'un thermocycleur Biometra T3™ (Biometra®) qui possède 3 blocs séparés ce qui présente l'avantage de pouvoir amplifier séparément les échantillons anciens et les contrôles positifs. Le volume réactionnel final était de 10µl pour le kit Profiler Plus™

contenant : de 1 à 3 μ l d'extrait d'ADN (ajusté en fonction des résultats de la quantification) ; 3,8 μ l de PCR Amplification Mix™ ; 2 μ l de Primer Set™ ; 0,2 μ l d'AmpliTaq Gold™ (soit une unité) et de 0 à 3 μ l d'H₂O QSP 10 μ l.

Pour le kit Identifiler™, le volume réactionnel final était de 12,5 μ l contenant : de 1 à 5 μ l d'extrait d'ADN ; 5,25 μ l de PCR Mix ; 2,75 μ l de Primer Mix ; 0,25 μ l de d'AmpliTaq Gold et de 0 à 4 μ l d'H₂O QSP 12,5 μ l.

Le nombre de cycles d'amplification a été modifié en tenant compte des résultats de la quantification réalisée en amont. L'amplification a été réalisée suivant les recommandations pour le typage d'extraits contenant un faible nombre de copies qui permettent théoriquement d'amplifier une copie de séquence cible (Gill *et al.*, 2000). Les conditions d'amplification utilisées sont les suivantes : 11 minutes de dénaturation à 94°C suivies par 34 cycles, comportant chacun 1 minute de dénaturation à 94°C, 1 minute d'hybridation à 59°C et 1 minute d'elongation à 72°C. Ensuite, une période d'elongation finale de 45 min à 60°C suivie d'une phase à température constante à 4°C. Les produits d'amplification sont ensuite conservés à 4°C. L'augmentation du nombre de cycles de PCR peut, cependant, entraîner une augmentation de l'apparition de certains artéfacts d'amplification. Il est donc crucial dans ces conditions de répéter les analyses afin de pouvoir comparer les résultats entre les différentes extractions pour s'assurer de la validité des profils obtenus.

L'analyse des produits de PCR a été réalisée par électrophorèse capillaire sur un séquenceur automatique ABI PRISM® 3100 (Applied Biosystems). Pour chacun des deux kits d'amplification des STR autosomaux 1,5 μ l de produit de PCR est mélangé à 9 μ l de HI-DI Formamide (Sigma) et à 0,5 μ l de marqueur de taille (400HD ROX size Standard (Applied Biosystems) pour le kit Profiler Plus™ et GS 500 LIZ size Standard (Applied Biosystems) pour le kit Identifiler™). Les échantillons sont dénaturés durant 4 minutes à 95°C puis placés immédiatement sur une plaque réfrigérée pour éviter toute ré-hybridation des brins d'ADN. De plus, pour chaque analyse une échelle allélique est ajoutée dans un des puits afin de s'assurer de l'attribution correcte des allèles aux différents pics. Les résultats de l'électrophorèse ont été analysés avec le logiciel GeneMapper v1.02 (Applied Biosystems).

III.3.2.2.4 Détermination des relations de parentés

Les relations de proche parenté entre sujets inhumés au sein d'un même site funéraire ou dans un même *alas* ont été testées grâce aux résultats des STR autosomaux tout en tenant compte des données obtenues pour les marqueurs à transmission uniparentale pour la formulation des hypothèses.

Comme étape préalable à tout calcul, il a été nécessaire de constituer une banque contenant les fréquences alléliques des différents loci étudiés pour la population Yakoute. En effet, suivant les populations considérées, les fréquences alléliques varient de manière importante et peuvent entraîner des biais dans les calculs des relations de parenté.

Lors des missions de la MAFSO, des prélèvements ont été réalisés au sein des populations habitant à proximité des sites archéologiques fouillées. Ces prélèvements ont permis de constituer une base de données de comparaison contenant les profils génétiques de 156 individus. Les prélèvements ont été réalisés de manière anonyme et les personnes ont été informées de l'utilisation strictement limitée au cadre de recherche qui serait fait de ces prélèvements. Nous avons interrogé les personnes sur les origines de leurs parents, grands-parents et dans la mesure du possible arrière-grands-parents. Les échantillons provenant de personnes attestant d'un ancêtre d'origine européenne n'ont pas été pris en compte dans l'analyse afin de limiter les variations par rapport aux fréquences de la population yakoute ancienne. De plus, les données provenant de sujets anciens ayant permis l'obtention de profils complets ont également été incluses dans la banque de fréquence dans le but d'augmenter l'effectif total et ainsi de limiter les biais d'échantillonnage et les variations par rapport à la populations ancienne.

Les fréquences alléliques pour chaque locus ont été calculées à l'aide du logiciel YCDMA (<http://perso.orange.fr/daniel.montagnon/YCDMADFra.htm>).

Nous avons pu évaluer trois types de relations de proche parenté : (i) les relations de filiation (y compris lorsque le génotype d'un des parents est manquant) en utilisant les formules proposées par le DNAview user's group (<http://www.dna-view.com/patform.htm> et http://www.dna-view.com/mothless.htm#N_1); (ii) les relations de type fratrie d'après les formules proposées par Buckleton (2005) et (iii) les relations de type oncle-neveu, demi-frères et cousins pour lesquels les formules de calcul sont identiques (Bukleton, 2005). Les

rapports de vraisemblance (ou *likelihood ratio*, LR) ont été calculés pour les relations de parenté pouvant exister entre les sujets.

III.3.3 Les marqueurs à transmission uniparentale

III.3.3.1 Haplotypes et haplogroupes

Contrairement aux STR autosomaux qui sont transmis pour moitié par chacun des deux parents, les marqueurs portés par le chromosome Y sont transmis exclusivement par le père à ses fils et le génome mitochondrial est transmis presque exclusivement par la mère à l'ensemble de ses enfants.

A une échelle plus large, il est possible de regrouper les haplotypes au sein d'haplogroupes. Ces ensembles d'haplotypes partageant des allèles ou des mutations en commun ont une histoire évolutive commune et permettent d'appréhender les migrations et l'histoire évolutive des populations humaines.

III.3.3.1.1 La notion d'haplotype

Un haplotype correspond au génotype haploïde présent à deux ou plusieurs sites polymorphes d'un même chromosome ou d'une même région d'intérêt (e.g. un génome entier comme le génome mitochondrial ou une portion d'un chromosome) (Templeton, 2005). Les sites polymorphes définissant l'haplotype peuvent être des polymorphismes ponctuels de séquence (SNP : *Single Nucleotide Polymorphism*), des insertions/délétions ou encore les allèles de différents STR. Les individus qui partagent le même état génétique pour l'ensemble des sites polymorphes de la région d'intérêt font partie du même haplotype. Ainsi il sera possible de définir des lignées paternelles et maternelles pour les individus qui présentent le même ensemble d'allèles pour les STR du chromosome Y ou pour les sujets possédant des mutations similaires dans les régions hypervariables de l'ADN mitochondrial.

Au sein d'une même population il est possible de comparer les haplotypes des différents individus afin de définir les lignées paternelles/maternelles présentes. Cette comparaison

intra-population permet, pour les individus inhumés à proximité, de mettre en évidence l'appartenance à une même cellule familiale ou d'envisager le type de structure sociale qui prévalait dans la population.

L'accumulation des mutations au cours du temps va conduire à la différenciation des haplotypes ce qui permet de proposer des schémas de dispersion en relation avec les trajectoires des différents flux de gènes et par inférence avec les mouvements des populations (Underhill, 2003). La comparaison de la distribution et de la fréquence des haplotypes de manière inter-populationnelle permet d'apporter des informations sur les affinités entre groupes ethniques. Cependant, suivant les marqueurs considérés, il est parfois difficile de retrouver des haplotypes partagés entre les populations il devient alors nécessaire de considérer les haplogroupes auxquels sont affiliés les haplotypes afin d'avoir une vision des événements à une échelle plus large.

Pour les études en anthropologie moléculaire, il est possible de comparer, pour une même population, les fréquences haplotypiques mais à des époques différentes. La mise en évidence des changements survenus au cours du temps permet de visualiser les modifications subies par la population ou au contraire la pérennité de certaines lignées. Il faut toutefois prendre en compte les effectifs étudiés et garder à l'esprit qu'un biais d'échantillonnage peut être responsable de l'absence d'un haplotype ancien dans la population moderne sans qu'il y ait eu disparition de celui-ci. Dans ce cas, il sera également intéressant de recourir à la distribution et aux fréquences des haplogroupes partagés entre les différentes périodes.

III.3.3.1.2 La notion d'haplogroupe

La notion d'haplogroupe peut renvoyer à différentes définitions : un haplogroupe peut correspondre à un ensemble d'haplotypes dérivant d'un haplotype ancestral commun et présentant des allèles ou des mutations en commun pour l'ADN mitochondrial. Cependant, concernant le chromosome Y, un haplogroupe correspond généralement à une lignée paternelle caractérisée par la présence d'un ou plusieurs SNP (de Knijff, 2000). Les haplogroupes sont généralement continents ou régions spécifiques (Pakendorf et Stoneking, 2005) et certains peuvent parfois être spécifiques à une population donnée (cf Figure 28).

Dix huit haplogroupes principaux sont définis pour le Chromosome Y et il existe plus de 25 haplogroupes pour l'ADN mitochondrial. Afin d'augmenter la résolution et les possibilités de différenciation des haplogroupes, des sous-haplogroupe sont généralement définis (cf Figure 29).

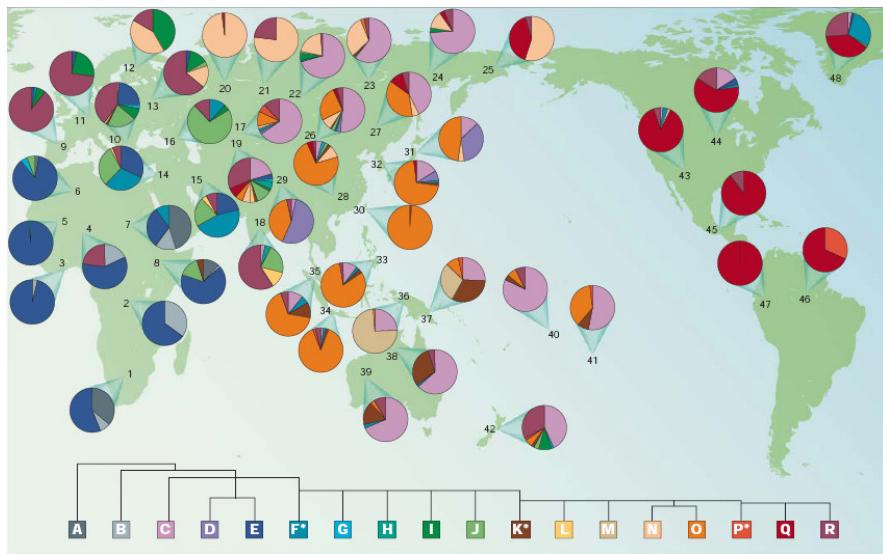


Figure 28 : Carte de la distribution globale des haplogroupes du chromosome Y (d'après Jobling et Tyler-Smith, 2003).

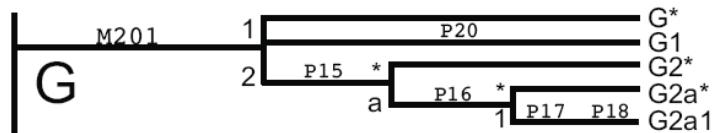


Figure 29 : Exemple de réticulation d'un haplogroupe en sous-haplogroupes. L'haplogroupe du chromosome Y, G, a été subdivisé en différents sous-haplogroupes en fonction des mutations.

L'affiliation à un haplogroupe se fait, pour le chromosome Y comme pour l'ADN mitochondrial, en considérant un ensemble de positions polymorphes dont les positions sont déterminées par rapport à une séquence de référence (*e.g.* séquence de l'ADN mitochondrial publiée par Anderson en 1981). La détermination fine de l'haplogroupe d'un individu ancien peut par conséquent présenter des difficultés du fait de la nature dégradée de la molécule d'ADN. Les études sur les populations actuelles utilisent en majorité des enzymes de restrictions permettant de différencier les bases présentes à une position donnée. Cependant ces méthodes ne sont pas adaptées à l'étude de l'ADN ancien en raison des quantités d'ADN nécessaires puisque chacune des positions doit être analysée à l'aide d'une enzyme de restriction spécifique ce qui implique des amplifications séparées. De plus, les risques de digestion partielle peuvent entraîner des erreurs d'interprétation des résultats (Kolman et

Tuross, 2000). D'autres méthodes applicables à l'ADN ancien sont aujourd'hui disponibles pour la détection de ces mutations ponctuelles: la méthode APLP (*Amplified Product Length Polymorphism*) (Adachi *et al.*, 2004), la technique d'extension d'amorce SNaPshot (Brion *et al.*, 2005 ; Bouakaze *et al.*, soumis (Annexe 4) ; Lee *et al.*, 2006) ou encore la spectrométrie de masse MALDI-TOF (Petkovski *et al.*, 2005).

Il est également possible d'affilier un haplotype à un haplogroupe par inférence en comparant par la méthode de *near-matching* (Yao *et al.*, 2002) l'haplotype du sujet à des haplotypes déjà publiés et pour lesquels toutes les informations sont disponibles. Dans notre étude, nous avons utilisé cette méthode afin de déterminer les haplogroupes des sujets anciens analysés. Dans le cas où la séquence exacte n'est pas retrouvée, il existe des positions dans les régions hypervariables que l'on peut qualifier de diagnostiques pour les différents haplogroupes. Il est possible de vérifier la présence de ces positions dans l'haplotype du sujet que l'on désire affilier et ainsi de rattacher de manière indirecte une séquence à un haplogroupe. La construction de Network (Bandelt *et al.*, 1999) dans lesquels les haplotypes des sujets anciens sont analysés conjointement avec des haplotypes dont l'affiliation est connue peut représenter une alternative aux méthodes précédentes.

III.3.3.2 Les marqueurs du chromosome Y

Les marqueurs présents sur le chromosome Y dans la région non recombinante (NRY) sont transmis de manière strictement paternelle de génération en génération et ce sans subir de modification excepté des mutations. Il est ainsi possible de suivre la transmission des lignées paternelles au cours du temps et d'avoir accès aux schémas de migrations et aux événements démographiques survenus dans les populations masculines. Il existe deux types de marqueurs sur le chromosome Y : les STR qui permettent de définir l'haplotype du sujet et les marqueurs bi-alléliques, SNP et insertion *Alu*, qui caractérisent les haplogroupes. L'étude conjointe de ces deux types de marqueurs, dont les taux de mutations sont différents, permet donc d'avoir une vue d'ensemble de l'évolution des lignées masculines (Jobling et Tyler-Smith, 1995).

III.3.3.2.1 Les STR

L'étude des lignées paternelles de nos échantillons anciens a été effectuée par l'analyse des STR localisés sur le chromosome Y. Les différents loci analysés sont localisés dans la région non recombinante du chromosome Y (NRY) qui représente 95% de celui-ci. Ces marqueurs sont liés et se transmettent par conséquent en un seul bloc aux sujets masculins de la génération suivante ce qui permet de suivre leur transmission sur de nombreuses générations. Les taux de mutation des différents loci ne sont pas identiques et une corrélation existe entre la taille du motif répété et le nombre moyen de répétitions (Forster *et al.*, 2000). Comme pour les STR autosomaux les changements d'une répétition sont plus fréquents que les modifications affectant plusieurs répétitions, tout comme les gains d'un allèle par rapport à une perte (Kayser *et al.*, 2000 ; Dupuy *et al.*, 2004). Les taux de mutation varient suivant les loci considérés, ils sont toutefois relativement élevés et le taux moyen est de $3,17 \times 10^{-3}$ (Kayser *et al.*, 2000). Il existe une contre partie à ces taux de mutation plus important : il est possible de retrouver des haplotypes identiques non pas par descendance mais par convergence (de Knijff *et al.*, 2000). Deux individus pourront par conséquent posséder le même haplotype sans pour autant partager un réel ancêtre commun ce qui peut représenter un biais lors de la recherche des haplotypes partagés et conduire à des erreurs importantes lors des analyses phylogénétiques.

Pour le typage des STR nous avons utilisés deux kits commerciaux : le kit PowerPlex® Y (Promega) qui permet d'amplifier simultanément 11 loci (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) puis le kit AmpF/STR® Y-Filer™ (Applied Biosystems) qui contient 5 loci supplémentaires (DYS448, DYS456, DYS458, DYS635 et GATAH4) (Figures 30a et 30b). L'utilisation dans le kit Y-Filer™ de cinq fluorophores permet d'augmenter le degré de multiplexage sans augmenter la taille des amplicons ce qui est un point crucial dans l'analyse d'ADN dégradé ou en faible nombre de copies. De plus, l'augmentation du nombre de loci confère une meilleure précision dans la définition des haplotypes. Le risque d'erreur d'obtenir des similitudes entre deux individus en raison d'évènements de mutation convergente et non pas d'une appartenance à une même lignée paternelle est par conséquent réduit.

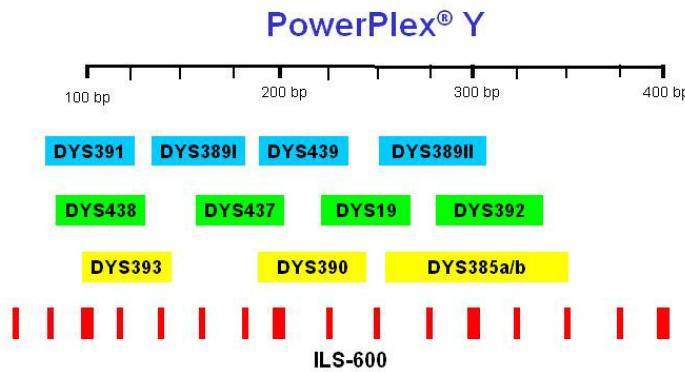


Figure 30a : Fluorophores et échelle de taille des produits de PCR des loci inclus dans le kit PowerPlex® Y (modifié d'après STR DNA Internet database : <http://www.cstl.nist.gov/div831/strbase/multiplx.htm>).

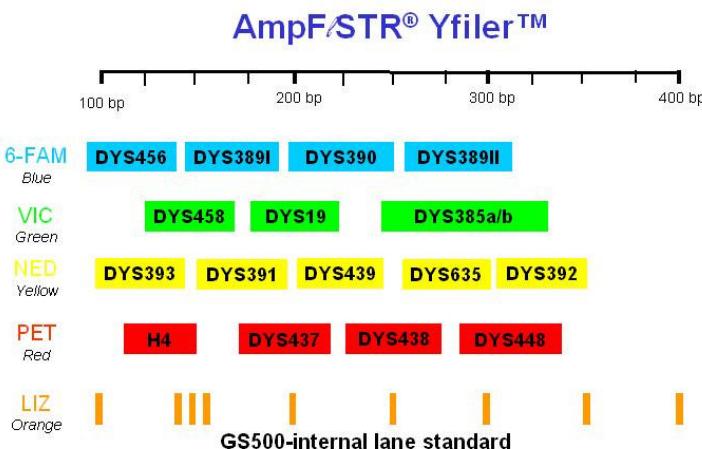


Figure 30b : Fluorophores et échelle de taille des produits de PCR des loci inclus dans le kit Ampf/STR® Yfiler™ (modifié d'après STR DNA Internet database : <http://www.cstl.nist.gov/div831/strbase/multiplx.htm>).

Considérant le faible nombre de copies de départ, nous avons appliqué les mêmes conditions d'amplification concernant le nombre de cycles que pour les STR autosomaux.

Pour le kit PowerPlex® Y, la PCR a été réalisée dans un volume final de 10 μ l contenant : 1,25 μ l de tampon ; 1,25 μ l de Primer Mix ; 0,3 μ l d'AmpliTaq Gold ; un volume d'extrait variant de 1 à 9,7 μ l suivant la concentration en ADN et un volume d'eau QSP 10 μ l. L'amplification a été réalisée sur un thermocycleur T3 (Biometra) dans les conditions suivantes : une première étape d'activation de l'enzyme durant 11min à 95°C ; 2min de

dénaturation à 96°C puis 10 cycles comprenant : 1min de dénaturation à 94°C, 1min d'hybridation à 60°C et 1min30sec d'elongation à 70°C ; suivis par 22 cycles comprenant : 10min de dénaturation à 90°C, 1min d'hybridation à 58°C et 1min30sec d'elongation à 70°C et enfin une phase d'extension finale de 30min à 60°C.

Pour le kit AmpflSTR® Yfiler™, la PCR a été réalisée dans un volume final de 12,5µl contenant : 4,6µl de PCR Mix, 2,5µl de Primer Mix, 0,4µl d'AmpliTaq Gold, un volume d'extrait variant de 1 à 5µl suivant la concentration et une volume d'eau QSP 12,5µl. L'amplification a été réalisée en appliquant les conditions suivantes : 11min d'activation de l'enzyme à 95°C ; 34 cycles comprenant 1min de dénaturation à 94°C, 1min d'hybridation à 61°C, 1min d'elongation à 72°C et enfin 80min d'extension finale à 80°C.

Pour les deux kits, l'électrophorèse et l'analyse de fragments ont été réalisées sur un séquenceur automatique ABI PRISM® 3100 (Applied Biosystems) en utilisant les paramètres recommandés par le fabricant. L'analyse des résultats a été réalisée à l'aide du logiciel Genemapper v1.02 (Applied biosystems).

III.3.3.2.2 Les SNP

Les SNP présents sur le chromosome Y sont des marqueurs plus stables que les STR en raison de leur taux de mutation moins élevé (2×10^{-8} mutation par base par génération (Nachman et Crowell, 2000)). Ces marqueurs permettent par conséquent de suivre des événements à une échelle de temps plus large et sont donc essentiels dans les études portant sur l'histoire évolutive des populations humaines (Underhill, 2003).

Il semble possible de suivre les événements de dispersion et de séparation survenus depuis la sortie d'Afrique des Hommes anatomiquement modernes en retracant l'apparition des différentes mutations, les SNP permettant de ponctuer les différentes étapes des grandes migrations humaines. Ainsi chacun des 18 haplogroupes majeurs peut être caractérisé par un SNP spécifique, il est par conséquent possible, en s'appuyant sur les données de la littérature, d'établir un panel de SNP qui correspondront aux haplogroupes les plus fréquemment rencontrés dans une population donnée (Brion *et al.*, 2005). Une caractérisation plus fine de chaque haplogroupe est également possible puisque 153 haplogroupes ou sous-haplogroupes ont ainsi été caractérisés par le typage de 245 SNP (http://ycc.biosci.arizona.edu/nomenclature_system/results.html).

L'analyse des SNP du chromosome Y a été réalisée au laboratoire en utilisant deux approches différentes : le typage par spectrométrie de masse MALDI-TOF et l'extension d'amorces à l'aide du kit SNaPshot™ (Applied Biosystems). Nous ne détaillerons pas ces techniques qui sont décrites respectivement dans la thèse de Doctorat du Dr Petkovski et dans l'article de Bouakaze et collaborateurs (Annexe 4).

La méthode MALDI-TOF a été utilisée afin de typer le SNP TAT qui est présent à une fréquence très importante dans la population Yakoute, de 80 à 100% suivant les études (Zerjal *et al.*, 1997 ; Karafet *et al.*, 1999 ; Lahermo *et al.*, 1999 ; Pakendorf *et al.*, 2002 et 2006) et qui détermine l'appartenance à l'haplogroupe N3.

Le kit SNaPshot a été utilisé pour l'analyse simultanée de 13 SNP (RPS4Y, M216, M217, M89, M9, TAT, M175, M242, M3, M173, M17) qui permettent de déterminer l'affiliation à 11 des principaux haplogroupes rencontrés en Sibérie, en Mongolie et chez les populations Amérindiennes (C, C3, F, K, N3, P, O, Q, Q3, R1, R1a1).

La quantité d'extrait d'ADN à disposition étant un facteur limitant pour l'analyse des SNP du chromosome Y, nous avons limité le typage en prenant en compte les haplotypes STR partagés dans nos échantillons anciens mais également avec les échantillons modernes. Les SNP étant des marqueurs signant des événements plus anciens que les STR, des haplotypes STR identiques appartiennent au même haplogroupe.

III.3.3.3 L'ADN mitochondrial

III.3.3.1 Généralités et propriétés

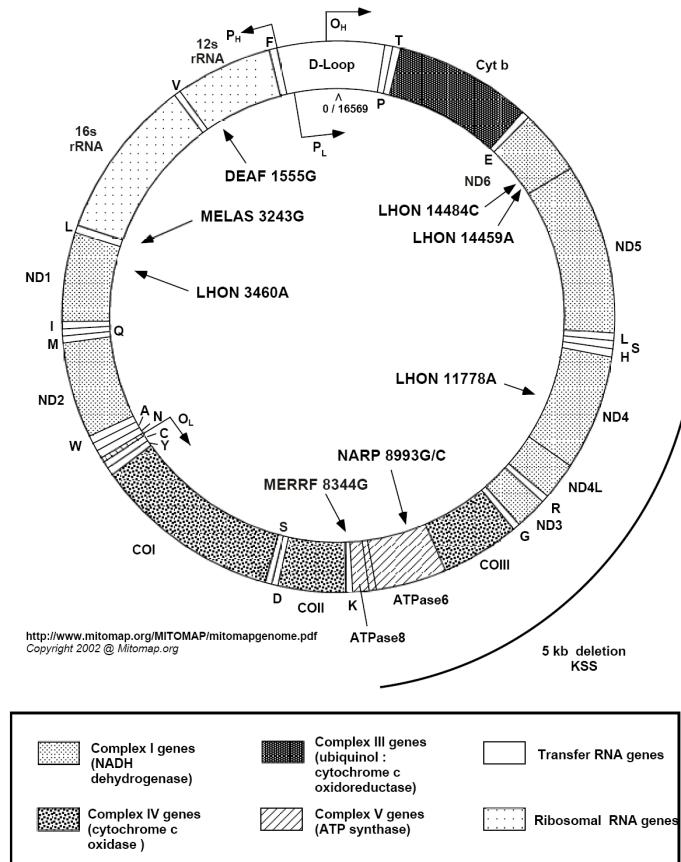


Figure 31 : Carte schématique de l'ADN mitochondrial humain (d'après MITOMAP: A Human Mitochondrial Genome Database. <http://www.mitomap.org>, 2006).

Le génome mitochondrial est principalement composé de régions codantes à l'exception d'un fragment d'une longueur d'environ 1100pb qui est impliqué dans les fonctions de régulation : la région contrôle ou D-loop, localisée entre les positions nucléotidiques 16024 et 576. La D-loop comprend trois régions hypervariables comprises entre les positions 16024 et 16400, 44-340 et 438 et 576 désignées respectivement par HV1 (Hyper Variable 1), HV2 et HV3 (Brandstatter *et al.*, 2004) (Figure 31).

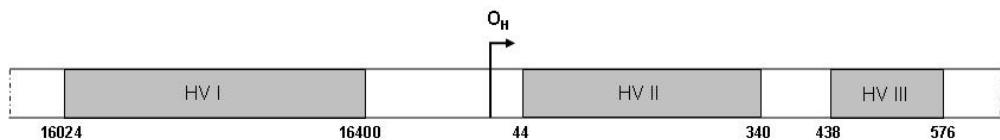


Figure 32 : Carte schématique de la région contrôle de l'ADN mitochondrial. Les positions nucléotidiques sont données en accord avec la numérotation de la rCRS (Andrews *et al.*, 1999). (O_H : origine de réPLICATION).

Depuis l'étude des variations au sein de l'ADNmt humain par Brown en 1980, ce génome a été utilisé dans de nombreuses études sur l'évolution, les migrations et les relations entre les différentes populations humaines actuelles mais aussi du passé. Cet usage largement répandu est en relation avec certaines spécificités qui font de l'ADN mitochondrial un outil particulièrement bien adapté. Ces propriétés incluent : une transmission quasi-exclusivement maternelle, une absence de recombinaison, un taux de mutation plus important que celui de l'ADN nucléaire et surtout un nombre de copies par cellules important.

Le mode de transmission uniparental fait de l'ADNmt un marqueur moléculaire particulièrement intéressant pour retracer les origines des lignées maternelles d'une population sans l'influence des recombinaisons propres à l'ADN nucléaire. Cette transmission uniquement maternelle provient du fait que seules les mitochondries présentes dans l'ovocyte seront transmises lors des divisions cellulaires. La tête du spermatozoïde étant, en effet, dépourvue de mitochondries et celles présentes dans le flagelle ne pénétrant que très rarement. De plus, les mitochondries provenant des spermatozoïdes sont éliminées dans l'ovocyte (Manfredi *et al.*, 1997 ; Shitara *et al.*, 1998). La présence de mitochondries d'origine paternelle est donc rare et on peut considérer à l'heure actuelle que l'ADNmt d'un individu est d'ascendance strictement maternelle (Giles *et al.*, 1980 ; Schwartz *et al.*, 2003; Pakendorf *et al.*, 2005).

L'absence de recombinaison est également une spécificité notable de l'ADNmt. Cette propriété fut remise en cause par la publication de quatre articles dans lesquels les auteurs avançaient avoir observé plusieurs événements de recombinaison au sein du génome mitochondrial (Awadalla *et al.*, 1999 et 2000; Eyre-Walker *et al.*, 1999; Hagelberg *et al.*, 1999). La vérification ultérieure de ces données permis de mettre en évidence que ces conclusions provenaient, en fait, d'erreurs lors de l'alignement des séquences ou de données

erronées (Arctander 1999 ; Jorde et Bamshad, 2000 ; Kumar *et al.*, 2000 ; Merriwether et Kaestle, 2000 ; Parsons *et al.*, 2000). Toutefois, deux équipes ont mis en évidence des recombinaisons chez des sujets dont l'ADNmt provenait des deux parents (Schwartz et Vissing, 2002 ; Kraytsberg *et al.*, 2004). Cette découverte souligne que les recombinaisons sont possibles puisque les mitochondries possèdent une recombinase fonctionnelle (Thyagarajan *et al.*, 1996). Toutefois, les modalités de fusion des mitochondries dans les cellules qui permettraient aux génomes de se recombiner restent à préciser (Enriquez *et al.*, 2000 ; Ono *et al.*, 2001 ; Legros *et al.*, 2002). En outre, ces événements sont restreints à des cas exceptionnels, la recombinaison ne représente, par conséquent, pas un problème majeur dans les études de génétique des populations ou d'anthropologie moléculaire.

Les taux de mutation de l'ADN mitochondrial sont beaucoup plus élevés que ceux du génome nucléaire. Cette différence serait due à un renouvellement des mitochondries plus rapide que celui des cellules (Brown *et al.*, 1982). De plus, la matrice mitochondriale est un milieu très oxygéné ce qui va favoriser les dommages oxydatifs subis par l'ADNmt qui ne possède pas d'histones protectrices et présente des mécanismes de réparation limités (Richter *et al.*, 1988 ; Wallace 1994). Enfin, la polymérase des mitochondries possède une fiabilité réduite (Budowle *et al.*, 2003). Le taux de mutation est d'environ $1,70 \times 10^{-8}$ substitutions par site et par an si on considère le génome entier et que l'on exclut la région contrôle (Ingman *et al.*, 2000). En effet, les taux de mutation observés dans la D-loop sont plus importants mais la valeur exacte est fonction de la position nucléotidique considérée. La région contrôle présente des positions pour lesquelles les taux de mutation sont hétérogènes, certaines positions, ou "*hot spot*" (cf *infra*), étant soumises à des mutations de manière récurrente qui vont entraîner des taux de mutations jusqu'à 5 fois supérieurs à ceux normalement observés (Meyer *et al.*, 1999 ; Heyer *et al.*, 2001). De plus, l'estimation précise du taux de mutation diffère entre les études basées sur l'étude de généralogies et celles basées sur la phylogénie (Pääbo, 1996 ; Macaulay *et al.*, 1997 ; Jazin *et al.*, 1998). Dans le premier cas les sites à évolution rapide sont principalement détectés alors que dans le second tous les sites polymorphes sont pris en compte mais les événements de dérive et de sélection vont influer sur les estimations (Howell et Smejkal, 2003). Les valeurs des dates de coalescence entre deux lignées maternelles peuvent par conséquent varier grandement suivant les taux de mutation pris en compte. Pour les études portant sur des événements anciens, il sera préférable d'utiliser les valeurs obtenues à partir des études phylogénétiques alors que pour l'histoire d'une population spécifique

l'utilisation de modèles prenant en compte des taux de mutation variables observées dans les généalogies sera préférable (Oota *et al.*, 2001 ; Heyer *et al.*, 2003).

Chaque mitochondrie possède plusieurs copies du génome et les cellules possèdent un grand nombre de mitochondries (jusqu'à plusieurs milliers, le nombre exact étant fonction du tissu considéré). Par conséquent, le nombre de copies du génome mitochondrial dans chaque cellule est très important. Ceci représente un avantage lors de l'amplification des séquences cibles et en fait un outil de choix pour les études portant sur des échantillons anciens ou pour l'analyse de cas médicaux légaux. En effet, il sera beaucoup plus aisément d'étudier l'ADNmt par rapport à l'ADN nucléaire dont chaque cellule ne présente qu'une seule copie. Les études portant sur les échantillons âgés de plusieurs dizaines de milliers d'années tels les Néandertaliens portaient, jusqu'à présent, exclusivement sur l'ADNmt (Krings *et al.*, 1997 ; Orlando *et al.* 2006). Il a également été rapporté dans une étude récente que la membrane des mitochondries pouvait représenter une barrière vis à vis des agents participant à la dégradation des acides nucléiques. De ce fait l'ADNmt, en plus de sa présence en un plus grand nombre de copies, pourrait être mieux protégé que l'ADN nucléaire (Gilbert *et al.*, 2006b).

Le nombre important de mitochondries dans la cellule peut néanmoins compliquer l'interprétation des analyses par l'apparition d'une variabilité individuelle et fait apparaître une des premières contraintes inhérentes à l'analyse du génome mitochondrial.

III.3.3.3.2 Les contraintes

III.3.3.3.2.1 *Les hétéroplasmies*

Une hétéroplasmie est définie par la coexistence dans une cellule ou un tissu de deux ou plusieurs populations de mitochondries possédant un génome dont la composition nucléotidique est différente. En prenant en compte cette définition il est possible de distinguer deux types d'hétéroplasmies : les hétéroplasmies ponctuelles, qui n'affectent qu'une position donnée et les hétéroplasmies de longueur, qui vont se manifester par l'insertion ou la délétion d'un ou plusieurs nucléotide.

Les hétéroplasmies ponctuelles peuvent provenir de la présence de mitochondries d'origine paternelles ; cependant comme nous l'avons vu précédemment ces cas sont extrêmement rares. Les cas les plus fréquents sont représentés par la transmission dans un tissu d'une

population de mitochondries dont le génome a subit une ou plusieurs mutations qui ne sont pas retrouvés dans les autres populations mitochondriales du même individu.

Les mécanismes ainsi que la fréquence des évènements conduisant à la persistance de ces différentes populations de mitochondries restent encore assez mal connue. Cependant il semble que des processus de ségrégation interviennent très tôt durant l'oogenèse ce qui entraînerait des phénomènes de goulot d'étranglement (Brown *et al.*, 2001 ; Poulton et Marchington, 2002). Lors de cette sélection certains variants pourraient être transmis dans des proportions plus ou moins importante définissant le taux d'hétéroplasmies chez le sujet adulte. En outre, certains tissus sont décrits comme présentant des taux d'hétéroplasmies important. Les cellules corticales de la tige des cheveux peuvent notamment posséder des mitochondries provenant des cellules germinales du bulbe et d'autres provenant des mélanocytes (Linch *et al.*, 2000) ce qui pourra entraîner des taux d'hétéroplasmies élevés si des mutations différentes ont affecté les populations de mitochondries présentes dans ces deux lignées cellulaires. Ainsi la fréquence des positions hétéoplasmiques dans les séquences obtenues à partir d'ADN extrait de cheveux sont particulièrement élevées (Grzybowski, 2000; Grzybowski *et al.*, 2003; Brandstätter *et al.*, 2003; Sekiguchi *et al.*, 2004) (cf Figure 33).

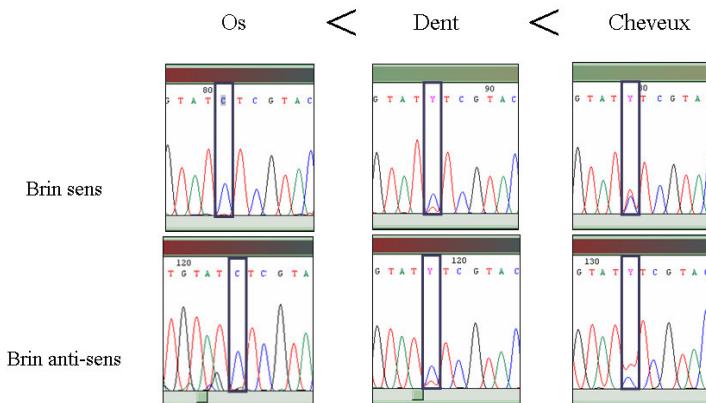


Figure 33 : Exemple d'hétéroplasmie en position 16093 observée sur des séquences obtenues à partir d'extraits d'ADN provenant de différents substrats (sujet Munur Urek 1).

La répétition de motifs mono-nucléotidiques (ou *stretch*) peut entraîner des erreurs de la polymérase lors de la réplication qui pourront être à l'origine d'hétéroplasmies de longueur. Les hétéroplasmies sont principalement rencontrées dans les régions HV1 et HV2 en raison de transitions T vers C à des positions spécifiques (*i.e.* 16189 et 310 respectivement). Ces mutations vont être responsables de l'apparition d'un *stretch* de C entre les positions 16384-16393 dans HV1 (Bendall and Sykes 1995) et 303-315 dans HV2 (Greenberg *et al.* 1983). La

proportion des hétéoplasmies de longueur peut varier entre les tissus d'un même individu (Pfeiffer *et al.* 2004) et au sein des descendants d'une même lignée maternelle (Parson *et al.*, 1998 ; Lutz *et al.*, 2000 ; Brandstätter *et al.*, 2004). Outre ces différences qui peuvent fausser les interprétations, les *stetches* de C sont responsables de la formation de motifs dits "out-of-phase" rendant l'interprétation des électrophorégrammes impossibles. Pour les sujets présentant ces deux transitions il est nécessaire de recourir à l'utilisation d'amorces s'hybridant à proximité de la zone de répétition nucléotidique (Szibor *et al.*, 1997 ; Rasmussen *et al.*, 2002) afin de pouvoir améliorer la qualité des séquences et les rendre ainsi interprétables.

L'âge du sujet peut également avoir une influence sur la fréquence observée des hétéoplasmies en raison de l'accumulation des dommages oxydatifs (Thèves *et al.*, 2006).

III.3.3.3.2 Les hot spots de mutation in vivo et postmortem

Plusieurs études ont rapporté que certaines positions spécifiques (e.g. 16093, 16126, 16129, 16172, 16183, 16189, 16192, 16261, 16311, 16362) présentaient des fréquences de mutations beaucoup plus élevées que celles observées pour le reste du génome mitochondrial (Hasegawa *et al.* 1993 ; Wakeley 1993 ; Excoffier and Yang 1999 ; Meyer *et al.* 1999 ; Heyer *et al.*, 2001). Ces positions présenteront plus fréquemment des hétéoplasmies et surtout pourront être le siège de mutations en retour pouvant compliquer considérablement les interprétations phylogénétiques, en particulier si ces positions correspondent à des sites diagnostic pour l'attribution des haplogroupes. Ces *hot spots* de mutation sont généralement plus fréquents dans les régions non codantes de l'ADN mitochondrial car moins soumis à la sélection. La fréquence des mutations serait également en relation avec la structure secondaire et tertiaire du génome mitochondrial qui entraînerait une sensibilité accrue de certaines zones aux dommages (Gilbert *et al.*, 2003a).

Lors de l'étude d'échantillons anciens, l'analyse des données peut se révéler encore plus complexe, en particulier si le nombre de copies de départ est inférieur à 1000 (Handt *et al.* 1996; Krings *et al.* 1997). Dans ce cas, l'influence des dommages *postmortem* sera beaucoup plus marquée sur les données obtenues. Tout comme pour l'ADN *in vivo*, il semble que la structure de l'ADN soit à mettre en relation avec les dommages subis par le génome mitochondrial à certaines positions spécifiques et par conséquent au taux de mutation plus

important observé dans ces régions. Ainsi, il existe une correspondance entre les sites fréquemment mutés *in vivo* et les hot spots de mutation post mortem (Gilbert *et al.*, 2003b). L'hétérogénéité des taux de mutations entre les différentes positions polymorphes du génome mitochondrial mais également au sein même des régions de la *D-loop* (Gilbert *et al.*, 2003b) va entraîner un biais très important dans les calculs des temps de coalescence entre deux lignées quand les calculs prennent en compte une valeur moyenne. De même, il semble difficile de proposer une calibration de l'horloge moléculaire fiable sans prendre en compte les variations importantes dans les taux de mutations aux différentes positions ou dans les différentes régions du génome mitochondrial.

III.3.3.2.3 Les mutations fantômes

Les procédés de séquençage peuvent parfois conduire à des séquences erronées pour lesquelles certaines positions seront mal analysées. Ces fausses mutations sont nommées mutations fantômes (Bandelt *et al.*, 2002). Plusieurs paramètres peuvent conduire à de telles erreurs : (i) un protocole de séquençage non optimisé : par exemple l'utilisation de la technique de PCR nestée qui est particulièrement propice à la génération des mutations fantômes (Brandstätter et Parson, 2003 ; Grzybowski *et al.*, 2003) ; (ii) la non-incorporation de ddNTP entraînant la présence de pics traînants ; (iii) les artefacts de compression liés à formation de structures secondaires stables lors de l'électrophorèse ; (iv) la non-attribution d'une base présente dans la séquence réelle mais quasiment invisible sur l'électrophorégramme sont des facteurs conduisant à l'apparition de mutations fantômes (Brandstätter *et al.*, 2005).

Certaines études de génétique des populations incluant un grand nombre d'échantillons ont souvent été critiquées en raison de la fréquence importante de mutations fantômes dans leurs données (e.g. Nasidze et Stoneking, 2001). Mais les domaines dans lesquels une attention particulière à la qualité des résultats est attendue n'en sont pas exempts. Les résultats des études portant sur l'ADN ancien notamment sont souvent remis en cause en raison des erreurs liées aux dégradations *postmortem*, aux contaminations par de l'ADN exogène, à la faible fiabilité de la Taq lors des clonages ou encore aux erreurs survenant lors de l'électrophorèse (Brandstätter *et al.*, 2005). Les mutations fantômes sont préférentiellement détectées à certaines positions, en particulier la position 16085, mais peuvent survenir sur des sites naturellement variables tels la position 16311 (Brandstätter *et al.*, 2005).

Notre étude a été réalisée en tenant compte de la possibilité de rencontrer de tels artefacts. Chaque séquence a, par conséquent, été validée après obtention des brins sens et anti-sens et chaque mutation a été contrôlée sur plusieurs extractions pour le même individu et sur différents substrats quand cela était possible.

III.3.3.2.4 Les pseudogènes

Les pseudogènes correspondent à des copies de séquences du génome mitochondrial insérées dans le génome nucléaire. De nombreux gènes mitochondriaux sont ainsi présents, mais non fonctionnels, dans certains chromosomes. La région HV1 notamment peut être présente sous forme de pseudogène dans le chromosome 11 (Zichler *et al.*, 1995). Ces séquences n'étant pas soumises au stress oxydatif qui règne dans les mitochondries ni à leur rythme de division, leur évolution va être beaucoup moins rapide (10 à 40 fois selon Ward *et al.*, 1991). L'analyse de ces copies permettrait par conséquent d'avoir accès à des séquences "fossiles" du génome mitochondrial qui peuvent être utilisées dans les études phylogénétiques comme séquence de comparaison, ou *out group*, à la place des séquences de chimpanzés (Zischler *et al.*, 1995). Cependant, l'amplification de pseudogènes peut se produire lors d'études portant sur l'ADN mitochondrial à proprement dit. Ces amplifications pourront entraîner des erreurs d'interprétations phylogénétiques importantes si elles ne sont pas détectées. Dans le cadre des études d'anthropologie moléculaire, il semble difficile d'amplifier de manière reproductible un pseudogène en raison du faible nombre de copies d'ADN nucléaire en comparaison avec le génome mitochondrial. La reproductibilité des résultats ainsi que la vérification phylogénétique des séquences est nécessaire afin de s'assurer de la validité des séquences et écarter la possibilité d'amplification de pseudogène.

III.3.3.3 Amplification, séquençage et analyse

L'amplification de la première région hypervariable de l'ADN mitochondrial a été effectuée en utilisant deux couples d'amorces qui permettent d'obtenir deux amplicons se chevauchant : L15989 (CCCAAAGCTAAGATTCTAAT)/H16239 (TGTGTGATAGTTGAGGGTTG) (Ivanov *et al.*.. 1996 ; Gabriel *et al.*.. 2001) et L16190 (CCCCATGCTTACAAGCAAGT)/H16410 (GAGGATGGTGGTCAAGGGAC) (Gabriel *et al.*.. 2001).

De plus, dans le cas d'un *stretch* de C lié à une substitution T/C en position 16189, une amorce alternative a été utilisée afin d'obtenir des séquences lisibles dans les régions 3' du stretch : R16167 (GGGTTTGATGTGGATTGGG).

Les réactions d'amplification ont été réalisées dans un volume réactionnel final de 50 µl contenant de 2 à 6 µl d'extrait d'ADN suivant la concentration, 10 mM Tris HCl pH 8.3, 50 mM KCL, 1.5 mM MgCl₂, 1 mg/ml BSA, 200 µM de chaque dNTP (dATP, dTTP, dCTP et dGTP), 0.25 µM de chaque amorce, et 2 U de Taq Gold Star (Eurogentec) et un volume d'H₂O QSP 50µl. La PCR a été effectuée dans un thermocycleur T3 (Biometra®) en utilisant les paramètres d'amplification suivants: 10 min de dénaturation à 94°C suivies par 38 cycles, chacun comportant 30 sec de dénaturation à 94°C, 30 sec d'hybridation à 48°C pour HV1a ou HV1total, ou 30 sec à 51°C pour HV1b et 45 sec d'elongation à 72°C. Ensuite une période d'elongation finale de 5 min à 72°C et la conservation des produits d'amplification à 4°C.

Afin d'estimer la concentration des produits d'amplification, 10µl de produits de PCR mélangés à 2µl de bleu de migration (6X loading dye solution de Fermentas®) sont déposés sur un gel d'agarose à 1% contenant 1g de poudre d'agarose pour 100ml de TBE (Tris base 0,45M ; acide borique 0,45M et EDTA 0,01M) et quelques gouttes de BET (bromure d'éthidium). Après électrophorèse, l'ADN est visualisé par exposition du gel sous une lampe à ultraviolet qui permet de rendre fluorescent le BET. La comparaison avec un marqueur de taille (GeneRuler 100bp DNA ladder de Fermentas®), qui a été déposé dans un des puits du gel d'agarose et a migré en parallèle avec les échantillons, permet de visualiser la taille des fragments amplifiés et de déterminer leur quantité en ng/µl. Ceci permet de déduire la quantité d'ADN présent dans les 40µl de réactions d'amplifications restantes.

Les produits d'amplification sont ensuite purifiés sur filtre Microcon®-PCR (Millipore). Cette étape permet (i) d'éliminer les composants autres que l'ADN utilisés lors de la PCR (dNTP, amorces, etc.) et (ii) de concentrer l'ensemble des échantillons à une même concentration de 2ng.µl⁻¹ qui est la concentration optimale pour effectuer la PCR de séquence. Brièvement, 400µl de TE 1X (250µl de Tris 2M, 250µl d'EDTA 0,2M dans 50ml d'eau bidistillée) ainsi que la totalité des 40µl restant des produits de PCR sont déposés sur un filtre Microcon®-PCR, et centrifugés 15 min à 3500 rpm. On rajoute pour chaque échantillon le volume de TE 1X nécessaire à l'obtention d'une concentration de 2ng.µl⁻¹, puis on transfère le filtre sur un nouveau tube et le tout est centrifugé 4min à 3500 rpm.

La PCR de séquence est réalisée sur chaque brin indépendamment avec le kit BigDye™ Terminator Cycle Sequencing v3.1 (Applied Biosystems). Le mélange réactionnel contient : 2µl de produit de PCR, 4µl de Big Dye™ Terminator Reaction Mix, 2µl d'amorce concentrée à 10µM et 2µl d'H₂O. Les paramètres d'amplification sont les suivants : 10 sec de dénaturation à 96°C suivis par 25 cycles, chacun comportant 35 sec de dénaturation à 96°C, 15 sec d'hybridation à 50°C et 4 min d'elongation à 60°C.

Les échantillons sont ensuite précipités dans une solution d'éthanol à 60%, contenant les 10µl de produits de PCR de séquence et 4µl d'éthanol à 60%. Les tubes sont centrifugés 30 min à 12000 rpm, le surnageant est prélevé et le culot resuspendu dans une solution de rinçage de 250µl d'éthanol à 70% (afin d'enlever les dNTP, amores, etc.) et centrifugé 5 min à 12000 rpm. Le surnageant est à nouveau prélevé et les culots mis à sécher.

Les produits de la réaction de séquence de l'ADN mitochondrial sont resuspendus dans 10µl d'H₂O, et 4µl sont ensuite soumis à l'électrophorèse de séquence sur un séquenceur automatique ABI PRISM® 3100 (Applied Biosystems). La qualité des séquences est ensuite visualisée à l'aide du logiciel Sequence Navigator (Applied Biosystems) puis les séquences sont éditées avec le logiciel Chromas Pro (Technelysium Pty Ltd).

III.3.3.3.4 Procédure de clonage

Afin de remplir les critères d'authentification concernant la publication de données portant sur des séquences anciennes (Cooper et Poinar, 2001) nous avons développé, au sein du laboratoire en collaboration avec le Dr Hatsch, un protocole de clonage à haut débit sur plaques 96 puits. Ce protocole avait pour but d'éviter l'utilisation de kits commerciaux extrêmement coûteux en utilisant des solutions préparées au laboratoire. L'ensemble de la méthodologie adoptée pour cette procédure est présenté dans l'article en Annexe 5.

Pour la validation des résultats obtenus à partir du clonage, nous avons utilisé le logiciel Confidence Consensus (<http://www.macdonald.cam.ac.uk>). Ce logiciel permet de déterminer la validité de la séquence consensus obtenue à partir des différents clones d'une seule PCR. Cette méthode ne permet pas de vérifier l'authenticité de la séquence en elle même mais prend en compte les différentes positions polymorphes détectées pour l'ensemble des clones et évalue si la séquence consensus est le résultat de l'amplification d'une ou d'un mélange de séquences (Bower *et al.*, 2005).

III.3.3.4 Analyse des données des marqueurs à transmission uniparentale

Comme nous l'avons mentionné précédemment, les données obtenues sur le chromosome Y et sur l'ADN mitochondrial ont dans un premier temps servit à la formulation des différentes hypothèses sur les liens de parenté existant entre les sujets inhumés à proximité les uns des autres. Nous avons également analysé ces données dans une approche plus populationnelle. Bien que la plupart des sujets soient issus de tombes isolées, il est néanmoins possible de regrouper les individus selon deux paramètres : la période durant laquelle les sujets ont vécu et la localisation géographique de leur tombe. Les différentes analyses statistiques ont été réalisées en regroupant les sujets suivant ces paramètres et en considérant ces individus comme appartenant à une même population archéologique.

III.3.3.4.1 Constitution et mise en place des bases de données

La nécessité de comparer les résultats obtenus sur nos échantillons avec des données publiées mais fréquemment absentes des bases de données disponibles en ligne telles Genbank (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>) a conduit à la création de bases de données internes à la fois pour les haplotypes du chromosome Y et pour les séquences HV1 de l'ADN mitochondrial. De plus, la réalisation d'une base de données centrée sur les populations d'intérêt du laboratoire permet un gain de temps dans la recherche des informations.

Concernant les données des STR du chromosome Y, les données disponibles au laboratoire et les données de la littérature portant sur des populations de Sibérie, de Mongolie, d'Asie Centrale et de l'Est ainsi que d'Europe (cf Annexe 6) ont été compilées dans une base de données inclue dans le logiciel YCDMA représentant près de 4000 haplotypes. L'utilisation de cette base de données vient en complément des banques disponibles en ligne et notamment du Y-HRD (*Y-STR haplotype reference database*, <http://www.ystr.org/index.html>) qui représente un outil performant de recherche mais dans lequel toutes les données de la littérature ne sont pas soumises de manière systématique.

La base de données mitochondriale est disponible sur le serveur de l'Institut de Médecine Légale de Strasbourg (<https://iml.u-strasbg.fr>) et comprend à l'heure actuelle plus de 2600 séquences HV1 provenant d'échantillons traités au laboratoire ou collectées dans la littérature (cf Annexe 8). Il est possible de faire des recherches afin d'identifier les correspondances

entre une séquence donnée et celles présentes dans la banque. D'autres fonctions ont également été implémentées : une fonction de recherche de contamination par comparaison avec des séquences de références correspondant aux membres de l'équipe, diverses fonctions d'exportation sous forme de tableaux, afin d'éviter les erreurs de retranscription des données, ou encore au format des principaux logiciels d'analyse statistique (ARLEQUIN et Network).

III.3.3.4.2 Haplotypes et haplogroupes partagés

Les fréquences de chaque haplotype et haplogroupe sont calculées pour les populations archéologiques définies suivant les données temporelles et géographiques. Ces fréquences vont permettre de réaliser des comparaisons avec les autres populations à disposition dans les bases de données.

Au niveau individuel il est possible de définir les affinités ethniques et génétiques d'un sujet par la recherche de son haplotype et de sa fréquence au sein de différentes populations. Cette recherche a été effectuée pour les haplotypes du chromosome Y à l'aide du logiciel YCDMA dans la base de données du laboratoire. Nous avons également utilisé les bases de données mise à disposition en ligne : Y-HRD (<http://www.ystr.org/index.html>), ainsi que deux bases de données de fournisseurs : Applied Biosystems AMPf/STR® Y-filer™ database (<http://www.appliedbiosystems.com/yfilerdatabase/>) et Promega PowerPlex® Y database (<http://www.promega.com/techserv/tools/pplexy/pplexy.asp>). Pour les séquences HV1 de l'ADNmt nous avons utilisé le logiciel BLAST 2 (*Basic Local Alignment Search Tool*) à la fois pour les recherches dans la base de données du laboratoire mais également dans les bases de données de séquences nucléotidiques disponibles en ligne sur le site NCBI (<http://www.ncbi.nlm.nih.gov/Database/index.html>). Ce programme compare les séquences entre elles et calcule le pourcentage d'identité existant entre les deux séquences.

A l'échelle d'une population il est possible d'appréhender les changements survenus au cours du temps et par conséquent les différentes influences subies par la population par l'observation des variations dans la fréquence des différents haplotypes. Cependant, les haplotypes présents dans la population étudiée ne sont pas toujours retrouvés dans les différentes populations de référence du fait de leur disparition au cours du temps ou simplement en raison de l'échantillonnage. Il faut donc recourir à l'haplogroupe auquel cet haplotype est affilié.

La détermination des haplogroupes du chromosome Y a été effectuée par l'utilisation du multiplex SNaPshot directement sur les échantillons anciens lorsque la quantité d'extrait d'ADN disponible nous le permettait ou sur des sujets modernes présentant le même haplotype. Pour certains échantillons il n'a cependant pas été possible de déterminer leur affiliation en raison du manque de matériel et de l'absence de correspondance dans les échantillons actuels à notre disposition.

Pour les séquences mitochondrielles l'attribution d'un haplotype à un haplogroupe donné a été réalisée par inférence en comparant les positions mutées observées dans la région HV1 de nos sujets avec les données publiées pour lesquelles l'analyse des positions présentes dans la région codante étaient disponibles. Cette méthode permet dans la plupart des cas de déterminer l'haplogroupe du sujet. Dans le cas d'haplotypes présentant des mutations spécifiques il est parfois possible d'affilier la séquence à un sous-haplogroupe. Il est toutefois important de garder à l'esprit que cette méthode reste empirique et peut entraîner certaines incertitudes dans le cas de séquences dont toutes les positions ne sont pas retrouvées.

L'étude de la répartition et des fréquences des différents haplogroupes présents dans un groupe d'individus permet d'avoir une vision globale de l'histoire et de l'évolution des populations. Le fait que les haplogroupes reflètent l'appartenance à une lignée commune va permettre d'estimer les mélanges survenus entre populations lors de migrations (Comas *et al.*, 1998 ; Tolk *et al.*, 2001) ou originaires de différentes régions géographiques (Pereira *et al.*, 2001 ; Salas *et al.*, 2001). De plus, l'observation des haplogroupes permet par exemple de mettre en évidence les événements de goulot d'étranglement pouvant survenir lors des migrations humaines, par exemple dans le cas de la colonisation de l'Amérique (Stone et Stoneking, 1998 ; Schurr *et al.*, 2004) ou encore d'apporter des informations sur le peuplement ancien de l'Europe (Haak *et al.*, 2005 ; Pereira *et al.*, 2005).

Ces deux paramètres estimés au sein de notre population archéologique ont été comparés avec les données publiées sur les populations originaires d'Asie Centrale, de Mongolie et de la région du Baïkal, régions supposées être le berceau de la population yakoute. Nous avons également réalisé des comparaisons avec les données disponibles sur les populations de l'Altaï, les populations de Sibérie du sud et également sur les populations vivant de nos jours en Yakoutie, Evenks et Evènes, et qui occupaient la Yakoutie Centrale vraisemblablement avant l'arrivée des Yakoutes. Quelques études portant sur l'ADN mitochondrial ont été publiées pour ces populations d'origines toungousses cependant il est difficile de vérifier la validité des échantillonnages réalisés en raison des difficultés qui peuvent être rencontrées sur

le terrain et des problèmes engendrés par les définitions des groupes ethniques proposées par les Russes qui ne correspondent pas toujours à l'appartenance réelle des individus.

III.3.3.4.3 Analyses statistiques

Une fois les différentes populations archéologiques définies au sein de l'échantillon ancien et en prenant en compte les liens de parenté éventuels entre les différents sujets, nous avons calculé les paramètres standards de diversité génétique intra-populations (diversité génétique, h : diversité haplotypique (Nei, 1987) et Pi : diversité nucléotidique (Nei, 1987)) à l'aide du logiciel ARLEQUIN v2.00 (Schneider *et al.*, 2000).

Afin de pouvoir visualiser les différences entre populations il était nécessaire d'utiliser une statistique qui inclue des informations sur des variables telles que la localisation géographique ou l'époque à laquelle ont vécu nos sujets, c'est à dire qui puisse prendre en compte les différences entre les groupes. Nous avons donc choisi d'utiliser différentes matrices de F_{st} par paire en fonction des paramètres considérés pour la définition des sous-groupes de population. L'indice F_{st} est un indice de fixation indiquant le degré de subdivision d'une population. Il est égal à la corrélation de deux gènes pris dans une subdivision de la population, par rapport à deux gènes pris au hasard dans la population totale (Ray, 2003).

Dans le cadre de nos analyses, les indices F_{st} ont été calculés avec les outils de l'analyse de la variance moléculaire (AMOVA : *Analysis of MOlecular VAriance*, Excoffier *et al.*(1992)), implémentée dans le logiciel ARLEQUIN (Schneider *et al.*, 2000), qui prend en compte les fréquences géniques et l'information moléculaire du polymorphisme considéré. L'AMOVA analyse la variance moléculaire d'une population structurée, c'est-à-dire comprenant des sous-populations d'individus séparées selon divers critères (par ex. environnementaux, linguistiques, culturels).

Les matrices de F_{st} obtenues à l'aide du logiciel ARLEQUIN ont été utilisées pour la réalisation d'analyses multidimensionnelles (MDS : *Multi Dimensional Scale*) qui permettent de visualiser de façon graphique les données des matrices. Les analyses MDS ont été réalisées à l'aide du logiciel SPSS 8.0. L'avantage de ce type d'analyse est qu'il est possible d'inclure des données sur de nombreuses populations avec un effectif important ce qui est difficile avec la méthode Network (cf *infra*). Néanmoins la représentation graphique peut présenter des

distorsions et par conséquent ne pas refléter exactement les valeurs des F_{st} . Par conséquent les graphiques MDS seront toujours interprétés en tenant compte de la matrice de F_{st} .

L'analyse phylogénétique des données obtenues pour les STR du chromosome Y et de l'ADN mitochondrial a été effectuée à l'aide du logiciel Network v4.1 (<http://www.fluxus-engineering.com>).

Considérant les erreurs pouvant survenir lors du séquençage de l'ADN mitochondrial, nous avons appliqué une méthode de vérification phylogénétique des séquences de la région HV1 en construisant un *Reduced Median Network* (Bandelt *et al.*, 1995). Cette méthode permet de mettre en évidence des incohérences dans les séquences obtenues par la présence de réticulations. Cependant il n'est possible d'utiliser que des matrices de données binaires (Bandelt *et al.*, 2002 ; Brandstätter *et al.*, 2005), certains sites polymorphes ne pourront par conséquent être pris en compte (e.g. 16232 qui peut présenter deux types de substitutions par rapport à la séquence de référence d'Anderson : soit une transition C/T soit une transversion C/A). Les positions impliquées dans les réticulations doivent être vérifiées sur les électrophorégrammes afin de s'assurer que les mutations sont bien présentes. Il faut ensuite s'assurer qu'il ne s'agit pas de hot spot de mutation, mutations fantômes ou d'erreur lors de la retranscription des données. Si nécessaire les étapes d'amplifications devront être répétées.

Pour évaluer les relations phylogénétiques entre nos sujets anciens et d'autres groupes de population nous avons utilisé la méthode dite de *Median Joining* (MJ) (Bandelt *et al.*, 1999) à la fois pour les données du chromosome Y et pour les séquences de la région HV1. La méthode MJ permet d'utiliser plusieurs types de données moléculaires en prenant en compte l'ensemble de la variabilité et pas uniquement les caractères binaires.

Cette méthode prend en compte les taux de mutations des différents loci pour les STR du chromosome Y ainsi que pour les positions polymorphes de la région HV1. Il est ainsi possible d'ordonner de manière hiérarchique les mutations en attribuant des poids différents soit en fonction des données de la littérature (Kayser *et al.*, 2000 ; Bandelt *et al.*, 2002) soit en tenant compte des fréquences de chaque position dans la population analysée. Cette méthode ne peut cependant s'appliquer que sur un nombre limité de données (au plus quelques centaines) à défaut de devenir ininterprétable. Ce type d'analyse est généralement utilisé à partir d'haplotypes provenant de différentes populations mais appartenant à un même haplogroupe afin d'évaluer les relations possibles entre ces haplotypes (e.g. Xue *et al.*, 2005 ; Pakendorf *et al.*, 2006).

IV RESULTATS

IV.1 Quantification de l'ADN nucléaire

L'ensemble des résultats obtenus pour la quantification de l'ADN nucléaire à l'aide du kit Quantifiler™ est présenté en Annexe 8. Les résultats ont été classés par type de substrat afin de visualiser les différences pouvant exister entre les types d'échantillons analysés. Le tableau 3 ci-dessous présente un récapitulatif des données de la quantification par PCR en temps réel.

	Substrats		
	Os	Dents	Cheveux
Nombre total d'échantillons	65	16	12
Nombre d'extraits quantifiés	133	18	41
Quantité moyenne d'ADN	0,23ng. μl^{-1} (0-2,32ng)	0,56ng. μl^{-1} (0,003-1,54ng)	0,12ng. μl^{-1} (0-0,945ng)
Fréquence d'extraits avec une concentration > 500pg. μl^{-1}	18,86%	50%	7,30%
Fréquence d'extraits avec une 500pg. μl^{-1} > concentration > 100pg. μl^{-1}	18,86%	38,90%	19,50%
Fréquence d'échantillons avec une concentration indéterminée	20,30%	0	19,50%
Fréquence d'échantillons présentant une inhibition	8,27%	0	0

Tableau 3 : Récapitulatif des données de la quantification par PCR en temps réel.

Concernant les échantillons osseux, nous ne disposions pas de prélèvement pour le sujet YAKa83. Pour le sujet YAKa42 les analyses ont été effectuées avant la mise en place du protocole de quantification et il ne restait plus assez de matériel osseux pour permettre la quantification à posteriori. Pour les échantillons dentaires, les dents prélevées sur le sujet YAKa55 ne satisfaisaient pas aux critères de sélection. Pour les sujets, YAKa40 et 42 une seule molaire ayant pu être sélectionnée nous n'avons pas effectué de quantification.

IV.1.1 Les substrats osseux

L'ensemble des échantillons osseux analysés pour les différents marqueurs moléculaires a fait l'objet d'une quantification absolue par PCR en temps réel. Les 133 extraits obtenus à partir

de 65 restes osseux ont été quantifiés avec le kit Quantifiler™ et ce pour les différentes extractions réalisées pour chaque sujet.

Pour environ 20% des extraits obtenus à partir d'échantillons osseux la sensibilité du kit Quantifiler™ n'a pas permis de déterminer une valeur de concentration. Les dégradations subies par les molécules d'ADN présentes dans ces échantillons ont pu donner lieu à cette absence d'amplification. De plus, la limite inférieure de détection de cette méthode étant de 23pg, ce qui correspond environ à trois copies de la séquence cible, il est probable que les variations stochastiques lors du pipetage soient également à l'origine de cette absence de détection. Les variations observées dans les résultats de la quantification dans cette gamme de concentrations montrent en outre que pour des quantités aussi faibles le système de détection du 7000 SDS se révèle peu fiable.

Il est important de considérer que pour le typage des STR autosomaux ou pour toute autre étude de l'ADN nucléaire, un nombre de copies de départ aussi réduit requiert une augmentation du nombre de cycles de PCR (Gill *et al.*, 2000, Whitaker *et al.*, 2001). Nous avons par conséquent considéré les extraits dont la concentration était inférieure à 100pg comme nécessitant ce type d'approche et une plus grande sensibilité dans la quantification n'était pas requise pour notre étude.

Les extraits d'ADN provenant d'échantillons osseux présentaient des concentrations variant de la limite de sensibilité du kit Quantifiler™ à des concentrations dépassant 1ng. μ l⁻¹ (1ng d'ADN nucléaire correspondant à environ 152 cellules diploïdes). Ces valeurs sont bien évidemment sans commune mesure avec des extraits provenant d'échantillons modernes. Cependant, ces concentrations sont plus comparables à celles rencontrées dans les études portant sur l'ADN dégradé et/ou des échantillons contenant de l'ADN en faible quantité comme par exemple dans les études de criminalistique (seuls cas pour lesquels il existe de nombreuses données portant sur la quantification absolue).

Nous avons également appliqué les protocoles recommandés pour les échantillons à faibles nombre de copies aux échantillons pour lesquels nous avons obtenus des extraits présentant des concentrations comprises entre 100 et 500pg. μ l⁻¹. Il faut en effet tenir compte du fait que la séquence cible du kit Quantifiler™ est de taille relativement réduite et inférieure à la plupart des marqueurs amplifiés par la suite.

Près de 19% de nos échantillons ont permis d'obtenir des concentrations supérieures à 0,5ng. μ l⁻¹ ce qui correspond à un nombre de copies comprises entre 76 et 152. Cette quantité est généralement celle préconisée pour l'analyse des STR autosomaux avec les kits d'amplification actuels sans qu'une augmentation du nombre de cycle de PCR soit nécessaire.

Pour moins de 10% des échantillons les valeurs de Ct des IPC indiquaient la présence d'inhibiteurs. Pour ces extraits, nous avons réalisé des gammes de dilutions en cascade afin de déterminer à quel degré de dilution il était possible d'obtenir une amplification sans réduire trop fortement la quantité d'ADN de départ.

Pour un même échantillon nous avons constaté que la quantité d'ADN pouvait varier du simple au double (YAKa16, YAKa20, YAKa36, YAKa37, etc.) à une relation de facteur 10 (YAKa22, YAKa46, YAKa51, YAKa59, YAKa64). Dans certains cas, la concentration n'a pu être déterminée pour un extrait alors qu'un résultat était obtenu lors de la quantification d'un autre extrait provenant du même individu (YAKa23, YAKa39, YAKa41, YAKa45, etc.). Il faut toutefois considérer que dans ces cas les concentrations étaient généralement faibles, de l'ordre de la dizaine de picogrammes. Dans ces gammes de concentrations, le pipetage ainsi que la détection peuvent être affectés expliquant ainsi ces variations.

Les échantillons osseux ont été classés selon la tranche d'âge et le type de prélèvement : (i) enfants de 0 à 2 ans (19 extraits) ; (ii) enfants de plus de 2 ans (16 extraits) ; (iii) os de petites tailles (os du tarse, du carpe, talus), rotules et fragments d'os longs de sujets adultes et soumis à une pulvérisation par cryobroyage (22 extraits) ; (iv) diaphyses d'os longs de sujets adultes pulvérisés par trépanation (77 extraits) (cf Figure 34).

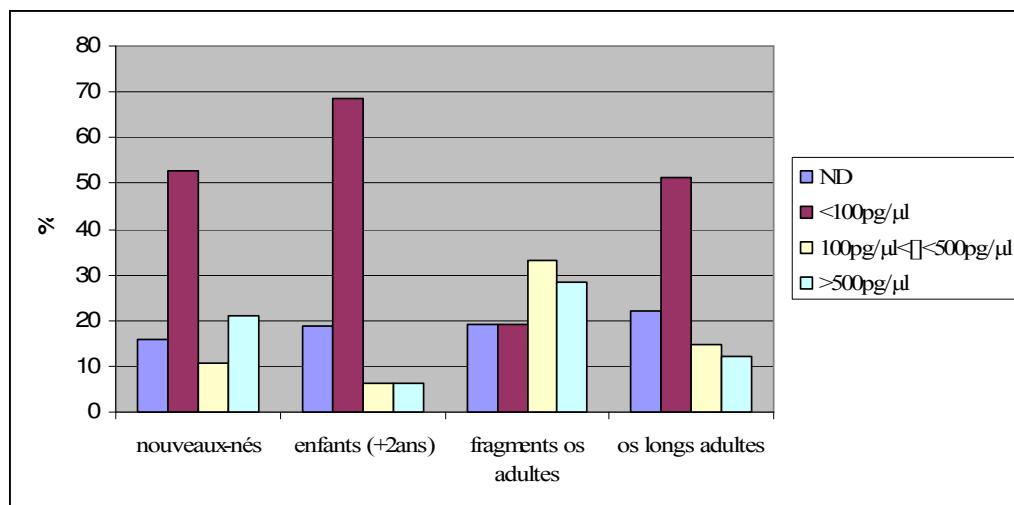


Figure 34 : Proportions pour chaque ordre de grandeur de concentration mesuré avec le kit Quantifiler™ observées pour les différents types d'échantillons osseux. ND : non déterminé ; <100pg/µl : concentration inférieure à 100pg.µl⁻¹ ; 100pg/µl < [] < 500pg/µl : concentration comprise entre 100 et 500pg.µl⁻¹ ; >500pg/µl : concentration supérieure à 500 pg.µl⁻¹.

Pour les différents types d'échantillons, environ 19% des extraits ne contient pas une quantité suffisante pour être détectée par la méthode employée, ou ne renferme plus d'ADN.

Les extraits dont la concentration est inférieure à $100\text{pg}.\mu\text{l}^{-1}$ sont les plus fréquents pour les différents types de prélèvement excepté pour la catégorie des fragments osseux provenant de sujets adultes traités par cryobroyage. Par ailleurs, ce type d'échantillons a permis d'obtenir les extraits les plus concentrés et la gamme de concentration $0,1/0,5\text{ng}.\mu\text{l}^{-1}$ représente la catégorie majoritaire. Cette catégorie d'échantillon a également présenté la plus forte proportion d'extraits de concentration supérieure à $0,5\text{ng}.\mu\text{l}^{-1}$.

Toutefois, il faut garder à l'esprit que les conditions taphonomiques jouent un rôle primordial dans la conservation de l'ADN et qu'elles vont influer sur tous les types de prélèvements.

IV.1.2 Utilisation du PTB

Pour 13 échantillons nous avons utilisé du Bromure de N-Phényl Thiazolium dans le tampon de lyse afin de diminuer l'importance des complexes de Maillard qui peuvent occasionner des échecs lors de l'amplification.

Les résultats de la quantification n'ont pas montré une augmentation nette des rendements au niveau de la quantité d'ADN obtenue (Tableau 4). Pour 4 des échantillons testés (YAKa 36, YAKa 37, YAKa 47, YAKa 72), les concentrations d'ADN dans les extraits se sont révélées inférieures à celles observées pour les extractions sans PTB. Cependant pour deux des échantillons les concentrations des deux types d'extraits se trouvent dans la gamme de concentrations pour lesquelles la méthode de détection employée se révèle peu fiable.

Pour 7 échantillons (YAKa35, YAKa68, YAKa74, YAKa75, YAKa76, YAKa80, YAKa81), les concentrations obtenues étaient équivalentes à celles observées avec le protocole d'extraction sans PTB. Mais comme précédemment, la sensibilité de la méthode ne permet pas d'établir clairement une distinction.

Enfin, deux extraits obtenus avec PTB ont montré des concentrations en ADN plus fortes avec le protocole d'extraction intégrant le PTB. Pour ces deux sujets (YAKa55 et YAKa58) trois extractions sans PTB avaient été réalisées et pour respectivement 1 et 2 extractions il n'avait pas été possible de détecter de l'ADN lors de la réaction de PCR en temps réel. Mais

là encore la faible concentration des échantillons peut induire des erreurs dans la détermination exacte de la concentration.

Cependant, il est important de considérer la qualité des résultats qui sera obtenue pour les typages subséquents et non pas simplement la concentration de l'extrait. En effet, nous avons constaté que la quantité d'ADN présente dans un extrait donnée ne reflétait pas toujours l'état de dégradation des acides nucléiques (cf chapitre IV.3.2.3 Influence du PTB).

n° Laboratoire	Ct	Quantité
YAKa36-E1	28,8	1,18
IPC	37,78	
YAKa36-E2	30,53	3,92E-01
IPC	ND	
YAKa36-PTB	31,12	6,35E-02
IPC	ND	
YAKa37-E1	28,32	1,66
IPC	ND	
YAKa37-E2	29,85	0,635
IPC	ND	
YAKa37-PTB	29,14	2,70E-01
IPC	ND	
YAKa47-E2	34,57	4,82E-02
IPC	27,25	
YAKa47-E3	32,25	3,90E-02
IPC	27,82	
YAKa47-PTB	33,9	1,35E-02
IPC	25,94	
YAKa72-E1	35,93	5,90E-03
IPC	27,89	0
YAKa72-E2	35,68	2,39E-03
IPC	27,73	
YAKa72-PTB	ND	ND
IPC	26,12	

Tableau 4a

n° Laboratoire	Ct	Quantité
YAKa35-E1	34,5	1,87E-02
IPC	30,84	
YAKa35-E2	34,08	3,11E-02
IPC	31,31	
YAKa35-PTB	32,66	2,08E-02
IPC	29,24	
YAKa68-E1	28,4	6,50E-01
IPC	31,61	
YAKa68-E2	30,47	8,65E-02
IPC	26,47	
YAKa68-PTB	31,08	1,01E-01
IPC	26,98	
YAKa74-E1	32,55	6,20E-02
IPC	26,18	0
YAKa74-E2	33,41	1,53E-02
IPC	26,53	
YAKa74-PTB	33,08	2,42E-02
IPC	26,31	
YAKa75-E1	33	2,14E-02
IPC	26	
YAKa75-PTB	33,9	1,34E-02
IPC	26,04	
YAKa76-E1	35,53	2,71E-03
IPC	26,38	
YAKa76-PTB	37,05	1,42E-03
IPC	25,98	
YAKa80-E1	31,78	5,75E-02
IPC	26,35	
YAKa80-PTB	32,37	4,01E-02
IPC	25,86	
YAKa81-E1	ND	ND
IPC	26,39	
YAKa81-PTB	ND	ND
IPC	25,98	

Tableau 4b

n° Laboratoire	Ct	Quantité
YAKa55-E1	36	2,51E-03
IPC	27,09	
YAKa55-E2	ND	ND
IPC	27,01	
YAKa55-E3	ND	ND
IPC	25,23	
YAKa55-PTB	36,81	1,68E-03
IPC	25,98	
YAKa58	35,98	2,56E-03
IPC	26,96	
YAKa58	34,96	2,07E-03
IPC	25,11	
YAKa58-E3	ND	ND
IPC	27,07	
YAKa58-PTB	32,2	4,10E-02
IPC	26,15	

Tableau 4c

Tableau 4 : Résultats de la quantification obtenus pour les sujets testés avec le protocole d'extraction contenant du PTB dans le tampon de lyse. Les échantillons sont répartis en 3 groupes suivant les résultats obtenus avec le PTB : quantité d'ADN mesurée plus faible (tableau 4a), quantité équivalente (tableau 4b) et augmentation du rendement (tableau 4c).

IV.1.3 Les dents

DENTS				OS		
N° Laboratoire	Dents disponibles	Ct	Quantité	N° Laboratoire	Ct	Quantité
YAKa34-E1	2 PM, 8 I+C	28,65	0,925	YAKa34-E1	30,93	2,50E-01
IPC		26,04		IPC	29,79	
YAKa34-E2		29,6	5,700E-01	YAKa34-E2	33,86	3,62E-02
IPC		29,62		IPC	29,69	
YAKa37	2 M très usées, 4 I	29,66	5,450E-01	YAKa37-E1	28,32	1,66
IPC		29,39		IPC	ND	
YAKa39	1 M	28,67	0,645	YAKa37-E2	29,85	0,635
IPC		26,21		IPC	ND	
YAKa41	1 M	31,81	1,100E-01	YAKa39-E1	31,84	0,054
IPC		29,31		IPC	24,63	
YAKa64	2M	31,91	1,025E-01	YAKa41-E1	34,64	4,60E-02
IPC		29,54		IPC	27,26	
YAKa65	1M+2I	29,42	0,54	YAKa41-E2	ND	ND
IPC		26,02		IPC	26,86	
YAKa66	3M	27,59	1,19E+00	YAKa64-E2	34,63	4,64E-02
IPC		26,89		IPC	27	
YAKa67	1 M	29,41	6,600E-01	YAKa64-E3	33,41	2,12E-02
IPC		29,35		IPC	26,86	
YAKa69	2M	31,5	6,75E-02	YAKa64-E1	35,18	9,000E-03
IPC		27,1		IPC	29,71	
YAKa78	M+I	28,01	1,44	YAKa65-E1	34,12	1,26E-02
IPC		26,02		IPC	27,12	
YAKa79	2M	29,99	4,270E-01	YAKa65-E2	32,52	4,06E-02
IPC		29,29		IPC	27,82	
YAKa80	2M	30,53	2,52E-01	YAKa66	28,47	6,25E-01
IPC		25,75		IPC	27,52	
YAKa81	2M+1C	31,59	1,20E-01	YAKa67	34,33	1,695E-02
IPC		25,82		IPC	29,43	
YAKa82	II	33,57	3,04E-02	YAKa69	28,25	7,30E-01
IPC		25,93		IPC	ND	
YAKa83-E1	3I	30,15	3,27E-01	YAKa78	32,27	3,89E-02
IPC		25,73		IPC	26,25	
YAKa83-E2		28,9	0,78	YAKa79DIL	33,17	4,020E-02
IPC		25,89		IPC	30,9	
YAKa86	3 M (petites)	31,09	1,70E-01	YAKa80-E1	31,78	5,75E-02
IPC		26,14		IPC	26,35	
				YAKa80-PTB	32,37	4,01E-02
				IPC	25,86	
				YAKa81-E1	ND	ND
				IPC	26,39	
				YAKa81-PTB	ND	ND
				IPC	25,98	
				YAKa82	31,34	8,30E-02
				IPC	26,35	
				YAKa86	31,85	5,50E-02
				IPC	25,94	

Tableau 5 : comparaison des quantifications des extraits obtenus à partir de dents et d'os. M : molaire, PM : prémolaire, I : incisive.

Nous disposions de dents pour 16 des sujets étudiés. Les volumes d'extraits obtenus ne nous ont pas permis de dupliquer l'étape de quantification pour chacun des extraits par conséquent seuls 18 extraits ont été quantifiés (Tableau 5).

Les résultats montrent que les extraits obtenus à partir des dents présentent des concentrations d'ADN supérieures à celles obtenues à partir des os pour des quantités de poudre parfois inférieures. Seul un extrait (YAKa82) a montré une concentration inférieure à $100\text{pg}.\mu\text{l}^{-1}$.

Ainsi pour 11 (YAKa34, YAKa39, YAKa41, YAKa64, YAKa65, YAKa67, YAKa78, YAKa79, YAKa80, YAKa81, YAKa86) des 15 échantillons testés, les quantités d'ADN obtenues à partir de dents étaient significativement supérieures à celles obtenues à partir d'os. Pour les sujets YAKa41 et YAKa81, les extractions à partir de poudre d'os n'avaient pas permis d'obtenir des quantités d'ADN détectables alors que les extraits obtenus à partir de dents ont montré des concentrations supérieures à $100\text{pg}.\mu\text{l}^{-1}$.

De plus, pour 2 des échantillons (YAKa37 et YAKa69), une inhibition a été observée lors de la quantification des extraits d'ADN à partir d'os. La présence d'inhibiteur n'a, en outre, été détectée dans aucun des extraits réalisés à partir de dents.

IV.1.4 Les cheveux

Chacun des extraits soumis ou non à la procédure de lavage, a été analysé avec le kit QuantifilerTM. Ainsi, 41 extraits obtenus à partir de cheveux ont été quantifiés.

Dans un premier temps, il est nécessaire de préciser que nous n'avons pas observé de nette différence entre les extraits obtenus à partir de cheveux lavés ou non (Annexe 3).

Pour 4 individus (YAKa37, YAKa45, YAKa55, YAKa58) les concentrations observées étaient inférieures à $10\text{pg}.\mu\text{l}^{-1}$ ce qui correspond à une copie d'ADN et pour certaines extractions aucune amplification n'était observée. Il est fort possible que ces extraits ne contenaient pas d'ADN et que les résultats proviennent d'amplifications de débris cellulaire ou de chromosomes isolés (Taberlet *et al.*, 1996 ; Gill *et al.*, 2000 ; Kloosterman *et al.*, 2003).

Les cheveux collectés sur 3 sujets (YAKa34, YAKa39, YAKa66) ont permis d'obtenir des concentrations supérieures à $100\text{pg}.\mu\text{l}^{-1}$ et entre 0,5 et $1\text{ng}.\mu\text{l}^{-1}$ pour 3 des extraits du sujet YAKa66.

La comparaison des valeurs moyennes de quantification des les échantillons pour lesquels nous disposions de cheveux (Figure 35) montre qu'il n'existe pas de relation entre la

conservation des acides nucléiques dans les os ou les dents et les concentrations en ADN dans les extraits d'ADN obtenus à partir de cheveux. Pour le sujet YAKa37 les concentrations dans les os et les dents étaient supérieures à $0,5\text{ng}.\mu\text{l}^{-1}$ alors que l'extraction à partir des cheveux n'a pas permis d'obtenir une quantité détectable d'ADN. En revanche, pour le sujet YAKa39 les concentrations obtenues à partir de l'extraction des cheveux se sont révélées supérieures à celles obtenues à partir des os.

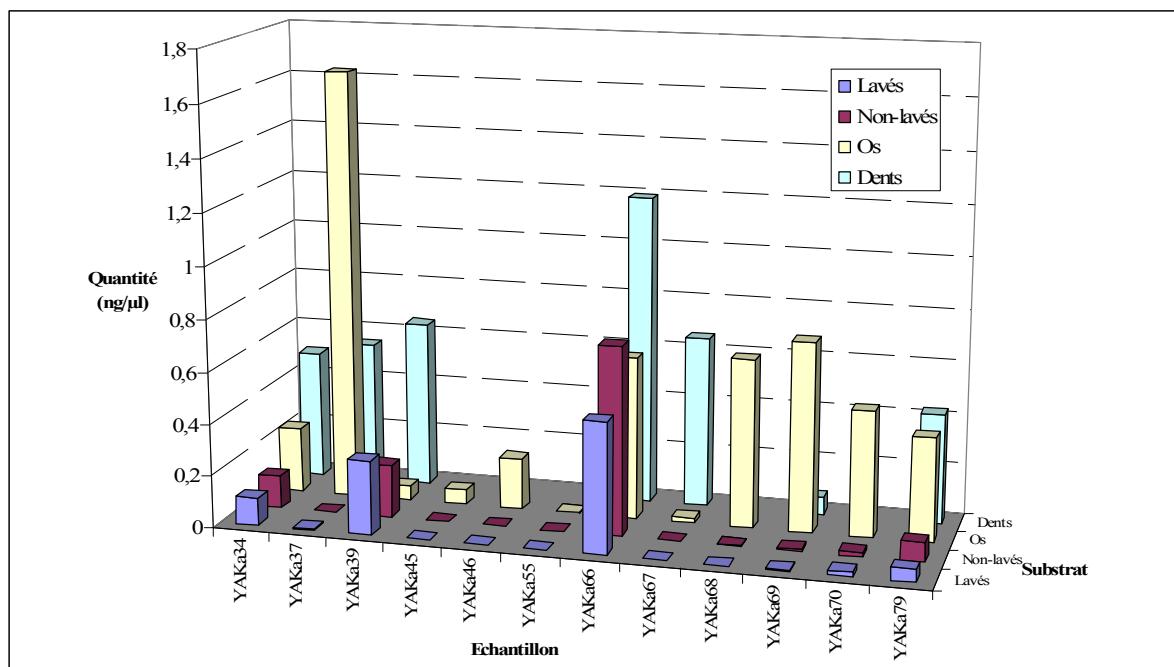


Figure 35 : Comparaison des valeurs moyennes de quantifications obtenues pour les différents types de substrats.

Aucun extrait n'a révélé la présence d'inhibiteurs car les Ct des différents IPC ont tous dépassé le seuil de détection. Les précautions prises quant à la quantité de cheveux utilisée se sont donc révélées efficaces pour éviter la présence d'un excès de mélanine pouvant entraîner une inhibition.

IV.2 Amplification totale du génome: kit Genomiphi™

Le kit Genomiphi™ a été testé sur 7 échantillons anciens sélectionnés suivant un gradient de concentration de $0,75\text{ng}.\mu\text{l}^{-1}$ à $2,5\text{ng}.\mu\text{l}^{-1}$ (cf Tableau 6) et pour lesquels des profils STR complets avaient été obtenus à partir des extraits originaux. Le but de ce test était d'établir dans quelle mesure ce type d'amplification était applicable à des échantillons anciens, de déterminer quelle quantité d'ADN il était possible d'obtenir après application de ce protocole et d'estimer la fiabilité des résultats obtenus par typage STR après l'amplification.

NOM	Quantité d'ADN ($\text{ng}.\mu\text{l}^{-1}$)
YAKa 33	0.765
YAKa 21	0.97
YAKa 31	1.19
YAKa 32	1.27
YAKa 18	1.28
YAKa 20	1.43
YAKa 19	2.61

Tableau 6 : Concentrations respectives des différents échantillons anciens utilisés pour le test Genomiphi™ (Amersham). Quantification réalisée grâce au kit Quantifiler™ (Applied Biosystems) sur 7000SDS.

Les produits d'amplification obtenus avec le kit Genomiphi™ ont été déposés sur gel d'agarose 1% afin d'évaluer la taille des produits obtenus. La figure 36 représente la photo du gel d'agarose sur lequel ont été déposés les produits d'amplification après purification sur microcons YM30.

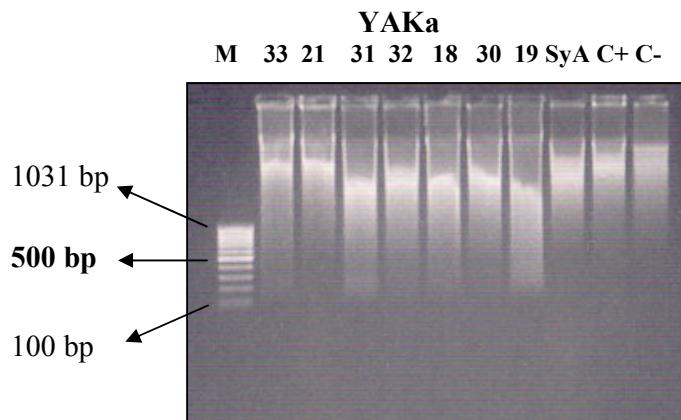


Figure 36 : Contrôle des produits d'amplification Genomiphi par dépôt sur gel d'électrophorèse à 1% d'agarose. Les dépôts correspondent à 6µl de produit d'amplification+2µl d'H₂O+2µl de bleu.

L'ensemble des échantillons montre le même type de résultats indépendamment de la quantité d'ADN génomique de départ. Il semble que l'amplification par la polymérase Phi29 ait permis de générer de l'ADN de haut poids moléculaire pour tous les échantillons. Le témoin négatif présente également une amplification. Cette amplification aspécifique vient de l'hybridation et de l'amplification de dimères d'amorces.

Les résultats de la quantification à l'aide du kit Quantifiler™ ont montré des rendements négatifs puisque la quantité mesurée après amplification était inférieure à celle mesurée avant le traitement avec Genomiphi™.

L'augmentation réelle de la quantité d'ADN n'a pas pu être appréciée, cependant les résultats du typage STR ont tout de même permis d'estimer l'applicabilité de la méthode à des échantillons anciens. L'annexe 9 présente les résultats obtenus et leur comparaison avec les profils établis à partir d'extraits non soumis au protocole Genomiphi. Les pertes d'allèles mentionnées dans différentes publications (Paulson *et al.*, 1999 ; Holbrook *et al.*, 2005 ; Sun *et al.*, 2005 ; Ballantyne *et al.*, 2006) ont également été observées sur nos échantillons. Des absences d'amplifications pour la plupart des loci ont également été constatées et dans certains cas nous avons observé l'amplification de faux allèles.

IV.3 Les STR autosomaux

IV.3.1 Résultats généraux

L'ensemble des individus prélevés a été typé pour les STR autosomaux à l'aide de deux kits d'amplification ; le kit AMPflSTR® Profiler Plus™ et le kit AMPflSTR® Identifiler™ (Applied Biosystems). Les profils consensus sont présentés dans le tableau 7 ci-dessous. Ces profils ont été déterminés à partir des amplifications réalisées sur les extraits provenant des différents substrats analysés pour chaque sujet et des multiples extractions réalisées pour chacun d'entre eux.

Les analyses effectuées ont permis d'obtenir des profils consensus complets, ou ne présentant qu'un seul locus non amplifié, pour 42 des sujets analysés, soit 63,6%. La comparaison des ces différents profils avec ceux des membres de l'équipe de terrain et du laboratoire exclut toute contamination concernant ce marqueur puisque aucune correspondance n'a été observée. Pour 8 sujets (YAKa35, YAKa43, YAKa48, YAKa51, YAKa52, YAKa53, YAKa60, YAKa84),

il n'a pas été possible d'établir de génotypes consensus à partir des différentes amplifications et la détermination du sexe à partir du locus de l'amélogénine n'a été possible que pour le sujet YAKa53. Les quantifications des extraits de ces différents sujets montrent des concentrations inférieures à $50\text{pg}.\mu\text{l}^{-1}$ pour YAKa35, YAKa51, YAKa52 et il n'a pas été possible de déterminer la concentration des autres échantillons.

n° Labo	Nom	Amel	CSFPO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	THO1	TPOX	vWA
YAKa15	ouhorai	XY	-	9-11	-	14-17	-	29-29	-	16-16	12-12	10-11	13-15	20-27	-	-	16-19
YAKa16	TD 1	XX	-	9-12	-	14-16	-	30-32	-	15-16	11-13	8-12	13-13	23-23	-	-	18-19
YAKa17	no 10	XY	-	8-12	-	15-19	-	29-30	-	15-15	11-12	10-11	13-16	23-24	-	-	18-18
YAKa18	KM 1	XX	-	9-11	-	15-18	-	29-30	-	14-17	11-11	8-10	10-13	24-24	-	-	16-17
YAKa19	no 9	XY	-	11-12	-	14-15	-	29-34,2	-	15-15	12-12	8-9	13-13	23-24	-	-	14-14
YAKa20	no 11	XX	-	8-13	-	14-15	-	30-31	-	15-16	11-13	8-11	13-13	23-24	-	-	18-19
YAKa21	TA 1	XY	-	8-12	-	14-14	-	30-30	-	15-15	12-12	8-11	13-13	21-24	-	-	17-17
YAKa22	12	XX	-	11-13	-	14-14	-	29-30	-	15-15	11-12	8-8	10-11	-	-	-	13-18
YAKa23	no 5	XY	-	-	-	-	-	-	-	15-17	11-11	-	12-13	23-23	-	-	16-19
YAKa24	no 6	XY	-	12-12	-	-	-	29-29	-	16-17	12-13	9-9	10-13	24-25	-	-	16-18
YAKa25	no 4	XY	-	9-12	-	13-17	-	29-29	-	14-15	11-12	10-13	13-13	22-23	-	-	17-18
YAKa26	BKH	XY	-	10-11	-	11-15	-	29-30	-	16-17	11-11	8-9	13-13	23-23	-	-	17-18
YAKa27	3	XY	-	8-10	-	14-14	-	29-31	-	15-15	10-10	-	13-14	22-23	-	-	17-19
YAKa28	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa29	STR1 a	XY	-	8-13	-	13-19	-	31-32	-	15-16	11-12	8-9	10-13	21-22	-	-	17-19
YAKa30	STR1 b	XX	-	12-13	-	-	-	29-29	-	15-15	13-13	-	-	19-19	-	-	14-18
YAKa31	AE 2	XY	-	8-14	-	15-18	-	29-29	-	16-17	7-10	8-8	10-16	22-23	-	-	16-19
YAKa32	SRT 2	XX	-	11-11	-	15-18	-	28-30	-	15-17	9-11	9-10	12-13	22-24	-	-	14-20
YAKa33	KM 2	XX	-	10-13	-	17-19	-	30-32,2	-	16-17	12-12	11-13	13-13	23-24	-	-	14-18
YAKa34	Arbre Chamanique ind1	XY	-	10-11	9-13	13-15	14-17,2	29-29	17-23	15-16	10-12	8-8	13-16	21-25	8-9,3	11-11	17-19
YAKa35	Arbre Chamanique ind4	XX	-	11	-	-	-	-	-	-	-	-	-	18	-	-	-
YAKa36	Arbre Chamanique ind2	XX	12-14	10-11	9-13	13-14	13-15	29-29	23-25	16-16	12-13	8-8	13-16	21-25	6-8	8-8	19-19
YAKa37	Arbre Chamanique ind3	XX	12-12	8-10	9-13	13-13	14-15	29-30	17-23	15-16	10-13	8-8	10-16	20-25	6-9,3	8-11	17-19
YAKa38	Arbre Chamanique ind5	XY	-	11	-	13-14	-	29	-	15-16	10-12	8-10	15-16	21-22	-	-	17-19
YAKa39	Munur Urekhl	XY	-	8-13	-	13-17	-	29-30	-	-	11-12	8-8	13-14	19-24	-	-	14-16
YAKa40	Doussoulen1	XY	-	12-13	-	13-13	-	30-31	-	15-16	11-12	10-12	13-13	21-22	-	-	16-18
YAKa41	Arbre Chamanique 2	XY	-	11-14	9-9	14-15	13-13	29-31	23-23	16-17	11-13	8-9	13-16	22-25	8-(9)	8-11	17-19
YAKa42	Koulousoun Nakh 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa43	Fedoseva Allalaika	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa44	Fedosseva Tchersinsky	XY	-	8-8	-	-	13-14,2	29-29	-	-	11-11	12-12	13-13	-	9-9,3	8-8	17-17
YAKa45	Arbre Chamanique 3	XX	11-12	8-12	9-13	14-17	14-14,2	31-32	19-20	14-15	11-12	8-9	10-13	23-24	7-7	9-10	16-19
YAKa46	Munur Urekhl bis	XX	9-10	10-11	11-13	13-17	15-16,2	29-32,2	18-22	15-15	11-12	12-12	13-13	20-23	7-8	8-8	14-16
YAKa47	Musee ethno	XY	12-12	8-10	9-12	-	13-14	29-30	20-20	15-16	11-12	12-12	13-13	22-22	9,3-9,3	8-11	16-19
YAKa48	Fedoseva Diring Urekhl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa49	Jardin Botanique	XY	-	-	-	-	-	-	-	-	-	-	13-13	-	-	-	-

n° Labo	Nom	Amel	CSFPO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	THO1	TPOX	vWA
YAKa50	Jarama 1	XX	11-12	8-9	9-9	14-19	13-16	30-32	17-17	15-17	11-13	8-12	13-13	-	7-9	9-12	14-14
YAKa51	Jarama 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa52	Jarama 3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa53	Jarama 4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa54	Toumousaktakh 1	XX	-	-	9-12	-	13-13.2	30-30	17-19	-	11-12	-	13-13	-	9-9.3	8-11	16-18
YAKa55	Koulousoun Nakh 1	XX	-	9-11	-	14-14	13-13	30-32.2	-	15-16	11-13	-	11-13	-	8-8	11-11	-
YAKa56	Dirilaa Sayliga 1	XY	13-13	11-12	13-14	17-18	13-13	29-30	18-19	14-15	13-13	8-12	11-13	21-21	6-7	11-13	16-19
YAKa57	Bekh Alaas 1	XY	11-11	9-11	9-12	14-19	13-15.2	29-30	24-24	15-16	7-13	11-12	10-12	24-24	7-8	9-11	14-18
YAKa58	Bekh Alaas 2	XY	11-11	10-10	-	-	13-13	30-30	-	15-15	-	13-17	24-24	-	11-12	17-19	
YAKa59	Bekh Alaas 3	XY	-	-	-	-	14-14.2	-	-	15-16	-	-	-	-	-	9-9	-
YAKa60	Bekh Alaas 4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa64	Sette Toumoul	XY	10-11	11-13	9-9	13-15	13-13	29-30	18-25	15-15	11-11	8-12	13-13	24-24	7-9	11-11	14-19
YAKa65	Oulakhans Alas 1	XY	10-11	11-11	10-13	14-15	13-14	30-31.2	22-24	15-16	11-13	10-12	10-13	21-23	7-9.3	8-11	16-16
YAKa66	Batta Tcharana	XY	11-12	11-11	9-9	13-13	13.2-15	29-31	18-18	16-17	12-13	10-11	11-16	20-25	9.3-9.3	8-8	17-19
YAKa69	Kous Tcharbyt	XY	11-12	8-10	9-9	14-20	13-13.2	30-30	17-19	15-17	10-13	8-10	13-16	24-24	9-9	8-9	14-18
YAKa70	Seden	XY	9-11	9-11	9-12	13-15	15-16	29-30	19-24	15-16	11-13	10-10	13-16	20-29	9.3-9.3	8-11	18-19
YAKa67	Ken Ebe 3	XY	10-12	9-11	12-12	15-16	12-17.2	29-30	23-24	15-15	11-11	11-11	10-13	23-23	9-9.3	8-11	14-16
YAKa68	Ken Ebe 2	XY	12-12	-	9-11	-	13-15;2	29-30	17-19	15-16	11-12	-	10-13	23-24	6-7	8-9	14-19
YAKa71	Ken Ebe 4	XY	11-14	9-12	9-11	13-14	13-13.2	30-31	17-23	16-16	11-12	8-11	13-14	22-24	7-7	9-11	16-19
YAKa72	Ken Ebe 5	XX	-	-	-	-	13-14	-	-	16-17	11-12	-	12-13	24-24	-	8-11	14-14
YAKa73	Ken Ebe 6	XY	-	10-11	11-12	-	13-13	30-30	-	-	-	-	10-15	22-24	7-8	11-11	16-16
YAKa74	Ken Ebe 7	XY	-	-	9-12	-	13.2-15	29-31	19-20	15-16	11-11	-	13-13	-	-	8-8	17-19
YAKa75	Ken Ebe 8	XX	-	-	9-12	-	13-14	30-30	-	15-16	7-13	-	11-13	-	7-9.3	8-8	16-17
YAKa76	Ken Ebe 9	XX	10-10	8-11	12-13	18-18	13-13	29-30	19-19	16-16	11-13	8-12	11-11	21-21	-	8-8	16-19
YAKa78	bouogaryma 1	XY	10-12	11-12	12-13	13-21	13.2-13.2	29-29	19-25	15-15	12-13	8-11	13-13	20-23	6-7	8-8	14-17
YAKa79	bouogaryma 2	XX	10-10	8-11	9-10	13-14	14-14.2	29-30	18-20	16-17	12-12	8-11	10-10	20-24	6-9.3	8-11	16-18
YAKa80	Nelegher	XY	11-11	9-11	11-12	13-14	13.2-15.2	29-30	19-24	16-17	11-11	8-8	10-14	23-24	9-9.3	9-9	19-19
YAKa81	Orto Aryy	XY	10-12	9-11	10-12	15-15	13-13	29-31.2	19-22	16-16	11-13	10-12	10-13	21-24	9-9.3	8-8	16-18
YAKa82	Okhtobout 2	XX	10-11	11-12	9-13	14-15	15.2-17.2	29-30	23-24	15-16	12-12	8-11	10-13	20-21	6-9	8-11	18-18
YAKa83	Kyys Ounouoga	XX	10-11	11-12	9-10	15-17	13-14	29-30	18-25	16-16	10-13	8-10	13-16	24-24	8-9	8-8	17-19
YAKa84	Kyys Ounouoga	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa86	Tyyt Bappyt	XY	12-12	10-12	10-11	14-22	15-15.2	29-30	18-23	15-15	11-13	10-10	13-13	21-24	9-9.3	8-11	17-18

Tableau 7 : Profils consensus obtenus à partir de l'ensemble des substrats. Les loci dans les cases grisées correspondent aux marqueurs supplémentaires du kit Identifier™ par rapport au kit Profiler Plus™. Les échantillons dont les noms de laboratoire sont grisés ont été typés avec Profiler Plus™.

IV.3.2 Efficacité de typage

Nous avons calculé les efficacités de typages des réactions en évaluant cinq critères par comptage sur l'ensemble des profils obtenus : (i) l'absence d'amplification ; (ii) un génotype correspondant parfaitement au génotype consensus ; (iii) la présence d'un allèle supplémentaire ; (iv) la perte d'un allèle par rapport au génotype connu et enfin (v) le typage d'un allèle absent dans le génotype consensus.

Cette efficacité de typage a été évaluée pour les deux kits d'amplification utilisés pour le typage à partir d'ADN extrait d'échantillon osseux, pour les résultats obtenus avec le kit Identifier™ à partir de substrats osseux et dentaire et enfin pour les échantillons dont l'ADN a été extrait avec le tampon de lyse contenant du PTB.

IV.3.2.1 Comparaison des kits utilisés sur les substrats osseux

Le kit AMPflSTR® Profiler Plus™ n'ayant été utilisé que sur des échantillons osseux nous n'avons pu comparer son efficacité avec le kit AMPflSTR® Identifier™ que pour ce type de substrat. Les comparaisons ont bien entendu porté sur les marqueurs présents à la fois dans les deux kits et nous avons décidé de ne pas tenir compte des résultats du locus de l'amélogénine qui est généralement amplifié (sauf cas cités ci-dessus).

Les résultats sont présentés dans les diagrammes ci-dessous ; les marqueurs sont classés par ordre croissant de taille des produits de PCR (Figure 37a et b).

Pour le kit Profiler Plus™ nous nous sommes basés sur 64 amplifications. Force est de constater que pour le kit Profiler Plus™ la proportion de génotypes complets reste dans une gamme comprise entre 85% et 75% ce qui correspond à des amplicons de 250pb au maximum. La proportion d'absence d'amplification augmente de manière notable pour les marqueurs D7S820 et D18S51 dont les allèles de plus grande taille sont comparables à ceux de FGA mais dont la taille des allèles les plus petits est supérieure.

La fréquence de perte d'allèles est inférieure à 10% pour tous les marqueurs et la présence d'allèle supplémentaire est inférieure à 5% excepté pour le locus vWA. La proportion de typage erroné est également inférieure à 10% pour tous les marqueurs. Les résultats obtenus pour les différents artefacts d'amplifications, à l'exception de l'absence d'amplification, ne permettent pas d'établir une relation claire entre leur fréquence et l'augmentation de la taille des loci pour ce kit.

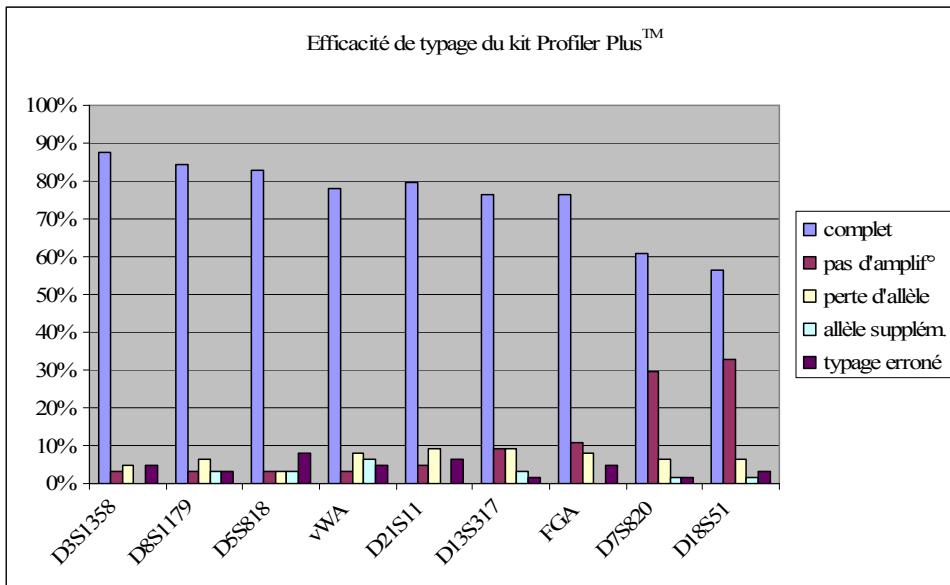


Figure 37a: Efficacité de typage du kit Profiler Plus™ évaluée sur 64 amplifications réalisées à partir d'échantillons osseux.

Pour le kit Identifiler™, nous nous sommes basés sur 168 amplifications. Comme pour le kit Profiler Plus™, la fréquence de génotypes complets diminue à mesure que la taille des marqueurs augmente. Cependant, pour le marqueur D13S317 la proportion d'absence d'amplification est plus élevée que celle observée pour FGA qui est pourtant de taille plus importante. Les marqueurs D7S820 et D18S51 présentent tous deux les proportions les plus importantes d'absence d'amplification.

La fréquences des pertes d'allèles est supérieure à 10% pour les loci vWA et D13S317 ; pour les autres marqueurs ces proportions varient entre 4% et 8%. Il semble que les pertes d'allèles soient légèrement plus importantes avec ce kit qu'avec le kit Profiler Plus. Les autres artefacts d'amplification sont néanmoins moins fréquents que pour Profiler Plus.

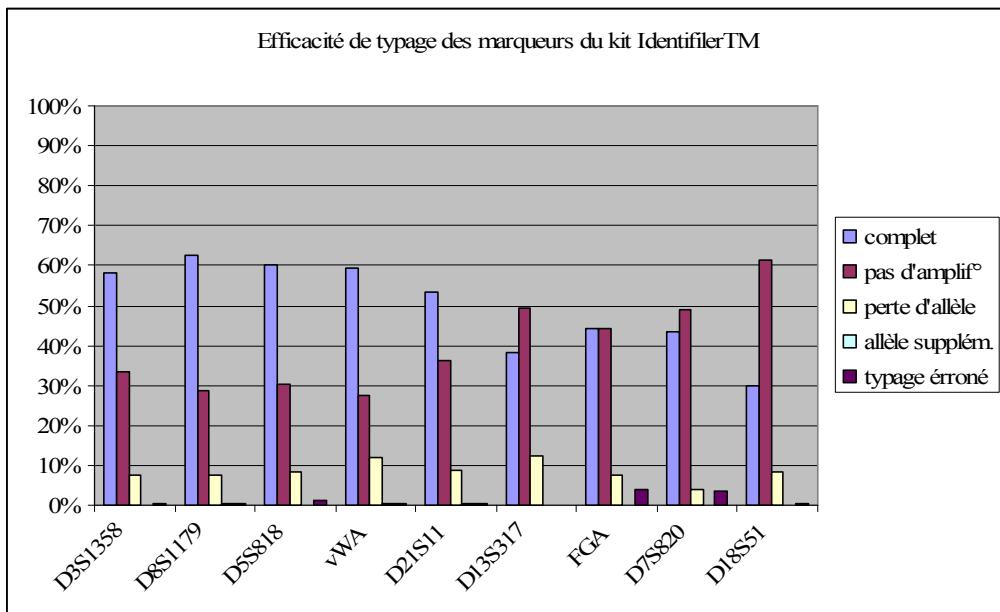


Figure 37b: Efficacité de typage du kit Identifiler™ évaluée sur 168 amplifications réalisées à partir d'échantillons osseux. Seules les données des marqueurs présents dans le kit Profiler™ Plus sont présentées.

IV.3.2.2 Comparaison des substrats analysés

Les deux diagrammes ci-dessous (Figures 38a et b) présentent les résultats de l'évaluation de l'efficacité de typage du kit Identifiler sur des extraits obtenus à partir d'échantillons osseux et dentaires.

Pour les échantillons osseux, nous avons repris les données des 168 amplifications précédentes en y incluant l'ensemble des marqueurs du kit. La même tendance se dégage que précédemment c'est à dire que les échecs d'amplifications sont plus fréquents pour les marqueurs de grande taille. Pour les loci dont les produits d'amplification approchent les 300pb, les pourcentages d'échecs sont supérieurs à 50% sauf pour D2S1338 (44%). Cependant pour ces marqueurs de haut poids moléculaire, la corrélation entre taille des amplicons et efficacité de typage est moins nette.

Nous avons observé que les pertes d'allèles étaient toutes inférieures à 10% sauf pour les deux marqueurs identifiés précédemment (vWA, D13S317). La fréquence de typage d'allèles supplémentaires ou de faux allèles correspond à 2% excepté pour FGA (3,57%) et D2S1338 (4,17%).

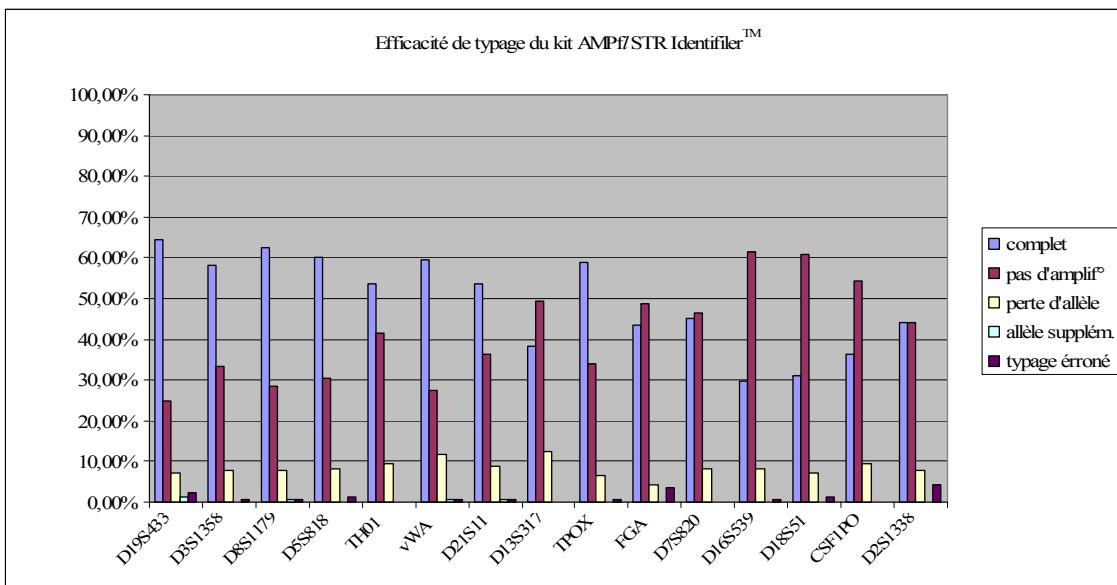


Figure 38a: Efficacité de typage du kit Identifier™ évaluée sur 168 amplifications réalisées à partir d'échantillons osseux

L'efficacité de typage du kit Identifier™ sur les extraits d'ADN provenant de dents a été évaluée à partir de 31 amplifications.

Le diagramme ci-dessous montre clairement la qualité de l'ADN extrait à partir de dents. En effet, sur les 15 loci analysés les typages ont permis d'obtenir un génotype complet pour 11 marqueurs et pour les 4 loci restant seul une perte d'allèle a été observée.

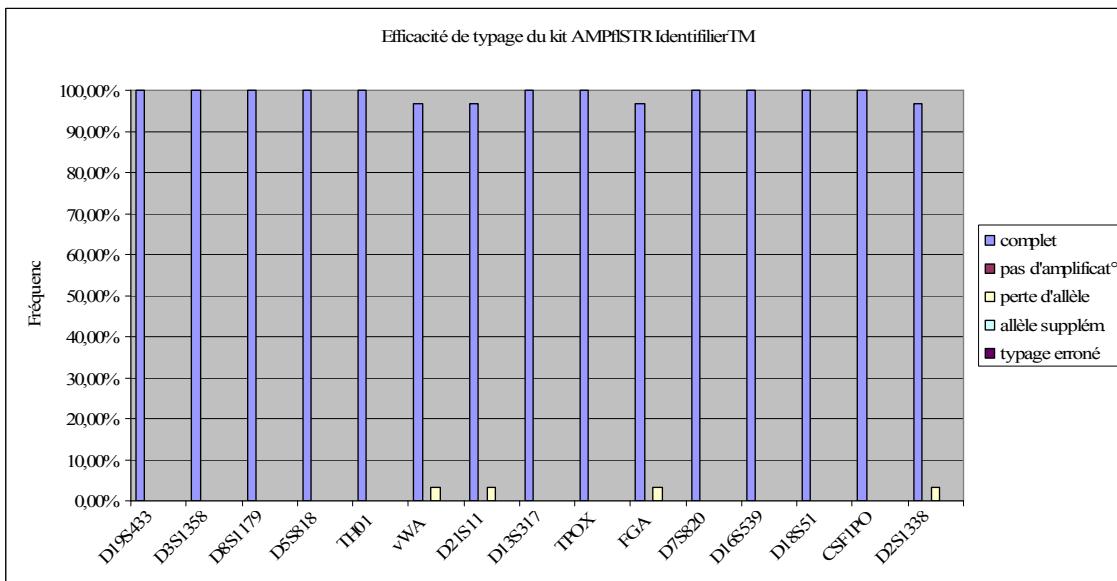


Figure 38b: Efficacité de typage du kit Identifier™ évaluée sur 31 amplifications réalisées à partir d'échantillons dentaires.

IV.3.2.3 Influence du PTB

Afin d'évaluer dans quelle mesure le protocole d'extraction incluant du PTB dans le tampon de lyse permet d'améliorer la qualité de l'ADN et ainsi d'obtenir de meilleurs résultats nous avons comparé les résultats obtenus lors du typage des STR autosomaux avec les kits Profiler PlusTM et IdentifilerTM. Notre analyse s'est basée sur les extraits provenant de poudre d'os ayant subit les deux approches. Les évaluations ont été réalisées sur un total de 64 amplifications en prenant en compte les neufs marqueurs en communs entre ces deux kits d'amplification.

Dans le premier diagramme ci-dessous (Figure 39) les cinq paramètres d'efficacité de typage ont été évalués pour 3 groupes d'échantillons : (i) les échantillons du premier groupe montraient des résultats de quantification faisant état de concentrations en ADN plus faibles avec le tampon de lyse contenant du PTB qu'avec le tampon habituellement utilisé ; (ii) le second groupe correspond aux échantillons pour lesquels les deux types d'extractions ont montré des rendements similaires ; (iii) enfin le dernier groupe rassemble les échantillons pour lesquels les extraits étaient plus concentrés après utilisation d'un tampon de lyse contenant du PTB.

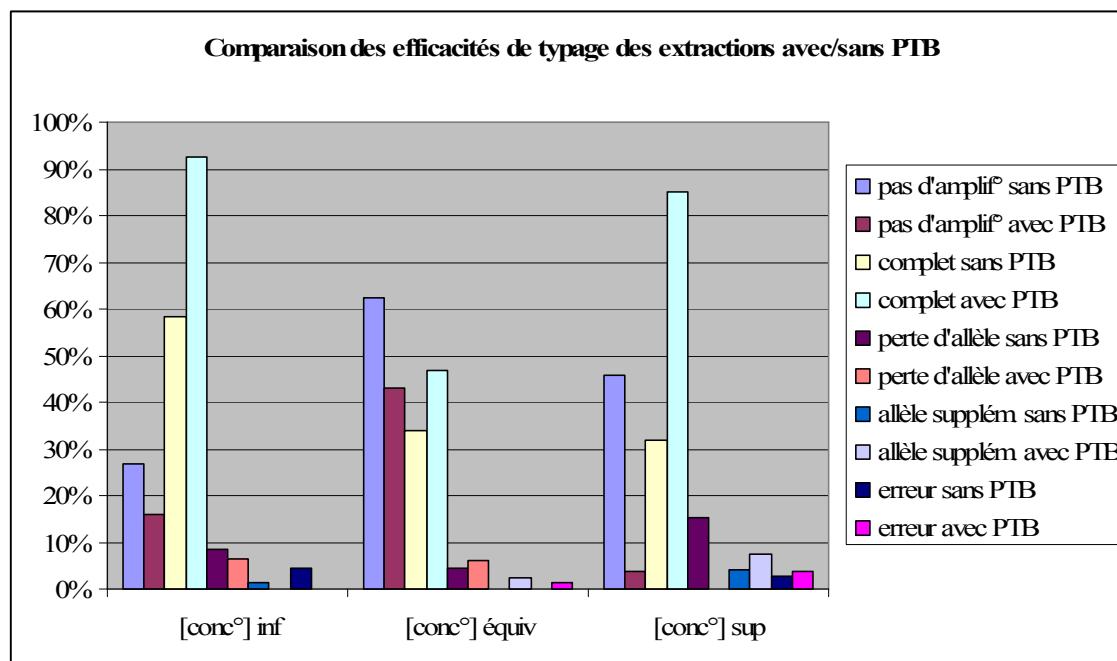


Figure 39 : Efficacité de typage des STR autosomaux pour les trois groupes d'échantillons. [conc°] inf : premier groupe ; [conc°] équiv : deuxième groupe ; [conc°] sup : troisième groupe.

Pour le premier groupe, on constate que les absences d'amplification sont moins nombreuses lors de l'utilisation du tampon contenant du PTB ; de plus la proportion de génotypes complets est plus importante. Concernant les artéfacts d'amplification, la fréquence des pertes d'allèles a diminué avec le PTB et la détection d'allèles supplémentaires ou les évènements de faux typage, n'ont pas été observés avec l'utilisation de PTB.

Une concentration plus faible des extraits de départ semble tout de même permettre un meilleur typage des STR autosomaux en utilisant le tampon contenant du PTB.

Les échantillons pour lesquels les rendements des deux protocoles d'extraction étaient équivalents ont montré le même type d'amélioration des résultats. Toutefois dans des proportions moins nettes que dans le premier groupe : la diminution des absences d'amplification étant de 1,69 dans le premier groupe et de 1,43 dans le second et l'augmentation des génotype complets étant respectivement de 1,58 et 1,38. Les différents types d'artéfacts d'amplification pour ce groupe d'échantillons étaient également plus nombreux avec le tampon PTB tout en restant à des fréquences faibles.

Enfin, les amplifications prises en compte pour le troisième groupe ont montré les plus grands ratios pour la diminution des absences d'amplification et l'obtention de génotypes complets : 11,5 et 2,65. Tout comme pour les échantillons du groupe précédent nous avons constaté une légère augmentation du nombre des différents types d'artéfacts d'amplification.

Il faut toutefois considérer que l'efficacité de typage dans ce groupe n'a été évaluée que pour 10 PCR, correspondant à 90 loci pris en compte, ce qui peut constituer un biais important.

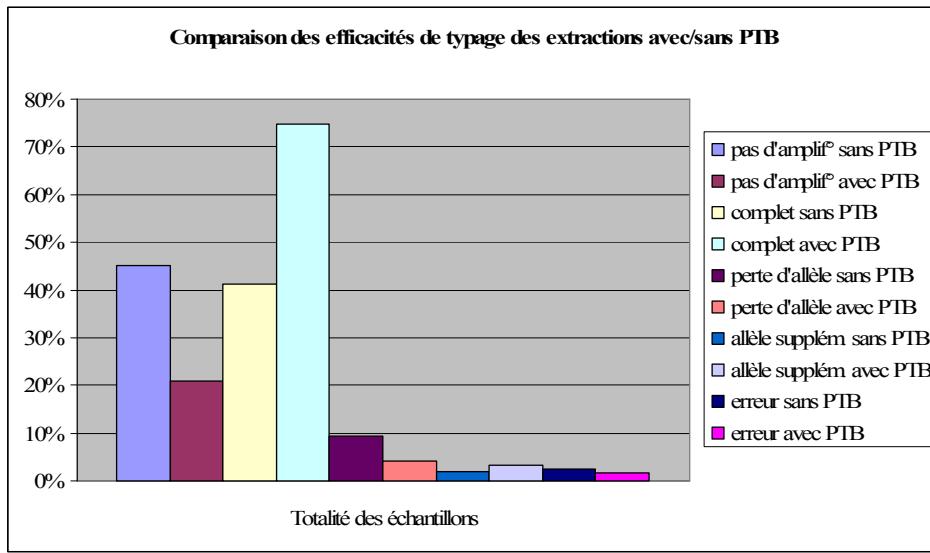


Figure 40: Efficacité de typage comparative pour l'ensemble des échantillons testés avec le tampon de lyse contenant du PTB.

Afin d'avoir une vision globale de l'influence du PTB, nous avons réuni l'ensemble des résultats dans un même diagramme (Figure 40).

Les résultats montrent une diminution des absences d'amplification, une augmentation du nombre de marqueurs avec des génotypes complets ainsi qu'un diminution des pertes d'hétérozygotie. La détection d'allèle supplémentaire semble légèrement augmenter et concernant la détection de faux allèles, la différence entre les deux valeurs n'est pas significative, la proportion de cet artéfact reste constante suivant les deux protocoles employés.

IV.3.3 Equilibre de Hardy-Weinberg

Afin de déterminer si l'ensemble de échantillons analysés pouvait être considéré comme constituant une population à l'équilibre de Hardy-Weinberg, nous avons réalisé un test exact d'équilibre pour les différents loci du kit AMPf/STR® Profiler™ Plus à l'aide du logiciel Genepop 3.4 (Raymond et Rousset, 1995). Nous n'avons pris en compte que ces neufs loci dans un souci d'homogénéisation des données entre les échantillons typés avec ce kit et ceux typés avec le kit Identifiler™.

L'estimateur F_{is} , reflétant la déviation des fréquences par rapport à l'équilibre a également été évalué selon les paramètres de Weir et Cokerham (W&C) (1984). Des valeurs positives pour cet estimateur traduisent un déficit d'hétérozygotes par rapport à la valeur attendue et des valeurs positives un excès d'hétérozygotes.

Les résultats montrent que seuls deux loci, D13 et D21, se rapprochent d'une P-value de 0,95 correspondant à l'équilibre en prenant comme hypothèse H1 un déficit d'hétérozygotes (tableau 8). Seuls ces deux marqueurs présentent en outre des valeurs de F_{is} indiquant un excès d'hétérozygotes.

LOCUS	P-val	Erreur Std	F_{is} W&C
D13	0,9308	0,0009	-0,089
D18	0,2603	0,0041	+0,033
D21	0,9309	0,0023	-0,110
D3	0,5505	0,0015	+0,049
D5	0,0677	0,0010	+0,090
D7	0,0761	0,0008	+0,097
D8	0,2913	0,0023	+0,020
FGA	0,0100	0,0005	+0,137
vWA	0,3082	0,0024	+0,024

Tableau 8 : Résultats obtenus pour le test exact de Hardy Weinberg.

Il semble que l'ensemble des sujets anciens analysés ne puissent être considérés comme une population à l'équilibre de Hardy-Weinberg. Ce résultat n'est pas étonnant attendu que pour plusieurs échantillons nous ne disposons pas de profils complets (les sujets pour lesquels moins de trois loci avaient été amplifiés n'ont toutefois pas été pris en compte). Par ailleurs, notre échantillon ne répond pas à la stricte définition d'une population archéologique puisque les échantillons ne sont pas regroupés de manière chronologique ou temporelle. Toutefois, nos sujets proviennent tous de Yakoutie Centrale ce qui constitue un même ensemble géographique. De plus, pour les différents tests statistiques nous avons tenu compte des relations de parenté pouvant exister entre les sujets.

IV.3.4 Analyse Moléculaire de Variance

Une analyse moléculaire de variance (AMOVA) a été réalisée à l'aide du logiciel ARLEQUIN afin de déterminer si des variations étaient observées en fonction de la répartition chronologique ou géographique des sujets étudiés.

Les échantillons ont été répartis en trois sous groupes chronologiques correspondant aux trois périodes ayant été mises en évidence en considérant les pratiques funéraires et les datations (cf chapitre II.1.2.2. Datations). Nous n'avons pas pris en compte les sujets appartenant à la période pré-médiévale car les profils obtenus ne comprenaient pas assez de marqueurs amplifiés.

Pour la répartition géographique, nous nous sommes servi de la localisation des tombes et avons divisé les échantillons suivant les trois *ulus* dans lesquels ont été faites la plupart des découvertes : Churapchinsky, Khangalassky et Tattinsky.

Répartition	Variance (%)	
	Au sein des populations	Entre les populations
Chronologique	100	0
Géographique	99,2	0,8

Tableau 9 : Résultats de l'analyse AMOVA basée sur les répartitions chronologique et géographique de l'ensemble des sujets anciens.

Les résultats de l'analyse, présentés dans le tableau 9 ci-dessus, montrent qu'aucun des deux facteurs retenus ne présentent un critère de différenciation pour notre échantillon. En effet, la totalité de la variabilité est présente au sein des sujets composant les différents sous groupes et non pas entre les sujets des différents ensembles. Il faut toutefois considérer que la taille des échantillons pourrait entraîner un biais ne permettant pas de différencier nos sujets selon ces critères.

IV.3.5 Relations de parenté au sein des différents ensembles funéraires

Les STR autosomaux ont été utilisés afin de calculer les probabilités des différents liens de parenté pouvant exister entre les sujets ayant été inhumés au sein d'un même site. Il a également été nécessaire de prendre en compte les résultats des marqueurs à transmission uniparentale pour évaluer les différentes hypothèses.

IV.3.5.1 Le site d'Arbre Chamanique (AC)

Ce site, situé dans l'*ulus* de Churapcha, a été nommé ainsi en raison de la présence d'un arbre chamanique près de l'emplacement des tombes. Le nom exact du lieu, déterminé ultérieurement, est « *okhtoubout* » ce qui signifie « mort au champs d'honneur ».

Le site comprend trois tombes proches, dont une sépulture multiple (AC1) dont le coffre contenait les corps de quatre sujets et sur lequel un enfant avait été déposé. Ces tombes sont situées au sommet d'une terrasse dominant un *alas* de taille importante. La tombe AC2 est en position intermédiaire entre l'*alas* et la lisière de la forêt et la troisième tombe est la plus éloignée de l'*alas*.

La tombe Arbre Chamanique 1



Figure 41 : Corps des 4 sujets inhumés dans le coffre de la tombe d'Arbre Chamanique 1. le sujet 4 (YAKa35) était déposé sur le coffre.

L'ensemble des sujets inhumés dans cette tombe présente le même haplotype mitochondrial et devaient par conséquent appartenir à la même lignée maternelle. Pour les deux sujets masculins : sujet **1** (YAKa34) et **5** (YAKa38) (dont le sexe a été déterminé de manière génétique uniquement puisqu'il s'agit d'un sujet immature) nous n'avons pas pu comparer les lignées paternelles puisque aucun résultat exploitable n'a été obtenu pour le sujet YAKa38.

Les probabilités obtenues à partir des différentes bases de données de fréquences alléliques mettent clairement en évidence une relation de parenté de type frère/soeur entre les sujets **1** (YAKa34) et **2** (YAKa36). En effet, les probabilités déduites des valeur des rapports de vraisemblance (ou LR : *Likelihood Ratio*) pour ces relations sont supérieures à 99,999% en utilisant les différentes bases de données.

Concernant l'unique sujet **5** (YAKa38), seuls des résultats portant sur le kit Profiler Plus™ ont été obtenus au cours de l'année 2005. Faute de matériel osseux, il n'a pas été possible de réaliser des extractions supplémentaires pour tester le kit Identifier cette année. Les calculs ont par conséquent été effectués sur 9 loci et non 15 comme précédemment. Les résultats sont donc moins discriminants que ceux obtenus pour les autres sujets. Cet enfant apparaît comme ayant 99,96% de chance d'être le frère du sujet **1** ; 99,79% de chance d'être le frère du sujet **2**. Il semble donc que ces 3 sujets, qui possèdent la même lignée maternelle, soient frères et soeurs.

Les différentes relations possibles entre les sujets **1**, **2**, **5** et le sujet **3** (YAKa37) ont été testées. Des rapports de vraisemblance relativement élevés ont été obtenus, cependant la relation qui présente les probabilités les plus importantes semble montrer que le sujet **3** était la mère des 3 autres individus avec des probabilités respectives de : 99,9989%, 99,997% et 99,95% (ce score est plus faible en raison d'un nombre de marqueurs restreint et d'une possible mutation +1 pour le locus FGA qu'il a fallu prendre en compte (Brinkmann *et al.*, 1998)).

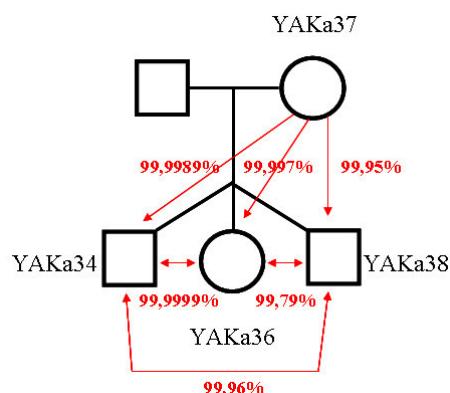


Figure 42 : Relations de parenté au sein de la tombe Arbre Chamanique 1.

Les valeurs des rapports de vraisemblance pour les relations entre le sujet 4 (YAKa35) qui a été retrouvé déposé à l'extérieur du coffre contenant les autres individus, ne font apparaître aucune parenté. Cependant ce résultat vient du fait que pour ce sujet les profils obtenus étaient très incomplets et par conséquent il n'a pas été possible de comparer les données obtenues sur les autres sujets. Il est toutefois permis de penser qu'une relation liait cet enfant aux autres sujets car il possède également le même haplotype mitochondrial.

Relations avec la tombe AC2

Le sujet découvert dans la tombe AC2 (YAKa41) présente le même haplotype mitochondrial que les sujets de la tombe n°1. Son profil STR exclut d'emblée une relation de type père-fils avec le sujet YAKa34 (loci permettant l'exclusion : D19, D5) ou une relation de type mère-fils avec YAKa37 (loci permettant l'exclusion : D13, D19 et TH01).

Une relation de type mère-fils est toutefois envisageable avec le sujet YAKa36 puisque la probabilité déduite du rapport de vraisemblance donne 99,97% de chance que ce type de relation existe entre ces deux individus. De plus, l'estimation de l'âge de ces deux sujets (25-30 ans et moins de 12 ans) pourraient être en accord avec cette hypothèse. La probabilité d'existence d'une relation de type demi-frères/oncle-neveu/grand-parent-petit-enfant est de 99,53%. La probabilité d'existence de ce même type de relation entre YAKa 34 et YAKa41 est de 96,3% et de 87,84% avec YAKa38. Il semble donc que le sujet de la tombe d'AC2 soit le fils du sujet YAKa36, les sujets YAKa34 et 41 étant probablement ces oncles.

Il se pose toutefois le problème de la période d'inhumation de ces sujets. Le mode de construction de la tombe dans laquelle est inhumé le sujet YAKa41 est du même type que celui de la tombe n°1. L'absence d'écorce de bouleau sur la surface du coffre pourrait signifier une période plus récente, mais cette absence peut également être en relation avec un sujet pour lequel il n'aurait pas été jugé nécessaire d'utiliser le revêtement d'une tente d'été.

Relations avec la tombe AC3 :

La tombe n°3 est la plus éloignée de la tombe n°1 et se rapproche de la lisière de la forêt. Le sujet féminin inhumé dans cette tombe ne présente pas le même haplotype mitochondrial que les autres individus de ce site. Les résultats des tests de parenté indiquent qu'aucune relation de proche parenté n'existe entre ce sujet et les individus des deux autres tombes. Les rapports de vraisemblance obtenus sont en effet tous inférieurs à 0,5 (probabilité inférieure à 33%). Il

n'est donc pas possible de déterminer quelles étaient les liens qui existait entre cette femme et les sujets des deux autres tombes.

IV.3.5.2 Le site de Jarama

Le site de Jarama a été découvert en sous bois, au sommet d'une bute surplombant un *alas*. Ce site comporte la tombe d'un sujet adulte et un ensemble de trois tombes qui contenaient des sujets immatures a été retrouvé à quelques mètres de la première sépulture.

A l'exception du sujet adulte de la tombe Jarama1 (YAKa 50) les sujets immatures n'ont pas permis d'obtenir de profils STR permettant la détermination des liens de parenté (cf tableau 7). L'haplotype mitochondrial du sujet adulte diffère de ceux des 3 enfants. Seuls les sujets des tombes 3 (YAKa52) et 4 (YAKa53) partagent une lignée maternelle commune. Malheureusement les profils STR obtenus étant très incomplets, il n'a pas été possible de tester les différentes hypothèses de parenté.

Ce site pourrait confirmer les conclusions proposées par Ricaut *et al.* (2005) portant sur le regroupement d'individus appartenant à une même structure clanique mais ne partagent pas forcément de liens de proche parenté.

IV.3.5.3 Le site de Bekh Alas

Le site de Bekh Alas regroupe 4 tombes qui étaient localisées à proximité d'une décharge publique à ciel ouvert. Ces 4 tombes avaient été déplacées et éventrées suite au passage des engins de chantier lors de la construction de la décharge. Nous n'avons par conséquent pu collecter que peu d'informations archéologiques (mode de construction des coffres et présence de restes de vêtements).

Seuls deux sujets présentaient des profils comportant un nombre de marqueurs permettant de tester les relations de parenté. Les haplotypes mitochondriaux des deux individus étant différents, une relation de type fratrie était exclue. Les LR obtenus pour les relations de type cousins ou demi-frères/oncle-neveu/grand-parent-petit-enfant ont permis d'exclure ce type de relations entre ces deux sujets. Les marqueurs amplifiés pour les STR du chromosome Y

montrent également des différences entre les allèles. Il semble donc que ces deux individus n'étaient pas apparentés.

IV.3.5.4 Le site de Ken Ebe

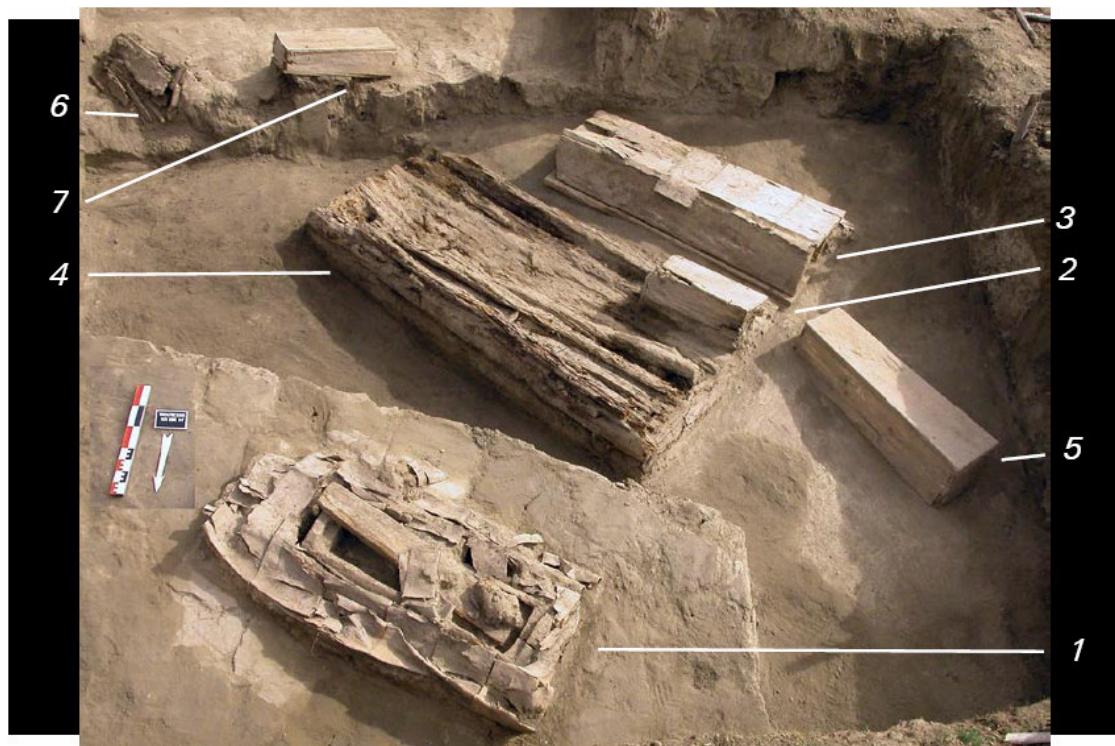


Figure 43 : Ensemble des tombes du site principal de Ken Ebe (photos MAFSO 2005).

Le site de Ken Ebe a livré 9 tombes au total, 7 étaient localisées à proximité (cf figure 43), une était située à quelques mètres du premier ensemble (tombe n°9, YAKa76) et la dernière était éloignée d'une centaine de mètres environ (tombe n°8, YAKa75).

Au sein du premier ensemble, la conservation des restes osseux du sujet de la tombe n°1 (YAKa71) ne permettait pas de réaliser de prélèvement pour l'analyse génétique. Concernant les tombes que nous avons pu analyser, cet ensemble est composé de 5 tombes de sujets immatures (n°2, YAKa68 ; n°3, YAKa67 ; n°5, YAKa72 ; n°6, YAKa73 et n°7, YAKa74) et de la sépulture d'une homme âgé (n°4, YAKa71). D'après les pratiques funéraires les tombes les plus anciennes semblent être les tombes 1, 4 et 3 ; viennent ensuite le dépôt des tombes 2, 5, 6 et 7.

Le sujet de la tombe n°8 était une femme sans doute âgée car presque totalement édentée ; quant au sujet de la tombe n°9 il s’agissait d’un enfant d’environ 7 à 8 ans.

Comme pour les autres sites la détermination au sein de cet ensemble funéraire complexe s’est avéré difficile du fait que les sujets mis au jour présentaient une conservation différentielle de l’ADN entraînant l’obtention de résultats très inégaux et de profils parfois très incomplets (YAKa72,73,74,75).

Les résultats obtenus mettent en avant que peu de sujets de ce site devaient partager des liens de parenté. En effet, seuls les sujets 3 (YAKa67) et 8 (YAKa75) possèdent le même haplotype mitochondrial. Les rapports de vraisemblance calculés pour les différents types de relations entre ces deux sujets excluent un lien de proche parenté de type frère-soeur, cousin-cousine ou demi-frères/oncle-neveu/grand-parent-petit-enfant.

Seuls les paires YAKa68-YAKa71 et YAKa71-YAKa73 pourraient être liés par une relation de type demi-frères/oncle-neveu/grand-parent-petit-enfant. Cependant, les probabilités déduites des rapports de vraisemblance restent tout de même faibles : respectivement 71,34% et 75,47%.

Tout comme pour le site de Jarama, nous nous trouvons donc en présence d’un site évoquant un regroupement des sujets par appartenance à un même groupe social plutôt que dans le cas de sujets partageant des liens de proche parenté. Considérant nos résultats, il est, pour l’instant, impossible de confirmer ou d’infirmer cette hypothèse.

IV.3.5.5 Le site de Bouogaryma

Le site de Bouogaryma était constitué par deux tombes de sujets adultes, un homme (n°1, YAKa78) et une femme âgée (n°2, YAKa79), éloignées de plusieurs mètres surplombant un *alas* en formation. Ce site est supposé correspondre à un cimetière de clan.

Les deux sujets mis au jour ne partagent pas le même haplotype mitochondrial et les relations de parenté du type cousin et demi-frère/oncle-neveu/grand-parent-petit-enfant ont donné des rapports de vraisemblance de 0,880 et 0,259 respectivement. Les probabilités correspondantes donnent 46,8% et 20,5% de chance pour que ces individus soient liés par ce type de relations. Ces deux individus n’étaient sans doutes pas liés par des relations de parenté.

IV.3.5.6 Les tombes n°9 et n°10

Ces deux sujets inhumés côte à côte présentent le même haplotype Y, tandis que leurs haplotypes mitochondriaux ne diffèrent que par une seule mutation (le sujet YAKa19 porte une substitution C/T en position 16298 par rapport à l'haplotype de YAKa17). Cette mutation aurait pu se produire en une génération.

La probabilité que ces individus soient frères étant de 9%, nous avons rejeté cette possibilité ; en revanche une relation de type cousin ou demi-frère/oncle-neveu donne une probabilité de 73%.

Ces résultats permettent donc d'exclure une relation frère-frère entre ces deux sujets. Les deux autres hypothèses sont envisageables et les résultats ne permettent pas de trancher en faveur d'une relation de type oncle neveu ou cousins germains paternels.

Le recrutement pour ces deux tombes a donc été fait suivant des critères de relation de parenté. L'appartenance à une même lignée paternelle semble être, en effet, le critère le plus probable puisque ces deux sujets présentent le même haplotype.

IV.3.5.7 Site de la rivière Tandy

Les relations de parenté entre les deux nouveaux-nés des tombes STR1a (YAKa29) et b (YAKa30) ont également été étudiées. La comparaison des allèles, pour les 6 marqueurs ayant donné un résultat, montre que seuls 2 allèles sont communs entre ces 2 sujets. De plus, les haplotypes mitochondriaux diffèrent par deux mutations et ne peuvent être rattachés à la même lignée maternelle. Il ne semble pas exister de relation de parenté entre ces deux sujets, l'inhumation de ces deux individus aurait donc été d'avantage motivée plutôt par un critère d'âge plutôt que par l'appartenance à une même famille.

IV.4 Les marqueurs du chromosome Y

IV.4.1 Les STR

La détermination du sexe réalisée par le typage de l'amélogénine a permis de diagnostiquer 38 sujets masculins. Les sujets étudiés lors de la première année ont été typés à l'aide du kit PowerPlex® Y (PP-Y, Promega) ; alors que les sujets de la deuxième année ont été typés dans un premier temps avec le kit PP-Y puis avec le kit AMPflSTR® Y-Filer™ (Applied Biosystems) à l'exception des sujets YAKa40, 41 et 49 pour lesquels nous n'avons pas pu répéter les analyses avec ce dernier kit. Enfin pour les sujets de la troisième année nous avons utilisé le kit Y-Filer™ exclusivement.

Les haplotypes consensus ont été déterminés par la comparaison des résultats des différentes amplifications et seuls les allèles présents au moins dans deux amplifications séparées ont été pris en compte pour le consensus. Les résultats sont présentés dans le tableau 10.

IV.4.1.1 Efficacité de typage

Nous avons pu déterminer des haplotypes consensus complets pour 10 sujets étudiés avec le kit PP-Y , soit 71,4% et pour 17 sujets avec le kit Y-Filer™, soit 70,8%. Les efficacités de typage de ces deux kits ont été évaluées par la même méthode que pour les STR autosomaux. Les résultats sont présentés dans les deux diagrammes ci-dessous (Figure 44a et b). Les STR du chromosome Y étant des marqueurs haploïdes les pertes d'allèles n'ont été évaluées que pour le marqueur DYS385 qui est le seul marqueur dupliqué inclus dans ces deux kits.

L'évaluation de l'efficacité de typage du kit PP-Y a été réalisée sur 31 PCR. Comme pour les STR autosomaux, il a été constaté que le succès d'amplification est fonction de la taille des produits de PCR des différents marqueurs. Tous les marqueurs rentrent dans ce schéma à l'exception de DYS389II qui présente une meilleure efficacité de typage que le marqueur DYS19 bien que présentant une taille supérieure.

Mis à part les absences d'amplification, la fréquence des artefacts de PCR est assez faible puisque des allèles supplémentaires ont été observés uniquement pour DYS438 et les typages

erronés n'ont affectés que DYS19 avec dans les deux cas des fréquences inférieures à 7%. La fréquence observée des pertes d'allèles pour DYS385 est de 12,9%.

Tableau 10 : Profils consensus établis à partir des différentes amplifications réalisées pour les STR du Chromosome Y (cf infra).

Echantillon	DYS456	DYS389I	DYS389II	DYS390	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA	DYS437	DYS438	DYS448
YAKa15		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa17		13	28	23		15	11 18	14	10	12		11		14	10	
YAKa19		13	28	23		15	11 18	14	10	12		11		14	10	
YAKa21		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa23		-	-	-		-	-	-	-	-		-		-	-	
YAKa24		14	32	-		-	-	14	11	-		-		-	11	
YAKa25		-	-	-		-	-	-	-	-		-		-	-	
YAKa26		13	29	23		14	12 13	13	10	10		14		14	10	
YAKa27		13	-	24			12 12	14	9	13		-		14	-	
YAKa29		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa31		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa34	14	14	32	23	16	14	11 13	14	11	10	22	16	12	14	11	19
YAKa38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa39	14	14	32	23	16	14	11 13	14	11	10	22	16	12	14	11	19
YAKa40		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa41		14	32	23		14	11 13	14	11	10		16		14	11	
YAKa47	14	13	31	24	18	15	11 14	14	10	11	21	12	12	15	10	-
YAKa49		14	31	23		14	11 13	14	11	10		-		14	11	
YAKa51	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa56	14	14	31	23	16	16	11 13	15	11	10	22	16	12	14	11	19
YAKa57	14	14	30	22	19	14	11 13	14	10	10	22	14	12	14	10	19
YAKa58	14	14	-	-	16	-	-	14	11	-	22	-	12	14	11	(19)
YAKa59	14	-	-	-	-	-	-	14	-	-	-	-	-	14	-	-
YAKa60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa64	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	19
YAKa65	14	14	-	23	(16/17)	14	(11 13)	14	10	10	22	15	12	14	11	19
YAKa66	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	20
YAKa67	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	20
YAKa68	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	19
YAKa69	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	19
YAKa70	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	20
YAKa71	14	14	31	23	16	14	11 13	13	11	10	22	15	12	14	11	19
YAKa73	-	(14)	-	-	16	-	-	14	10	-	22	-	-	14	-	19
YAKa74	14	-	-	-	16	-	-	14	-	-	-	-	-	14	-	19
YAKa78	14	14	31	23	16	14	11 13	14	11	10	22	15	12	14	11	20
YAKa80	14	14	32	24	16	14	11 13	14	11	10	22	16	12	14	11	19
YAKa81	14	14	31	23	14	14	11 13	14	10	10	22	15	12	14	11	19
YAKa86	14	14	31	23	16	14	11 13	14	11	10	22	16	12	14	11	19

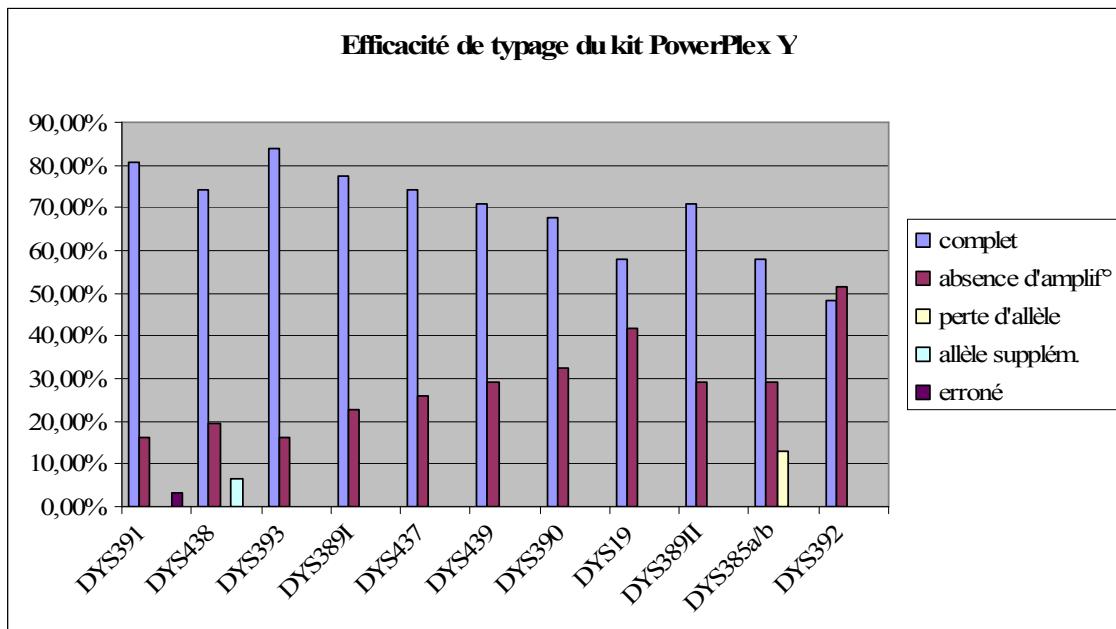


Figure44a : Efficacité de typage du kit PowerPlex® Y évaluée à partir de 31 amplifications.

Pour le kit Y-Filer™ nous avons évalué l'efficacité de typage sur 65 amplifications. La corrélation entre la taille des marqueurs et l'efficacité de typage est beaucoup moins nette pour ce kit. En effet, tous les marqueurs montrent une fréquence de typages corrects supérieure à 50% et on constate pour la majeure partie des loci que l'efficacité de typage est supérieure ou égale à 60%. Seuls DYS19 et DYS439 présentent des proportions plus faibles. Mis à part DYS393, qui présente une fréquence de typages corrects supérieure à 90%, l'efficacité de typage est relativement homogène pour l'ensemble des marqueurs.

Des artefacts d'amplifications ont été observés dans un plus grand nombre de marqueurs sans doute en raison du nombre plus important de PCR prises en compte. Toutefois les valeurs des fréquences correspondent généralement à un ou deux faux allèles ou allèles supplémentaires observés pour l'ensemble des analyses. Par ailleurs, pour le locus DYS385 la fréquence de pertes d'allèle observée est inférieure à 8%.

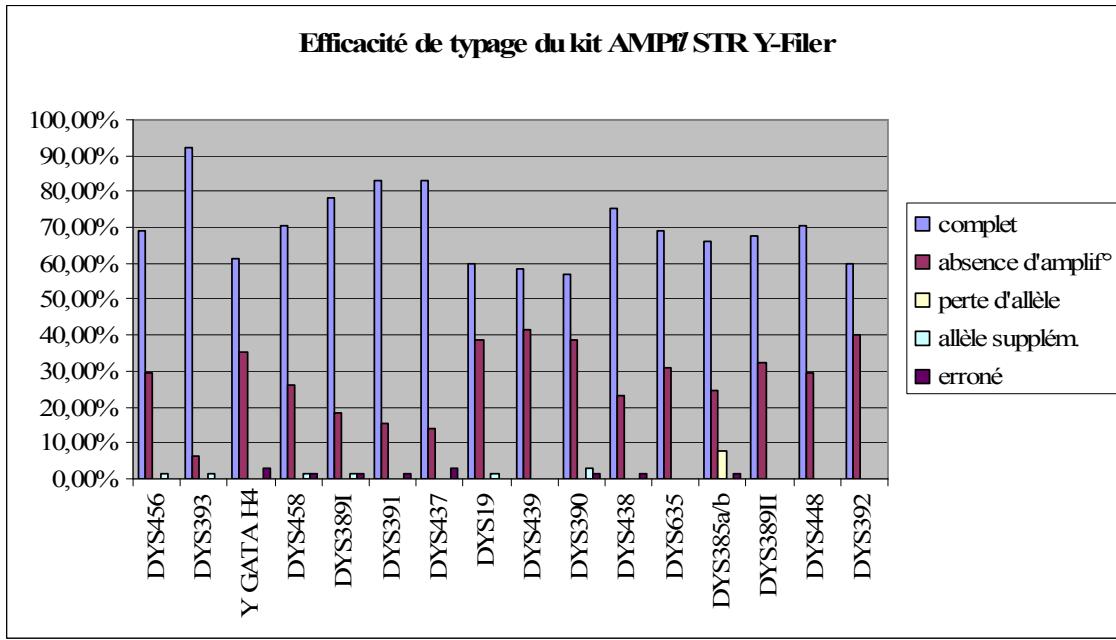


Figure44b : Efficacité de typage du kit AMPf/STR® Y-Filer™ évaluée à partir de 65 amplifications.

IV.4.1.2 Analyse des données

Paramètres de diversité

Nous avons évalué la diversité génétique pour chacun des loci au sein de l'ensemble des sujets anciens étudiés ainsi que pour la population moderne composée de 97 individus pour lesquels des prélèvements génétiques ont été effectués. Ces valeurs ont été comparées aux diversités calculées pour les populations présentes dans notre base de données. Le tableau présentant les valeurs chiffrées se trouve en Annexe 10, mais pour plus de lisibilité nous avons choisi de représenter ces comparaisons sous forme de diagrammes (Figure 45).

Le résultat le plus marquant de cette comparaison est que pour l'ensemble des marqueurs la diversité génétique observée dans l'échantillon ancien étudié est supérieure à celle de la population moderne.

La diversité génétique de 5 marqueurs (DYS19, DYS389I, DYS390, DYS393 et DYS 439) dans la population Yakoute, aussi bien ancienne que dans les différents échantillons modernes pris en compte, est inférieure à celle observée dans les autres populations.

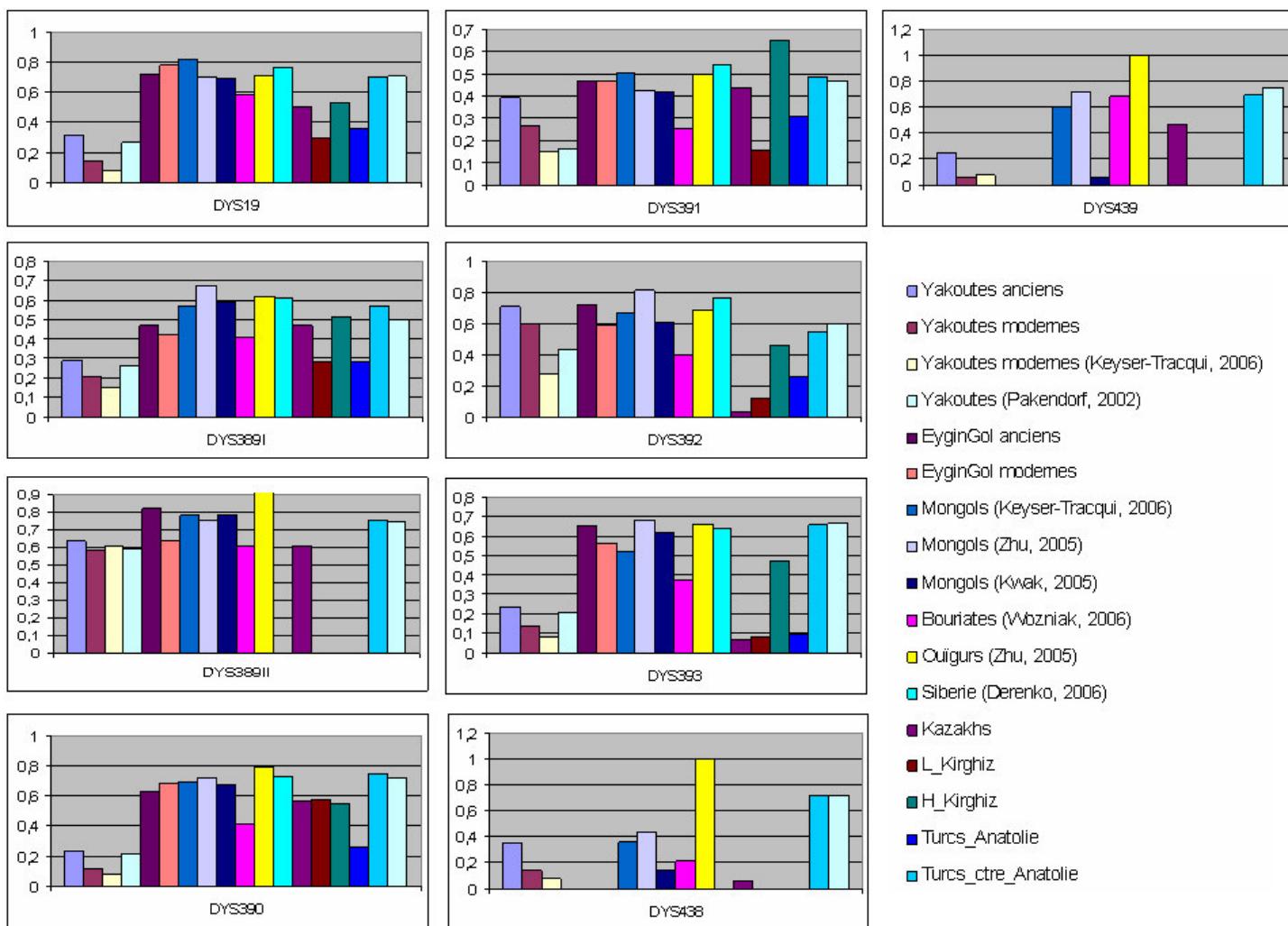


Figure 45 : Comparaison des indices de diversité génétique pour les STR du chromosome Y entre différentes populations.

Pour le locus DYS389II, la diversité observée est équivalente à celle des habitants actuels de la vallée d’Eygin Gol et de la population Bouriate. Enfin pour DYS391, DYS392 et DYS438, les diversités génétiques sont comprises entre les valeurs observées pour les différentes populations.

Nous avons calculé la diversité haplotypique pour l’ensemble des sujets anciens analysés et actuelle selon la formule de Nei (1987). Ce paramètre a été comparé aux données disponibles dans la littérature. Les résultats sont présentés dans le tableau 11 ci-dessous.

Population	Diversité haplotypique
Yakoutes anciens	0,89
Yakoutes actuels	0,90
Yakoutes actuels (Pakendorf <i>et al.</i> , 2006)	0,79
Bouriates (Wozniak <i>et al.</i> , 2006)	0.8691
Ouigours (Zhu <i>et al.</i> , 2006)	0.9993.
Ouigours*	0,99
Mongols (Zhu <i>et al.</i> , 2006)	0.9987
Mongols*	0,99
Kirghiz*	0,94
Kazakhs*	0,94
Uzbeks*	1,00

Tableau 11 : Diversités haplotypiques basées sur les STR du chromosome Y de différentes populations. * données d’après Zerjal *et al.*, 2002.

La diversité haplotypique de notre échantillon ancien est relativement basse comparée aux populations d’Asie Centrale et équivalente à celle observée dans la population actuelle étudiée au laboratoire. Elle est, en revanche, supérieure à celle observée dans la population de Yakoutie Centrale analysée par Pakendorf *et al.* (2006). La diversité observée chez les Bouriates est comparable à celle de notre population archéologique.

Haplotypes partagés

La population prise en compte pour les différentes analyses se compose des 26 sujets ayant permis d’obtenir des profils complets pour les loci composant l’haplotype minimum (i.e. DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392). Nous avons choisi d’effectuer les différents calculs et comparaisons à partir de ces marqueurs car la plupart des

données de la littérature sont disponibles uniquement pour ces loci et parfois pour un nombre plus réduit.

Au sein de l'ensemble des sujets anciens étudiés, nous avons pu distinguer 12 haplotypes différents présentés dans le tableau 12 ci-dessous. Deux haplotypes sont retrouvés avec des fréquences élevées dans notre échantillon : l'haplotype (H1) des sujets YAKa64, YAKa66, YAKa67, YAKa68, YAKa69, YAKa70, YAKa78 avec une fréquence de 26,85% et l'haplotype (H2) des sujets YAKa15, YAKa21, YAKa29, YAKa31, YAKa34 et YAKa39, YAKa40, YAKa41 qui représente environ 30% des lignées paternelles de notre population. Aucun des sujets présentant ces haplotypes n'est apparenté (cf Résultats analyse des liens de parenté) et tous proviennent de sites différents à l'exception de YAKa67 et 68. Un troisième haplotype est retrouvé chez plusieurs sujets de notre échantillon ancien : il s'agit de l'haplotype des sujets YAKa17 et YAKa19, pour lesquels il n'a pas été possible d'exclure une relation de parenté. Les sept autres haplotypes sont uniques dans notre échantillon ancien.

Echantillon	Frq abs	Frq rel	DYS 19	DYS 385	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393
YAKa64	7	26,85	14	11 13	14	31	23	11	15	14
YAKa15	8	30,8	14	11 13	14	32	23	11	16	14
YAKa17	2	7,7	15	11 18	13	28	23	10	11	14
YAKa26	1	3,85	14	12 13	13	29	23	10	14	13
YAKa47	1	3,85	15	11 14	13	31	24	10	12	14
YAKa56	1	3,85	16	11 13	14	31	23	11	16	15
YAKa57	1	3,85	14	11 13	14	30	22	10	14	14
YAKa65	1	3,85	14	11 13	14	-	23	10	15	14
YAKa71	1	3,85	14	11 13	14	31	23	11	15	13
YAKa80	1	3,85	14	11 13	14	32	24	11	16	14
YAKa81	1	3,85	14	11 13	14	31	23	10	15	14
YAKa86	1	3,85	14	11 13	14	31	23	11	16	14

Tableau 12 : Haplotypes minimum des STR du chromosome Y observés dans la population Yakoute ancienne. Frq abs : fréquence absolue ; Frq rel : fréquence relative en %.

Ces différentes lignées paternelles ont été comparées avec les données obtenues sur nos échantillons modernes afin d'évaluer les haplotypes conservés. Les haplotypes H1 et H2 sont retrouvés avec des fréquences respectives de 27,8% et 16,5% dans notre échantillon moderne et sont également présents dans les données de Pakendorf *et al.* (2006) avec une fréquence de 10% et 43%. Les haplotypes des sujets YAKa17 et YAKa19 ainsi que ceux des sujets YAKa81, YAKa86, sont également présents parmi nos 97 échantillons modernes. Les lignées paternelles des sujets YAKa47, YAKa71 et YAKa80 ont été retrouvées dans notre échantillon

moderne avec une mutation +/-1 tandis que les haplotypes des sujets YAKa26, YAKa56, YAKa57 n'ont pas été observés dans notre échantillon moderne.

Ces données ont été comparées à l'aide du logiciel YCDMA aux populations présentes dans notre base de données. Nous avons également réalisé une recherche sur les bases de données disponibles en ligne (YHRD, Y-Filer database et PowerPlex Y database). Les résultats sont présentés dans le tableau 13 ci-dessous.

Les résultats de ces comparaisons montrent que seuls les haplotypes des sujets YAKa17 et YAKa 26 sont retrouvés fréquemment dans d'autres populations que les Yakoutes. Peu de données sur les STR du chromosome Y sont disponibles pour les populations Toungousses (Evenks et Evènes). Nous avons tout de même pu avoir accès à des données concernant les haplotypes appartenant à l'haplogroupe N3 de la population Evenks (Karafet *et al.*, données non publiées et Xue *et al.*, données non publiées). La comparaison avec ces haplotypes a permis de mettre en évidence que l'haplotype H2 était retrouvé chez 2 Evenks et que l'haplotype du sujet YAKa86 était présent chez 8 Evenks.

	Base de données du laboratoire	Bases de données en ligne
YAKa64 (H1)	NR	NR
YAKa15 (H2)	2 Evenks	NR
YAKa17	3 Mongols	71 Kalmyks, 19 Bouriates
YAKa26	1 Turc Centre Anatolie, 4 Mongol	1 Mongol, 1 Kazakh, 10 Turcs
YAKa47	NR	NR
YAKa56	NR	NR
YAKa57	NR	+/-1 31 Bouriates
YAKa65	NR	NR
YAKa71	NR	NR
YAKa80	NR	NR
YAKa81	NR	1 russe Vladivostok
YAKa86	8 Evenks	NR

Tableau 13 : Comparaisons des haplotypes du chromosome Y dans notre base de données (références en Annexe 6) et dans les bases de données en ligne. NR : haplotype non retrouvé lors de la comparaison.

IV.4.2 Les SNP

Les résultats des typages des SNP obtenus à l'aide du protocole SNaPshot sont présentés dans le tableau 14 ci-dessous. Onze sujets anciens ont été sélectionnés pour représenter l'ensemble des haplotypes présents dans notre population et ainsi pouvoir déterminer par comparaison les affiliations aux différents haplogroupes de l'ensemble des individus anciens. L'état de dégradation de l'ADN de certains sujets anciens et le manque de matériel disponible ont entraîné l'impossibilité de diagnostiquer directement leur haplogroupe, aussi avons-nous sélectionné des sujets modernes qui présentaient le même haplotype. Cependant, il n'a pas été possible de déterminer l'haplogroupe des sujets YAKa17/YAKa19 et YAKa47. Pour les sujets YAKa17/YAKa19, la détermination de la mutation TAT avait été effectuée par la méthode MALDI-TOF et a montré que ces sujets ne portaient pas l'allèle C, ils n'appartiennent donc pas à l'haplogroupe N3. Concernant le sujet YAKa 47 la construction d'un *median-joining* network (cf Annexe 12) comprenant l'ensemble des sujets anciens a montré qu'il appartient vraisemblablement à un haplogroupe différent de N3.

Echantillon	Haplogroupe
YAKa64 (H1)	N3
YAKa15 (H2)	N3
YAKa17	? (pas TAT-C)
YAKa26	K(xN,O,P)
YAKa47	? (pas TAT-C)
YAKa56	N3
YAKa57	P
YAKa65	N3
YAKa71	N3
YAKa80	N3
YAKa81	N3
YAKa86	N3

Tableau 14 : Affiliation des sujets anciens aux différents haplogroupes du chromosome Y.

La majorité des haplotypes appartiennent à l'haplogroupe N3, caractérisé par la mutation TAT-C, dont la fréquence est de 78,3% dans l'échantillon ancien étudié.

Cette fréquence a été comparée aux données de la littérature pour différentes populations (Tableau 15). Les fréquences les plus élevées sont observées chez les Yakoutes et on retrouve des fréquences importantes de cet haplogroupe chez les Khantys (populations vivant dans le nord-ouest de la Sibérie dans la région de la rivière Ob) et les Bouriates ainsi que chez les Finlandais.

Cette fréquence élevée contraste avec les populations Toungousses environnantes qui présentent des fréquences beaucoup plus faibles : 16,8% chez les Evenks, 22% chez les Evènes et également chez les Youkaguirs : 25%. Très peu de lignées paternelles appartenant à cet haplogroupe sont retrouvées chez les Mongols puisque la fréquence de l'haplogroupe N3 est de seulement 2% environ et cet haplogroupe est absent des populations d'Asie Centrale (Kirghiz et Uzbeks) mais rencontré avec une fréquence de 10% chez les Kazakhs.

Population	Fréquence (%)	Référence
Russes	15	Zerjal <i>et al.</i> , 1997
	14	Derenko <i>et al.</i> , 2006
	3.3	Karafet <i>et al.</i> , 1999
Finlandais	52.4	Zerjal <i>et al.</i> , 1997
	61.1	Lahermo <i>et al.</i> , 1999
Saami	25.0	Zerjal <i>et al.</i> , 1997
	48.9	Lahermo <i>et al.</i> , 1999
Karelies	39.6	Lahermo <i>et al.</i> , 1999
Mari	30.6	Zerjal <i>et al.</i> , 1997
	33.3	Lahermo <i>et al.</i> , 1999
Mansis	18.2	Lahermo <i>et al.</i> , 1999
Khanty	63.2	Lahermo <i>et al.</i> , 1999
Koryaks	20.0	Lahermo <i>et al.</i> , 1999
	33.0	Karafet <i>et al.</i> , 1999
Youkaguirs	25.0	Karafet <i>et al.</i> , 1999
Esquimaux sibériens	50.0	Karafet <i>et al.</i> , 1999
Evenks	16.8	Karafet <i>et al.</i> , 1999
	16	Derenko <i>et al.</i> , 2006
Evènes	22.0	Karafet <i>et al.</i> , 1999
Tuvas	9.7	Derenko <i>et al.</i> , 2006
Tofalars	25	Derenko <i>et al.</i> , 2006
Bouriates	57.7	Zerjal <i>et al.</i> , 1997
	18.9	Derenko <i>et al.</i> , 2006
	28.4	Karafet <i>et al.</i> , 1999
Mongols	2.1	Zerjal <i>et al.</i> , 1997
	2.7	Karafet <i>et al.</i> , 1999
Ouigours	0	Karafet <i>et al.</i> , 2001
Kazakhs	10.0	Karafet <i>et al.</i> , 2001
Altaï	0	Karafet <i>et al.</i> , 2001
Uzbeks	1.9	Karafet <i>et al.</i> , 2001
Kirghiz	0	Karafet <i>et al.</i> , 2001
Yakoutes	85.7	Zerjal <i>et al.</i> , 1997
	100.0	Lahermo <i>et al.</i> , 1999
	75.0	Karafet <i>et al.</i> , 1999
	85.7	Pakendorf <i>et al.</i> , 2002
	74	Pakendorf <i>et al.</i> , 2006
Yakoutes anciens	78,3	notre étude

Tableau 15 : Fréquences de l'haplogroupe N3 dans les populations eurasiennes.

IV.4.3 Analyse Network

Nous avons réalisé la construction d'un *Median Joining* network avec les différents haplotypes appartenant à l'haplogroupe N3 provenant de l'ensemble des sujets anciens étudiés et de la population moderne ainsi qu'avec des données disponibles pour d'autres populations (Figure 46). Les haplotypes utilisés sont les suivants : 2 haplotypes bouriates (Derenko *et al.*, unpublished data) ; 10 haplotypes mongols (Xue *et al.*, 2006 ; Zerjal *et al.*, 2002) et 2 haplotypes ouigours (Xue *et al.*, 2006).

Nous avons utilisé les données des loci suivant : DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393. Seuls ces marqueurs ont été retenus car, en raison de l'hétérogénéité des taux de mutations entre les loci et les populations, nous ne disposons de valeurs de pondération fiables que pour cet ensemble de marqueurs (Kayser *et al.*, 2000).

La taille de chaque cercle correspond à l'effectif relatif des différents haplotypes. Le network obtenu montre clairement une structure bipolaire, les haplotypes yakoutes étant clairement distincts de ceux observés chez les Mongols et les Ouigours. L'haplotype nodal semble être Br134 qui est un haplotype retrouvé fréquemment dans la population bouriate (puisque il représente environ 14% des lignées paternelles (Wozniak *et al.*, 2006)) et également dans la population mongole mais à une fréquence moindre.

L'ensemble des haplotypes yakoutes, et notamment ceux découverts dans l'échantillon ancien analysé, ne sont séparés les uns des autres que par une seule mutation à l'exception du sujet YAKa56. Cet haplotype est en effet séparé par 4 mutations du reste des haplotypes. Les lignées paternelles yakoutes apparaissent donc séparées des autres haplotypes appartenant à l'haplogroupe N3. Cette caractéristique semble indiquer une expansion rapide à partir d'un nombre de fondateurs relativement réduit (Pakendorf *et al.*, 2006). L'haplotype du sujet moderne YAK42 est lié par une seule mutation à la fois à l'haplotype nodal bouriate, la branche des haplotypes yakoutes et partage également les mêmes mutations que deux sujets mongols.

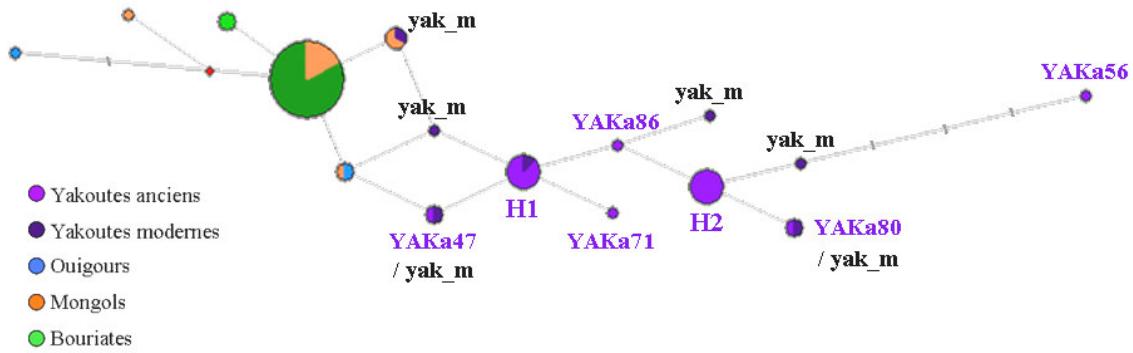


Figure 46 : *Median Joining* network des haplotypes STR N3 de différentes populations. (yak_m : yakoute moderne).

IV.4.4 Distances génétiques et analyse multidimensionnelle

Afin de pouvoir effectuer des comparaisons élargies à l'ensemble des sujets anciens analysés et inclure un nombre de populations de référence plus important, nous avons réalisé des calculs de distances génétiques entre populations à l'aide du logiciel ARLEQUIN. Ces calculs de paire de distance F_{st} ont été effectués pour 10000 permutations en utilisant la méthode de Reynolds (1983). Les loci pris en compte dans cette analyse sont identiques à ceux utilisés lors de l'analyse network afin de conserver une homogénéité et une cohérence entre les différents tests statistiques.

Nous avons sélectionné 12 populations pour ces comparaisons en fonction de leur localisation géographique, des affinités possibles avec la population Yakoute mais également de la disponibilité des données.

Afin de tester différentes hypothèses nous avons réalisé trois calculs successifs en prenant en compte : (i) l'ensemble des sujets anciens (Annexe 12) ; (ii) la répartition chronologique de nos échantillons (Annexe 13) et (iii) la répartition géographique des tombes (Annexe 14).

La matrice de F_{st} réalisée avec les sujets anciens pris dans leur ensemble montre des distances significatives avec toutes les populations incluses dans l'analyse. En revanche la comparaison avec la population yakoute actuelle donne une valeur négative ? que l'on peut considérer comme nulle. Il apparaît donc que ces deux populations ne sont pas différenciables par cette méthode révélant une grande homogénéité des lignées paternelles au cours du temps.

Afin de visualiser ces résultats sous forme graphique une analyse multidimensionnelle a été réalisée à l'aide du logiciel SPSS8.00. Les résultats sont présentés dans la figure 47 ci-dessous. Il apparaît clairement que la population yakoute ancienne ainsi que la population actuelle sont séparées des autres groupes. Les deux points ne sont pas confondus car la représentation graphique prend en compte l'ensemble des distances. Celles-ci étant plus importantes entre la population Yakoute moderne et les autres populations que celles obtenues pour les sujets anciens, le point représentant la population moderne s'en trouve excentré.

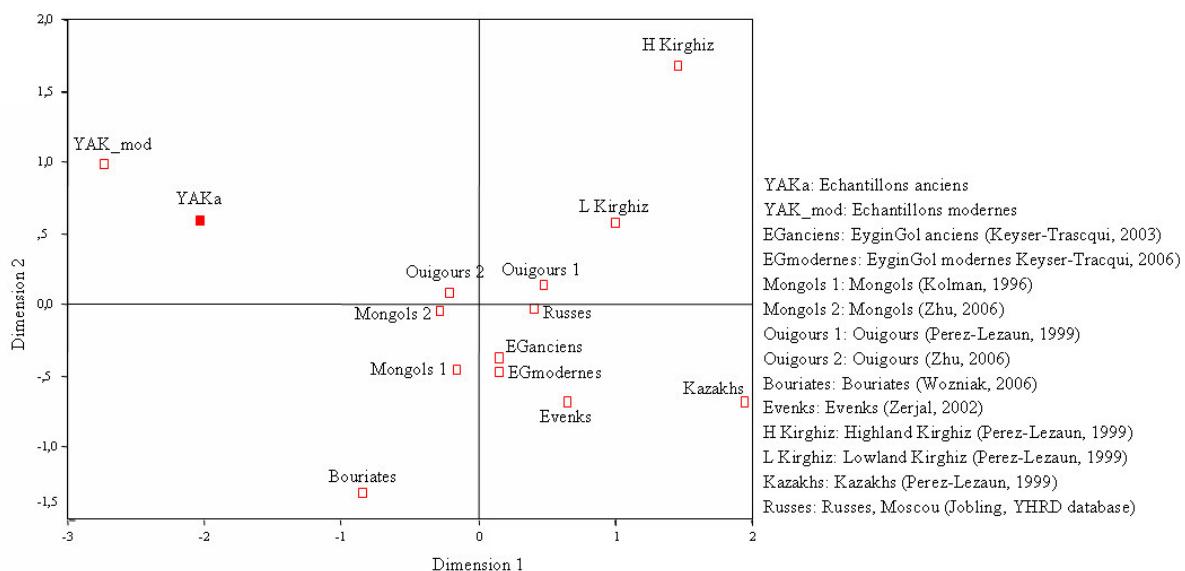


Figure 47 : Analyse multidimensionnelle basée sur les paires de distances F_{st} calculées à partir des haplotypes de 7 STR du chromosome Y.

Le calcul des paires de distances F_{st} dans le test réalisé sur la répartition chronologique de nos échantillons a mis en évidence des distances significatives pour toutes les comparaisons à l'exception des comparaisons au sein de la population yakoute ancienne et entre ces différentes périodes et la population actuelle.

L'analyse multidimensionnelle (Figure 48) met en évidence la séparation entre les yakoutes et les autres populations. Les différences entre le groupe ancien de la deuxième période et les valeurs obtenues pour la population moderne sont plus importantes que celles des périodes 1 et 3 ; il en découle une séparation dans la représentation graphique bien que les différences ne soient pas significatives entre ces 4 groupes.

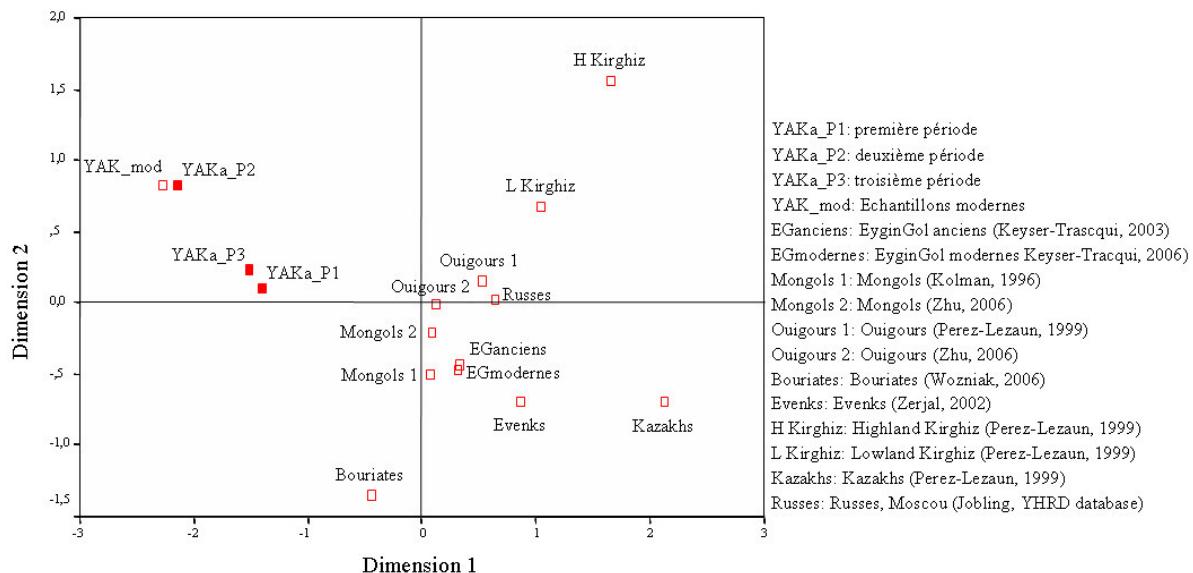


Figure 48 : Analyse multidimensionnelle chronologique basée sur les paires de distances F_{ST} calculées à partir des haplotypes de 7 STR du chromosome Y.

Les analyses effectuées en tenant compte de la répartition géographique des tombes mettent en évidence la même différenciation vis à vis des populations comparées sauf avec la population yakoute actuelle avec laquelle les sous-populations provenant des trois *ulus* ne montrent pas de différences significatives.

En revanche, la comparaison entre les différentes sous-populations archéologiques montre des distances significatives entre les populations de l'*ulus* Tattinsky et les sujets provenant des *ulus* Churapchinsky et Khangalassky. La représentation graphique de l'analyse multidimensionnelle présentée dans la figure 49 montre bien la séparation entre les différents sous-groupes géographiques au sein de l'ensemble des sujets anciens et la position intermédiaire des Yakoutes actuels reflétant l'absence de distance significative avec les Yakoutes anciens.

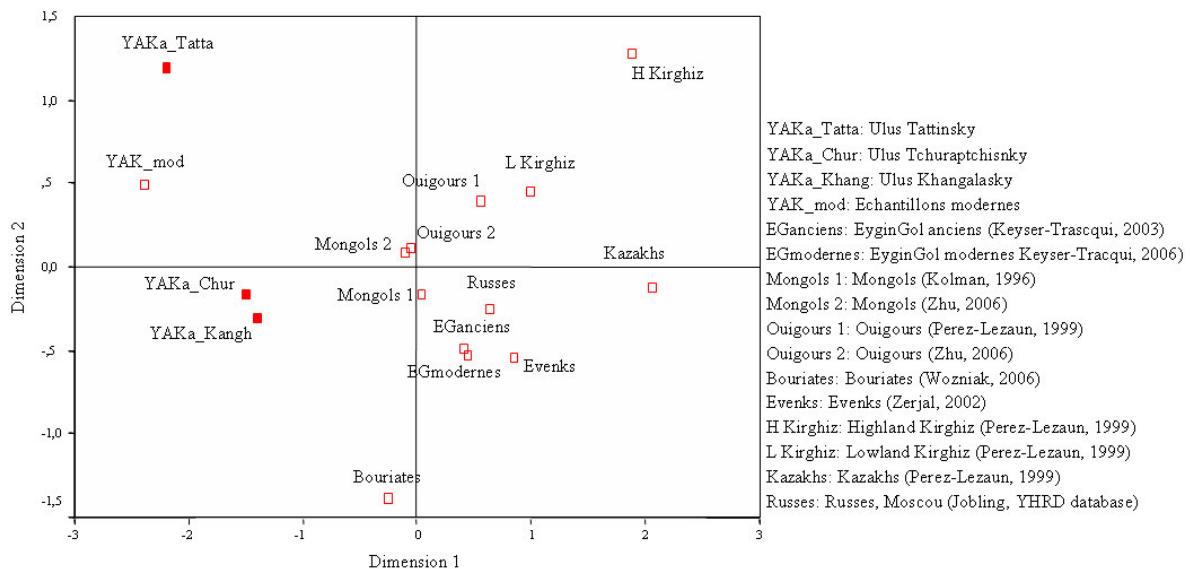


Figure 49 : Analyse multidimensionnelle géographique basée sur les paires de distances F_{st} calculées à partir des haplotypes de 7 STR du chromosome Y.

Les sujets provenant de l'*ulus* Khangalasky appartiennent en majorité à l'haplogroupe H2 alors que les lignées paternelles de l'*ulus* de Tattinsky sont en majorité affiliées à l'haplogroupe H1. Les sujets de la région de Churapchinsky présentent pour quatre sujets des haplotypes appartenant à l'haplotype H2 et deux lignées paternelles qui ne sont pas présentes dans la population yakoute actuelle.

IV.5 L'ADN mitochondrial

IV.5.1 Résultats généraux et paramètres de diversité

Le séquençage de la région HV1 a permis d'obtenir des séquences validées sur différentes amplifications et plusieurs extractions indépendantes pour 61 de nos sujets anciens. Nous avons décidé d'exclure 6 sujets de l'analyse. En effet, pour YAKa28, YAKa42 et YAKa84 il n'a pas été possible d'obtenir de séquence amplifiable ; pour les sujets YAKa49 et YAKa72 la région HV1a n'a pu être amplifiée et enfin pour YAKa48 la séquence obtenue présentait 6 hétéroplasmies impliquant une possible contamination. De plus, ce sujet provenant d'un musée il était impossible de comparer la séquence obtenue avec l'haplotype des personnes ayant été en contact avec les restes osseux. Il apparaissait, par conséquent, plus judicieux d'écartier cet échantillon.

Les hétéroplasmies observées dans notre échantillon ancien se sont limitées à certaines amplifications de 4 échantillons (YAKa45, YAKa48, YAKa58, YAKa59). Pour le sujet YAKa48 il n'a pas été possible d'obtenir de séquence fiable comme nous l'avons précisé plus haut. Pour les sujets YAKa45, 58 et 59 il n'a pas été possible de mettre en évidence si la présence de plusieurs hétéroplasmies était liée à une contamination puisque la comparaison entre les haplotypes des membres de l'équipe et la séquence obtenue n'a révélé aucune correspondance. Les séquences obtenues pour les autres extraits et dans le cas du sujet YAKa45 avec un autre substrat n'ont pas montré d'hétéroplasmies.

Lors du séquençage de certains échantillons il est arrivé qu'une hétéroplasmie soit détectée de manière ponctuelle (e.g. pour le sujet YAKa81 lors d'un séquençage réalisé à partir d'ADN extrait de restes osseux). Cependant, ces hétéroplasmies ponctuelles n'étant pas retrouvées dans les autres séquences du même sujet, elles n'ont pas été prises en compte pour la détermination de la séquence consensus.

La position 16093 :

Nous avons observé dans 2 de nos échantillons, YAKa20 et YAKa39, la présence d'une hétéroplasmie en position 16093. La duplication des amplifications à partir d'extractions différentes a confirmé la présence de deux populations mitochondrielles portant un C ou T pour cette position polymorphe. Pour YAKa20, l'ensemble des extractions a été réalisé à

partir d'os alors que pour YAKa39 nous disposions de trois types de substrats. La position 16093 s'est révélée être hétéoplasmique pour les séquences obtenues à partir des dents et des cheveux mais pas pour les extraits obtenus à partir d'os. Les séquences obtenues à partir de cheveux prélevés sur le sujet YAKa67 ont également révélé la présence d'une hétéoplasmie pour cette position.

Les données de la littérature (Meyer *et al.*, 1999 ; Stoneking *et al.*, 2000) révèlent que cette position est un *hot spot* de mutation, en outre les travaux de Brandstätter *et al.* (2003) ont confirmé que la fréquence des hétéoplasmies à cette position était importante dans les extraits d'ADN provenant de cheveux.

L'ensemble de ces données amène à penser que ces hétéoplasmies sont le fait de la dégradation de la molécule d'ADN plutôt que le fait de contaminations.

Les diversités haplotypiques et nucléotidiques observées chez les sujets anciens ont été comparées à celles observées dans différentes populations de la littérature (tableau 16).

La diversité haplotypique de l'ensemble des sujets anciens étudiés est plus faible que celle observée dans la population actuelle. Elle est également inférieure à la majorité des valeurs observées dans les populations comparées à l'exception des Evenks de l'Ouest, Itelmen et des Koryaks. De manière générale, il semble que les populations des steppes et les populations d'Asie Centrale présentent les valeurs de diversité haplotypique les plus importantes (environ 0,99), que les populations sibériennes comme les Evenks et les Koryaks présentent les valeurs les plus faibles (0,93-0,95) et que les Yakoutes et les Tuvas (population sibérienne vivant à l'ouest du lac Baïkal et présentant une forte influence turcique) présentent des valeurs intermédiaires (0,96-0,98) (Pakendorf *et al.*, 2003).

En considérant ce paramètre, l'ensemble des sujets anciens étudiés se rapprochent plus des populations sibériennes que des populations des steppes ou d'Asie Centrale. Il faut toutefois prendre en considération l'effectif de l'échantillon étudié qui pourrait être un facteur de biais. La diversité nucléotidique chez les sujets anciens se rapproche des valeurs observées dans la population yakoutes actuellemais également chez les Evenks, les Kirghiz et les Ouigours.

Population	Diversité haplotypique	Diversité nucléotidique
Altaï (Shields <i>et al.</i> , 1993)	0,993	0,0153
Bouriates (Pakendorf <i>et al.</i> , 2003)	0,996	0,0207
EyginGol anciens (Keyser-Tracqui <i>et al.</i> , 2006)	0,982	0,0176
EyginGol modernes (Keyser-Tracqui <i>et al.</i> , 2006)	0,991	0,0206
Evenks (Torroni <i>et al.</i> , 1993;Kaessmann <i>et al.</i> , 2002)	0,936	0,0186
Evenks de l'Ouest (Pakendorf <i>et al.</i> , 2006)	0,92	ND
Evènes (Derenko <i>et al.</i> , 1997)	0,950	ND
Itelmen (Schurr <i>et al.</i> , 1999)	0,930	0,0152
Kazakhs (Comas <i>et al.</i> , 1998)	0,990	0,0200
Kirghiz (Comas <i>et al.</i> , 1998)	0,988	0,0190
Koryaks (Schurr <i>et al.</i> , 1999)	0,934	0,0180
Mongols (Keyser-Tracqui <i>et al.</i> , 2006)	0,995	0,0208
Tuvas (Derenko <i>et al.</i> , 2000)	0,978	0,0190
Tuvas (Pakendorf <i>et al.</i> , 2006)	0,96	ND
Ouigours (Comas <i>et al.</i> , 1998)	0,993	0,0191
Yakoutes (Keyser-Tracqui <i>et al.</i> , 2006)	0,962	0,0185
Yakoutes (Pakendorf <i>et al.</i> , 2006)	0,96	ND
Yakoutes anciens	0,937	0,0186
Youkaguir (Pakendorf <i>et al.</i> , 2006)	0,940	ND

Tableau 16 : Comparaison des diversités haplotypiques et génétiques basées sur la région HV1 de l'ADN mitochondrial de différentes populations d'Asie Centrale et de Sibérie. ND : donnée non disponible.

IV.5.2 Vérification des séquences : construction d'un *Reduced Median network*

L'ensemble des séquences a été analysé dans un *Reduced Median network* afin de mettre en évidence les possibles erreurs de séquençage (figure 50). Les valeur de pondération des *hot spots* de mutations et des sites à évolution rapide ont été modifiées en tenant compte de la fréquence de chaque position dans notre population et les positions 16093 et 16362 ont été exclues de l'analyse suivant les recommandations de Bandelt *et al.* (2002).

Nous avons mis en évidence 3 réticulations liées à 4 positions polymorphes : 16129, 16171, 16298 et 16327. Pour chacune des positions concernées nous avons vérifié avec attention les électrophorégrammes. Les positions 16129, 16298 et 16327 sont des mutations qui surviennent fréquemment dans notre population, leur poids a donc été réduit ; cependant nous ne les avons pas supprimé car elles ne sont pas décrites comme des *hot spots* de mutation *postmortem* (Gilbert *et al.*, 2003) ou considérées comme des mutations fantômes fréquentes

(Brandstätter *et al.*, 2005). Ces mutations sont, en outre, présentes pour plusieurs individus dans des séquences obtenues à partir de différents substrats, validant ainsi leur présence.

Pour la position 16171, qui est présente chez cinq sujets anciens, ces mutations ont été contrôlées sur plusieurs extractions de différents extraits et les haplotypes comportant cette position sont également retrouvés dans plusieurs populations actuelles. Nous avons donc validé ces deux positions polymorphes qui sont également observée dans la population actuelle.

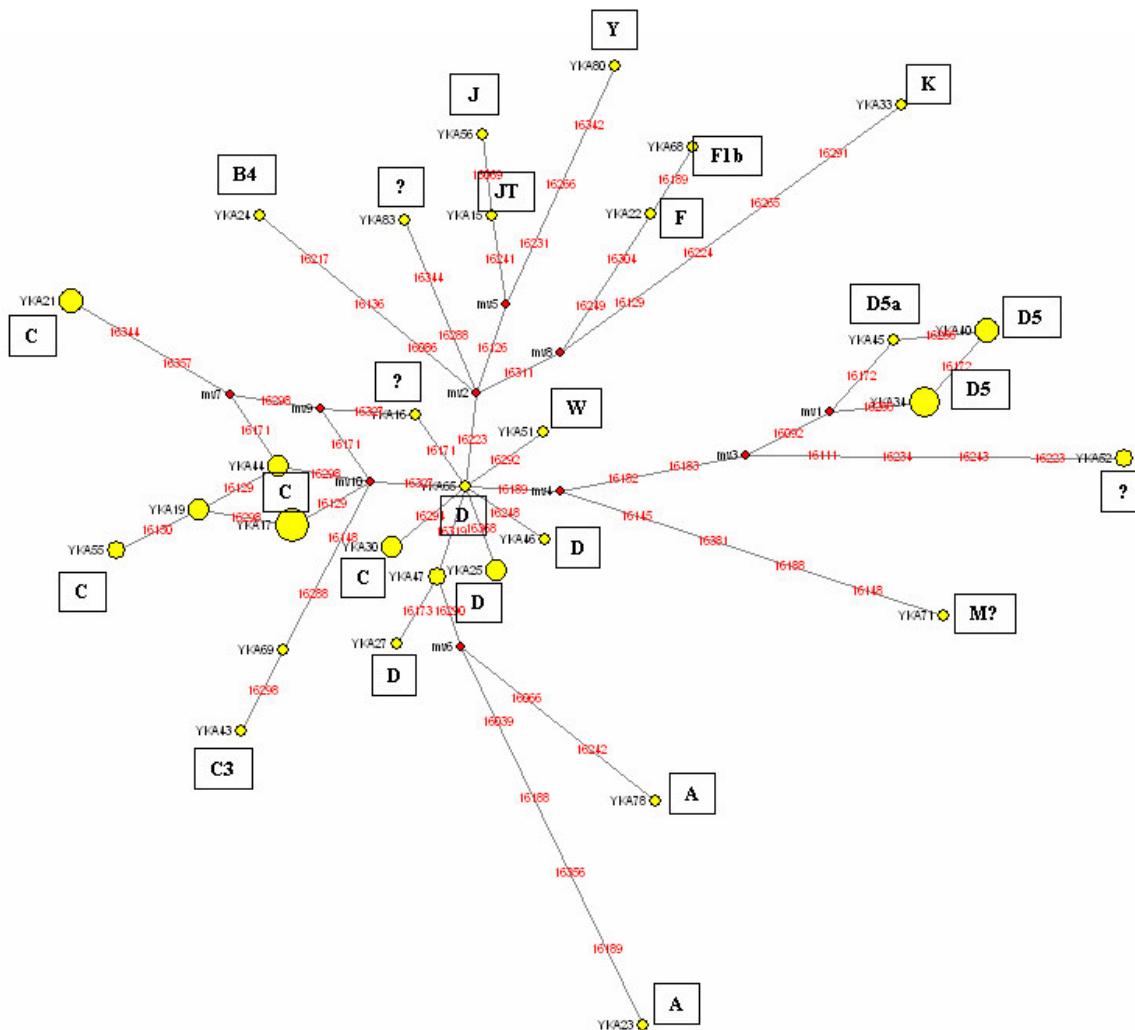


Figure 50 : *Reduced Median* network construit à partir des haplotypes mitochondriaux de la population Yakoute ancienne.

IV.5.3 Haplotypes partagés

Au sein de l'ensemble des sujets anciens nous avons trouvé 38 haplotypes différents parmi lesquels 11 sont partagés entre plusieurs sujets et 27 sont uniques. Les positions polymorphes sont résumées dans le tableau 17 ci-dessous.

n° Laboratoire	16092	16093	16111	16129	16150	16171	16172	16182	16183	16189	16223	16234	16243	16266	16294	16298	16319	16327	16344	16357	16362	16368		
CRS	T	T	C	G	C	A	T	A	A	T	C	C	T	C	C	T	G	C	C	T	T	T		
YAKa17, 31, 39, 50, 67, 75		C		A						T							T							
YAKa19, 60		C		A						T							C		T					
YAKa21, 32, 70, 86					G					T							C		T	T	C			
YAKa25, 29										T									C	C				
YAKa30, 64, 79										T							T				C			
YAKa34, 35, 36, 37, 38, 41	C						C	C	C	T				T						C				
YAKa40, 57, 59, 82	C						C	C	C	C	T			T						C				
YAKa44, 66, 73										T							C		T					
YAKa47, 54										T							A			C				
YAKa52, 53		T					C	C	C		T			C										
YAKa55, 76		A	T							T							C		T					

Tableau 17 : Haplotypes mitochondriaux partagés au sein de l'ensemble des sujets anciens.

L'ensemble des haplotypes obtenus a été recherché dans la base de données du laboratoire ainsi que dans les bases de données disponibles en ligne afin de déterminer si ils étaient retrouvés au sein d'autres populations. Les résultats synthétiques de cette recherche sont présentés ci-dessous (Tableau 18) et ne prennent en compte que les séquences présentant 100% d'homologie.

Haplotype	Origine des populations présentant des haplotypes identiques	Haplogroupe
YAKa15	NR	JT?
YAKa16	NR	D
YAKa17, 31, 39, 50, 67, 75	1 Bouriate, 15 Yakoutes	C
YAKa18	2 Bouriates, 3 Tofalars, 3 Yakoutes	H
YAKa19, 60	1 Sibérien ancien, 5 Bouriates, 2 Kazakhs, 1 Mansi, 1 Oroquen, 1 Tuvanian, 1 Ouigour, 4 Yakoutes	C2
YAKa20	10 Yakoutes, 1 YSE	C
YAKa21, 32, 70, 86	4 Bouriates, 3 Evenks, 1 Kazakh, 1 Mongol, 5 Oroquens, 5 Tuvas, 9 Yakoutes	C2
YAKa22	NR	F?
YAKa23	1 Kazakh, 1 Yakoute, 1 Evenk	A
YAKa24	1 Yakoute	B4
YAKa25, 29	2 Bouriates, 1 EyginGol ancien, 2 Kirghizs, 2 Mansis, 1 Oroquens, 2 Khantis, 2 Yakoutes	D
YAKa26	1 Sibérien ancien, 1 EyginGol ancien, 3 EyginGol, 1 Kirghiz, 1 Koryak, 1 Mansi, 1 Mongol, 1 Ouigour, 4 Yakoutes	C
YAKa27	1 Tchuche, 6 Esquimaux Sibériens, 2 Koryaks	D
YAKa30, 64, 79	1 Bouriate, 3 Evenks, 1 Tuvanian, 8 Yakoutes	C
YAKa33	1 Yakoute	K
YAKa34, 35, 36, 37, 38, 41	1 Chinois (Yao, 2003), 1 YSE	D5
YAKa40, 57, 59, 82	1 Daur, 21 Yakoutes	D5
YAKa43	2 Bouriates, 2 Kirghizs, 1 Oroquen, 1 Tuvanian, 3 Yakoutes	C3
YAKa44, 66, 73	1 Sibérien ancien, 8 Bouriates, 3 EyginGol, 1 Tchuche, 2 Daurs, 4 Evenks, 1 Kazakh, 2 Kirghizs, 6 Koryaks, 5 Mansis, 2 Mongols, 2 Tuvas, 1 Udegey, 1 Uzbek, 8 Yakoutes	C2
YAKa45	1 Chinois QuinDao (Yao, 2002), 3 Yakoutes, 1 YSE	D5a
YAKa46	NR	D
YAKa47, 54	1 Altaï, 1 EyginGol, 1 Evenk, 1 Kazakh, 1 Mansi, 3 Oroquens, 1 Tuvanian, 4 Yakoutes	D
YAKa51	1 Bouriate, 1 Kazakh, 3 Ouigours, 3 Yakoutes	W
YAKa52, 53	NR	?
YAKa55, 76	1 Bouriate, 4 Evenks, 1 Udegey, 4 Yakoutes	C2
YAKa56	3 Yakoutes	J?
YAKa58	sans hétéoplasmie en 16293: 7 Evenks, ,5 Tuvas, 11 Yakoutes , 1 YSE, 4 Youkaghirs	C
YAKa65	3 Yakoutes	D
YAKa68	1 Mansi, 3 Mongols (sans mutation en 16232: 3 Yakoutes)	F1b
YAKa69	2 Tuvas, 2 Tofalars (Starikovskaya, 2005), 3 Yakoutes	C3
YAKa71	NR	M?
YAKa74	1 Mansi	D
YAKa78	1 Bouriate, 1 EyginGol, 2 Yakoutes	A
YAKa80	NR	Y
YAKa81	1 Evenk (Starokovskaya, 2005)	C2
YAKa83	NR	?

Tableau 18 : Correspondances obtenues lors de la comparaison des haplotypes mitochondriaux des sujets anciens avec les bases de données. NR : non retrouvé ; YSE : Yakoute de langue evenk (*Yakut Speaking Evenk*) (Pakendorf et al. 2006).

Parmi les 38 séquences, 28 sont retrouvées dans la population yakoute actuelle soit 73% des haplotypes qui ont été retrouvés dans les échantillons modernes étudiés ou disponibles dans la littérature. Au sein des 10 haplotypes qui ne sont pas retrouvés dans la population yakoute actuelle, 8 ne sont présents dans aucune des bases de données dans lesquelles nous avons effectué nos recherches. L'ensemble des critères d'authentification mis en place, et notamment la validation par l'analyse network, permettent de s'assurer que les mutations mis en évidence pour ces différents haplotypes ne proviennent pas d'erreurs survenues lors de l'amplification. Par conséquent, l'absence de ces séquences dans les populations actuelles pourrait être en relation avec une disparition des ces haplotypes au cours des générations. Il faut également considérer que la non exhaustivité des bases de données disponibles pourrait être responsable de ce résultat.

IV.5.4 Détermination des haplogroupes

La méthode de comparaison par *near matching* a été utilisée pour déterminer l'affiliation aux différents haplogroupes de nos sujets. Pour les séquences qui n'ont pas été retrouvées lors des comparaisons nous nous sommes servi des positions diagnostiques des différents haplogroupes. Toutefois pour les sujets YAKa52-53 et YAKa83 il n'a pas été possible de déterminer leur appartenance à un haplogroupe par cette méthode (Tableau 18). Néanmoins nous avons reporté sur le network (Figure 50) les différents haplogroupes que nous avions pu déterminer. Il est ainsi possible de proposer l'affiliation de ces sujets à leur macro-haplogroupe : l'haplotype des sujets YAKa52-53 doit appartenir au macro-haplogroupe est-eurasiatique M alors que le sujet YAKa83 semble plutôt être rattaché au macro-haplogroupe ouest eurasiatique R.

La distribution des fréquences des différents haplogroupes est présentée dans le tableau 19. Cette distribution est comparable à celle observée au sein de la population Yakoute actuelle. D'une manière générale, les haplogroupes classiquement retrouvés dans les populations asiatiques représentent 90% des haplotypes des sujets anciens. Les haplogroupes majoritaires sont les haplogroupes C et D qui sont généralement les plus fréquents dans les populations de Sibérie Centrale et Méridionale mais également en Mongolie et Bouriatie (Starikovskaya *et*

al., 2005 ; Pakendorf *et al.*, 2006). Les individus les plus anciens que nous avons étudiés (YAKa26, 27, 43, 46, 47) appartiennent à ces deux haplogroupes.

Haplogroupe	Sujets anciens		Population actuelle	
	Effectif	Fréquence (%)	Effectif	Fréquence (%)
"Asiatiques"		91,09%		92,3%
A	2	3,57	2	2,2
B	1	1,78	4	4,4
C	27	48,26	35	38,5
D	8	14,28	17	18,7
D5 et D5a	7	12,5	13	14,3
F	2	3,57	6	6,5
M	3	5,35	1	1,1
G			3	3,3
Y	1	1,78	2	2,2
Z			1	1,1
"Européens"		8,91%		7,7%
H	1	1,78	1	1,1
J	2	3,57	1	1,1
K	1	1,78	1	1,1
T			1	1,1
U			2	2,2
W	1	1,78	1	1,1
TOTAL	56	100	91	100

Tableau 19 : Distribution des haplotypes mitochondriaux chez les sujets anciens et la population moderne.

L'haplogroupe C est l'haplogroupe majoritaire dans de nombreuses populations sibériennes : Evenk, Tuvianian, Yukhagir, Nganassan et Tofalar et également dans la population Bouriate (Starikovskaya *et al.*, 2005 ; Pakendorf *et al.*, 2006). Les haplotypes de 12 sujets peuvent être regroupés dans le sous haplogroupe C2 qui apparaît comme le plus fréquent dans toute la Sibérie (Starikovskaya *et al.*, 2005) et majoritaire dans les populations citées ci-dessus. Ce sous haplogroupe représente 44% des séquences appartenant à l'haplogroupe C des sujets anciens étudiés. Les sujets YAKa43 et YAKa69 ont pu être affiliés au sous-haplogroupe C3, qui est retrouvé en Sibérie Centrale et Méridionale mais également parmi les populations de l'extrême nord et du nord-est (Starikovskaya *et al.*, 2005).

L'haplogroupe D, qui est le second haplogroupe le plus fréquent en Sibérie (Starikovskaya *et al.*, 2005), représente environ 27% de nos séquences anciennes. Le résultat le plus marquant est la fréquence très importante des haplotypes pouvant être classés dans le sous-haplogroupe D5 ou D5a qui correspondent à 87% des haplogroupes D de notre échantillon. En effet, les fréquences de D5 et D5a sont très élevées chez les sujets anciens et on retrouve des

fréquences similaires dans la population actuelle. La prépondérance de ce sous-haplogroupe contraste avec les populations de Sibérie puisque la fréquence de D5 et D5a est généralement assez faible : 1% chez les Mansis (Derbeneva *et al.*, 2002), 4,2% chez les Tubalars, 1,1% chez les Tuvas, 1,4% chez les Evenks (Starikovskaya *et al.*, 2005), 4,5% chez les Youkaguirs (Pakendorf *et al.*, 2006) voire inexistante. Les haplogroupes D5 et D5a sont également rares dans les populations des steppes : 2% chez les Mongols (Yao *et al.*, 2004), 4,1% chez les Ouigours (Comas *et al.*, 1998) et en Asie Centrale : 1,8% chez les Kazakhs, absent chez les Kirghiz (Comas *et al.*, 1998) et entre 1,8% et 3,4% chez les Uzbeks (Yao *et al.*, 2004). Les sujets appartenant à ce sous-haplogroupe sont présent dès le 15^{ième} siècle (Djoussoulen) et durant toutes les périodes jusqu'au 19^{ième} siècle.

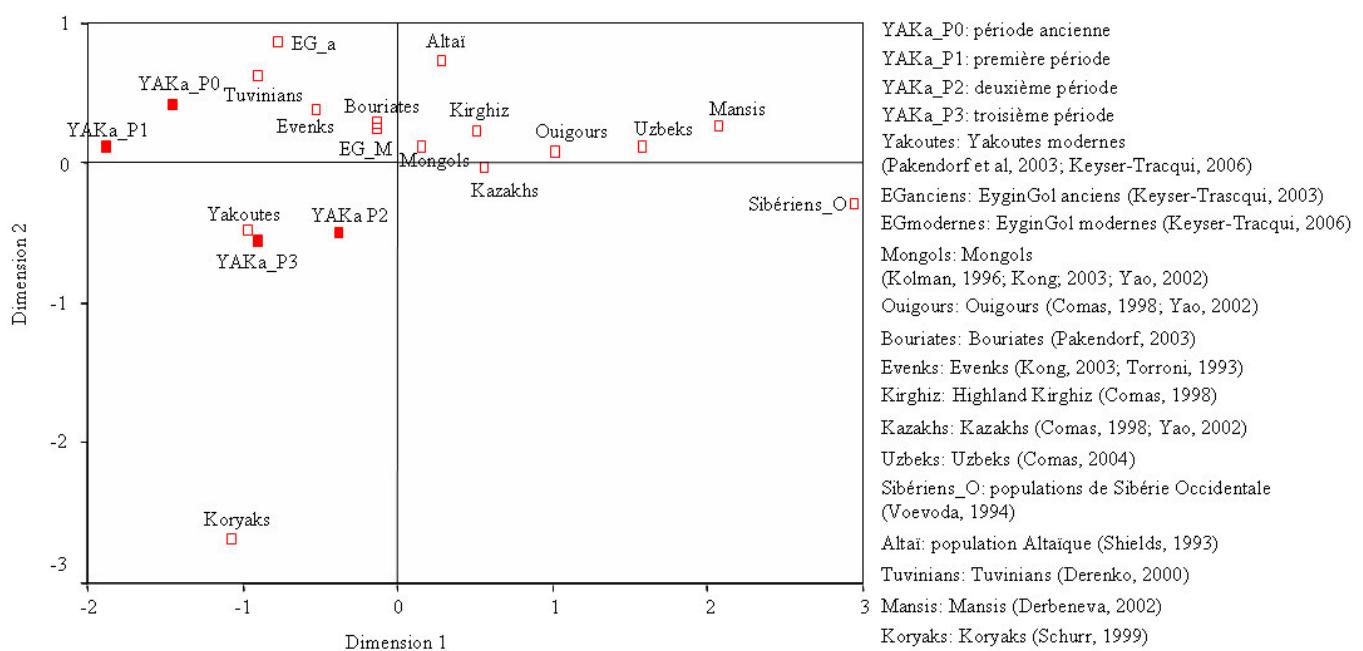
Les fréquences des haplogroupes A, B, F, G et Y sont relativement faibles dans notre échantillon. Ces haplogroupes sont répartis dans les différentes populations sibériennes avec des fréquences généralement peu élevées pour les haplogroupes A (sauf chez les Chukches 68,2% et les Eskimos 77,2%), B, F (excepté chez les Kets 23% pour l'haplogroupe F), G (sauf chez les Itelmens 68,1% et les Koryaks 41,9%) et Y (qui est restreint aux populations du Kamchatka et de l'Amour inférieur) (Starikovskaya *et al.*, 2005). Les populations des steppes du sud et d'Asie Centrale présentent également des fréquences faibles pour ces différents haplogroupes dont les fréquences sont toujours inférieures à 10% excepté pour l'haplogroupe G chez les Ouigours dont la fréquence atteint 10,8% (Yao *et al.*, 2004).

Enfin, nous avons observé la présence de quatre haplotypes répartis chez cinq sujets appartenant à des haplogroupes dits européens, tels que H, J, K, T et W. L'haplogroupe H présente des fréquences décroissantes d'ouest en est de 14,3% chez les Mansis à 4% chez les Bouriates (Starikovskaya *et al.*, 2005). Cet haplogroupe est retrouvé à des fréquences d'environ 10% chez les Ouigours et les Kazakhs (Yao *et al.*, 2004). Les autres haplogroupes sont généralement observés à des fréquences peu élevées dans les populations de Sibérie Occidentale et sont plus rares, voire absents, chez les populations toungousses et de l'extrême nord et nord-est (Starikovskaya *et al.*, 2005).

IV.5.5 Analyse multidimensionnelle

Des calculs de distances génétiques entre la population archéologique divisée suivant les critères temporels et géographiques et différentes populations disponibles dans notre base de données ont été réalisés. Nous avons appliqué la même méthode que celle utilisée pour les données des STR du chromosome Y. L'obtention de séquences exploitables pour les échantillons les plus anciens, qui n'avaient pas été inclus dans les analyses précédentes, nous a amené à définir une période supplémentaire dans notre répartition temporelle des sujets.

Les matrices de distances Fst sont données en Annexe 15 et la représentation de l'analyse multidimensionnelle est présentée en Figure 51 ci-dessous. Comme pour les lignées paternelles, les Yakoutes des différentes périodes ne présentent pas de différences significatives en fonction de l'époque considérée, ce qui semble attester d'une stabilité au cours du temps.



La comparaison avec les autres populations montre que pour les échantillons les plus anciens (YAKa_P0) les distances les plus faibles sont obtenues avec les populations suivantes : Bouriate, Egyn Gol ancien et moderne, Evenk, Mongol et Tuvinian. Les haplotypes présents

dans ce sous groupe sont parmi les séquences qui sont les plus fréquemment rencontrés dans les différentes populations de Sibérie. Il est intéressant de noter que ces échantillons appartiennent à une période antérieure à l'époque médiévale pour laquelle aucun contact entre les populations de Sibérie Centrale n'a été attesté.

Les sujets classés dans la première période médiévale montrent des affinités avec les populations Bouriate, Eygin Gol (ancienne et moderne), Evenks, Ouigours et Tuvas.

Pour la seconde période médiévale, nous avons observé des distances non significatives avec les mêmes populations que pour la période antérieure. Cependant les distances Fst font apparaître des affinités avec les populations Mongoles, Kazakhs et Kirghiz.

Cette proximité des lignées maternelles rencontrées dans notre échantillon ancien avec les Bouriates est renforcée par le fait que plus de la moitié des haplotypes anciens de notre population sont retrouvés chez les Bouriates actuels.

Enfin, les sujets classés dans la période postérieure au XVIII^{ème} siècle présentent des distances significatives avec l'ensemble des populations comparées dans cette analyse à l'exception des différents sous groupes Yakoutes, comme nous l'avons déjà mentionné, et des Tuvas. Cette distanciation par rapport aux différentes populations tend à rapprocher ce groupe de la population actuelle qui ne présente des affinités qu'avec les Yakoutes anciens, les Evenks et les Tuvas.

La deuxième analyse multidimensionnelle a été effectuée en prenant en compte la répartition géographique des sujets (figure 52), les matrices de distances sont présentées en Annexe 16.

Contrairement à l'analyse prenant en compte la répartition temporelle, les sujets découverts dans l'*ulus* Churapchinsky diffèrent des groupes anciens des deux autres régions. En outre, ce groupe présente des différences significatives avec l'ensemble des populations utilisées dans la MDS.

Les sous groupes des *ulus* de Tattinsky et Khangalassky ne présentent des différences significatives ni entre eux, ni avec la population actuelle. Leur comparaison avec les autres populations montre des différences non significatives avec les Bouriates, les Evenks et les Tuvas.

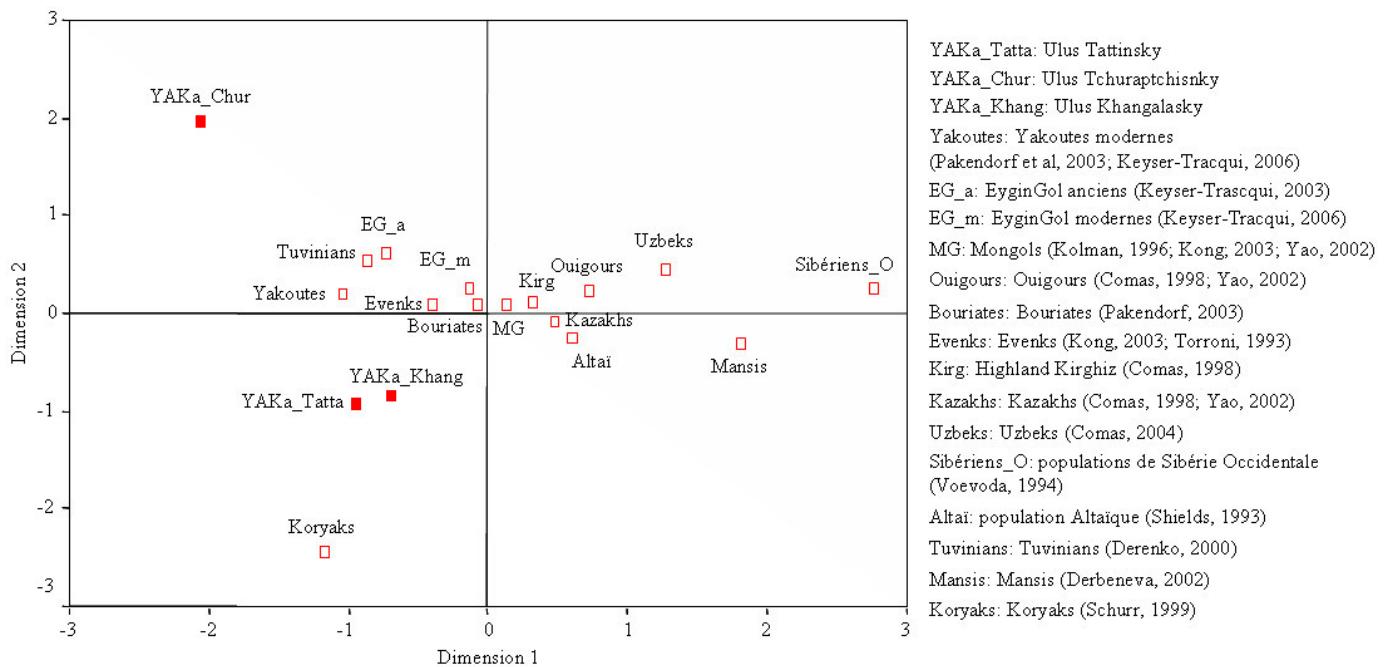


Figure 52 : Analyse multidimensionnelle basée sur les paires de distance F_{st} calculées à partir des haplotypes de la région HV1 des sujets anciens répartis en sous groupes géographiques.

V DISCUSSION

V.1 L'ADN ancien : approche fondamentale

V.1.1 Spécificités des échantillons anciens de Yakoutie Centrale

V.1.1.1 Authentification des résultats

La totalité des données obtenues sur les sujets anciens analysés dans cette étude a été comparée de façon systématique aux génotypes de l'ensemble des membres de l'équipe ayant participé à la fouille mais également aux analyses en laboratoire. Pour l'ensemble de nos échantillons ces comparaisons n'ont montré aucune correspondance, laissant supposer que les profils STR et les séquences mitochondrielles obtenus sont spécifiques des individus anciens. De plus, lors du clonage du sujet de Pokrovsk nous n'avons pas détecté de mélange de séquences parmi les clones amplifiés. Il semble que, même pour les échantillons provenant de musées et pouvant se révéler problématiques, le protocole de décontamination se soit avéré efficace.

Concernant l'ensemble des échantillons prélevés lors des fouilles réalisées par la MAFSO, le recours aux procédures de clonages n'a pas été jugé nécessaire. D'une part, en raison des précautions prises lors des fouilles qui permettent de réduire au maximum les possibilités de contamination durant cette phase, souvent considérée comme la plus problématique (Pääbo *et al.*, 2004; Willerslev and Cooper 2005 ; Sampietro *et al.*, 2006). D'autre part, grâce à la comparaison entre les profils obtenus pour les STR autosomaux dont le pouvoir de discrimination très élevé permet de garantir que l'ADN amplifié ne provient pas des membres de l'équipe ou d'une contamination par un autre sujet étudié au sein du laboratoire. Enfin, l'analyse de plusieurs types d'échantillons prélevés sur un même sujet constitue également une assurance concernant l'authenticité des résultats. Ces deux derniers critères, rarement présents dans les études d'anthropologie moléculaire, apparaissent comme de solides gages de validité.

Pour certains sujets anciens, nous avons observé une hétéroplasmie en position 16093. Or, un des expérimentateurs possède une transition T/C pour cette position. Les autres sites polymorphes présents dans la région HV1 du manipulateur n'ont toutefois pas été retrouvés

dans les séquences des sujets portant l'hétéroplasmie. Par ailleurs, aucune correspondance n'a été retrouvée pour les autres marqueurs entre cet expérimentateur et les sujets anciens. Cette position est, en outre, décrite comme un *hot spot* de mutation à la fois *ante* et *postmortem* ; nous en avons conclu que le taux de mutation élevé de la position 16093 devait être responsable de la présence de cette hétéroplasmie dans les séquences anciennes.

Lors de l'évaluation de l'efficacité de typage des différents kits d'amplification des STR autosomaux et du chromosome Y, nous avons observé de manière récurrente une diminution de l'efficacité de typage en fonction du poids moléculaire du marqueur. Cette caractéristique de l'ADN ancien, liée aux dégradations subies au cours du temps, est un des critères d'authentification des résultats (cf partie II.4. Critères d'authentification des résultats). Nos résultats satisfont également l'ensemble des différents paramètres nécessaires à la validation des résultats tels que : la concordance entre le sexe morphologique et génétique pour les sujets matures, l'absence de contamination dans les différents témoins négatifs d'amplification et la reproductibilité entre les différentes extractions et amplifications.

Concernant ce dernier critère, nous avons comparé les données provenant d'extractions réalisées à partir de différents substrats pour un même individu. De plus, les expériences pour les sujets de la deuxième et troisième année ont été réalisées par deux expérimentateurs différents, les extractions et amplification sur les échantillons pileux ayant été traités à part. La concordance des données obtenues à partir des échantillons osseux, dentaires et pileux d'un même sujet confirme encore l'authenticité de nos résultats.

Cette approche mettant en parallèle différents substrats permet, en outre, d'augmenter la proportion de résultats exploitables puisque la comparaison des données obtenues sur les différents marqueurs permet de générer un nombre plus important de profils et de séquences consensus pour les individus pour lesquels cette approche est permise.

Le sens phylogénétique des données apparaît également primordial pour s'assurer de l'authenticité des résultats. Les comparaisons avec les données portant sur les populations locales ainsi que la vérification des séquences mitochondriales par la méthode de *reduced median network* ont, de plus, permis de s'assurer que nos résultats n'étaient pas le fait de l'amplification d'ADN exogène ou de la recombinaison entre les deux demi-régions HV1.

V.I.1.2 Importance des conditions taphonomiques

Comme nous l'avons souligné dans le chapitre II.3.1.2, les conditions environnementales vont grandement influer sur la conservation ou la dégradation des acides nucléiques.

La Yakoutie Centrale se trouvant dans une zone de permafrost continu, la plupart des tombes étudiées sont gelées la majeure partie de l'année. Par ailleurs, suivant la profondeur de la fosse de creusement et l'exposition du site il arrive que les corps retrouvés soient partiellement gelés au plus fort de l'été. Cette température annuelle basse et surtout constante est responsable de l'état de préservation exceptionnel de l'ADN qu'il a été possible d'extraire à partir de nos échantillons. Nous avons ainsi pu obtenir des profils STR autosomaux complets pour plus de la moitié de nos échantillons anciens y compris pour les tombes de plusieurs milliers d'années telles la tombe de Pokrovsk. Il faut souligner que ces résultats sont uniques en comparaison des données qu'il est habituellement possible d'obtenir à partir d'échantillons anciens.

Le permafrost présente également d'autres conditions favorables à la conservation de l'ADN endogène des sujets anciens. Les dommages hydrolytiques, qui peuvent être responsables d'erreurs lors de l'amplification, semblent réduits du fait de la présence d'eau sous forme gelée et non liquide dans le permafrost. Ainsi, peu d'erreurs ont été observées lors du séquençage des échantillons anciens et il semble qu'aucune mutation fantôme ne soit présente dans les séquences mitochondriales obtenues.

Les fouilles de sauvetage réalisées sur le site de Bekh Alas pourraient confirmer cette hypothèse puisqu'un nombre important d'hétéoplasmies ont été observées pour deux des échantillons de ce site. Les tombes ayant été découvertes éventrées, les os avaient séjourné au contact des eaux de ruissellement durant une période indéterminée ce qui a pu favoriser ce type de dégradation.

Les pratiques funéraires des Yakoutes pourraient également avoir joué un rôle dans la préservation de l'ADN. La présence de multiples couches d'écorce de bouleau recouvrant les cercueils ou les coffres permet une isolation du corps vis à vis de l'eau mais également de l'oxygène. Par ailleurs, le permafrost présent en Sibérie Orientale a été décrit comme anaérobie du fait d'une concentration en méthane importante (Willerslev *et al.*, 2004) ce qui permettrait de limiter les lésions oxydatives mais également le développement de certains

champignons qui sont parfois responsables de fragmentations du brin d'ADN lors de la décomposition des corps. On observe d'ailleurs, après l'ouverture des tombes, le développement rapide de moisissures sur les parois des cercueils et les vêtements des sujets inhumés (Crubézy, données non publiées).

Lors de la mise au jour des tombes, les corps de 11 sujets ont été retrouvés lyophilisés en raison des basses températures et du faible taux d'humidité. Les processus de décomposition pour ces tombes doivent être particuliers et il est possible que cette lyophilisation, qui permet une préservation exceptionnelle des tissus mous, soit également impliquée dans la préservation de l'ADN. Il est en effet reconnu que la dessiccation des tissus permet une meilleure préservation de l'ADN (Pääbo, 1989 ; Lindhal, 1993) Ces 11 sujets ont tous présenté des profils STR autosomaux complets indiquant la bonne préservation de l'ADN. Il a également été possible de comparer les résultats entre les corps lyophilisés et les corps squelettisés au sein d'un même site, limitant ainsi le facteur de biais dû à des variations dans les conditions taphonomiques. Pour les sites d'Arbre Chamanique, Jarama et Ken Ebe, différents types de conservation ont été observés. Les corps lyophilisés ont permis d'obtenir de meilleurs résultats montrant, pour ces sites, une corrélation entre la préservation générale du corps et celle de l'ADN.

Nous avons vérifié, de manière fortuite, que les conditions de stockage des échantillons après la mise au jour et le prélèvement des échantillons étaient également primordiales. En effet, les échantillons pileux du sujet YAKa34 ayant été stockés à température ambiante pendant la mise au point du protocole d'extraction, nous avons pu observer une différence notable entre la qualité des profils STR obtenus lors de la première extraction et celle obtenue lors de la seconde extraction survenue trois mois plus tard. La première extraction a permis d'obtenir des profils complets alors que les amplifications pratiquées sur le deuxième extrait présentaient des absences d'amplification pour plusieurs loci. Une troisième extraction a, par conséquent, été pratiquée sur les cheveux qui avaient été conservés à -20°C et les résultats obtenus étaient de qualité similaire à ceux de la première extraction.

L'ensemble des prélèvements et des poudres obtenues à partir des os ou des dents doivent être conservés à -20°C de manière à prévenir les risques de dégradations postérieurs à la fouille.

V.1.2 Types de prélèvements et qualité des résultats

V.1.2.1 Les dents : "coffre fort" à ADN

En dépit de quantité de matériel de départ parfois réduite à un gramme de poudre, les dents ont permis d'obtenir les extraits les plus fortement concentrés en ADN puisque 89% des extraits ont montré une concentration supérieure à $100\text{pg}.\mu\text{l}^{-1}$ contre 38% pour les échantillons osseux et seulement 27% pour les cheveux. Nos résultats confirment, par conséquent, le postulat selon lequel les dents permettent une meilleure préservation de l'ADN que les autres tissus pour des conditions taphonomiques identiques (Kurosaki *et al.*, 1993 ; Oota *et al.*, 1995 ; Zierdt *et al.*, 1996 ; Ricaut *et al.*, 2005).

Les données obtenues à partir d'extraits provenant de dents montrent également des efficacités de typage proches de 100%. La taille des fragments qu'il est possible d'amplifier à partir de l'ADN extrait de dents semble supérieure à celle habituellement attendue pour des échantillons anciens (300pb) puisque les marqueurs de plus haut poids moléculaire sont amplifiés avec succès. L'incidence des inhibiteurs paraît également réduite puisque pour deux sujets, dont la quantification révélait la présence d'inhibiteurs dans les extraits obtenus à partir d'os, aucune inhibition n'a été observée pour ceux provenant de dents.

V.1.2.2 Les os

Concernant les prélèvements osseux, cette étude a mis en évidence que la fréquence des échantillons pour lesquels les concentrations d'ADN étaient inférieures à la limite de détection de l'appareil ($23\text{pg}.\text{l}^{-1}$) est constante suivant le type d'échantillons osseux pris en compte. Par conséquent il semble que les dégradations affectent l'os indépendamment de l'âge du sujet ou de la nature de l'os.

Toutefois, nous avons observé des écarts notables entre les rendements d'extraction des différents types de prélèvements. En effet, pour les sujets jeunes, nouveaux-nés et jeunes enfants, les quantités d'ADN qu'il est possible de récupérer sont moindres par rapport aux restes osseux provenant de sujets adultes. Le degré de minéralisation plus faible des os chez les enfants pourrait être responsable de cette différence. Si les interactions entre les cristaux

d'hydroxyapatite et l'ADN sont un facteur déterminant de la préservation de l'ADN, une relation entre le degré de minéralisation des tissus osseux et la quantité d'ADN récupérable pourrait exister. Cependant, des travaux plus approfondis sur ces interactions seraient nécessaires pour valider cette hypothèse.

Pour les sujets adultes, nous avons, en outre, mis en évidence que les fragments osseux permettaient d'obtenir de plus grandes quantités d'ADN que les diaphyses d'os longs. Ces deux types d'échantillons présentant globalement des degrés de minéralisation similaires, on peut penser que la différence doit porter sur un autre paramètre. Notre attention s'est portée sur la méthode de pulvérisation utilisée car seule cette étape diffère lors de l'extraction. Pour les diaphyses d'os longs, nous employons la méthode de trépanation alors que les fragments osseux sont cryobroyés. Malgré les précautions prises lors de la trépanation il se pourrait que des échauffements se produisent entraînant une dégradation de l'ADN par rupture des liaisons phosphodiesters (Lindhal, 1993). En revanche, la pulvérisation dans l'azote liquide permettrait de s'affranchir totalement de cette augmentation de température délétère. Le broyage dans l'azote liquide pourrait également renforcer les différences entre les résultats obtenus à partir d'os ou de dents. Des expériences mettant en parallèle les deux procédures pour les mêmes échantillons seront menées afin de vérifier ces premiers résultats.

Pour 8% des extraits d'ADN obtenus à partir d'os, une co-purification d'inhibiteurs a été mise en évidence lors de la quantification. La présence de substances inhibitrices n'a été observée que pour les substrats osseux indiquant une certaine protection des dents et des cheveux vis à vis des inhibiteurs. Le fait que les dents utilisées dans notre étude aient été extraites directement des alvéoles dentaires lors de la mise au jour et non pas retrouvées dans le sédiment peut également être à l'origine de cette absence d'inhibiteurs.

La dilution des échantillons inhibés a permis d'obtenir des résultats pour l'ensemble des sujets qui présentaient une inhibition à l'exception du sujet YAKa53. Ce résultat négatif pourrait être en relation avec le principal désavantage de la dilution des inhibiteurs : la dilution concomitante de l'ADN présent dans l'extrait. Pour des sujets dont l'ADN est trop dégradé ou trop faiblement concentré cette technique se révèle par conséquent inappropriée.

V.I.2.3 Les cheveux

Les analyses réalisées à partir de cheveux ont montré, pour la première fois, qu'il était possible d'extraire de l'ADN nucléaire analysable à partir de tiges dépourvues de bulbe. Ces résultats attestent de l'état de préservation de l'ADN des sujets anciens découverts en Yakoutie Centrale. En effet, bien que le développement d'un protocole d'extraction adapté joue un rôle dans ces résultats, il faut noter que la qualité des échantillons de départ est un facteur important dans cette réussite. Cette étude a souligné, grâce à la mise en parallèle de deux protocoles d'extraction incluant, ou non, une étape de lavage, que de l'ADN nucléaire devait être présent dans la tige du cheveux. Ceci va à l'encontre des études précédentes (McNevin *et al.*, 2005b) qui postulaient que seules les cellules piégées dans les écailles de la cuticule permettaient d'obtenir de l'ADN nucléaire amplifiable.

Comme nous l'avons souligné plus haut, aucun inhibiteur n'a été détecté lors de la quantification des extraits issus de cheveux. Nos résultats sont en accord avec les études précédentes qui mettaient en avant une résistance accrue des cheveux par rapport aux contaminants (Gilbert *et al.*, 2006b).

V.I.2.4 Influence du bromure de N-Pénacyl Thiazolium lors de l'extraction

La comparaison des valeurs de concentration pour les échantillons ayant été traités par le tampon de lyse incluant, ou non, du PTB montre que ce réactif ne semble pas affecter le rendement d'extraction.

En revanche, concernant l'efficacité de typage, nous avons observé une augmentation notable de la qualité des résultats lors de l'analyse des échantillons extraits avec ce tampon. Nous avons constaté une diminution des absences d'amplification d'un facteur deux et une augmentation du nombre de marqueurs avec des génotypes complets dans des proportions équivalentes. La diminution des pertes d'hétérozygotie est patente, toutefois les autres artéfacts d'amplification, typages erronés et amplification de faux allèles, conservent des fréquences similaires suivant les deux protocoles employés. On peut considérer que ces deux types d'erreurs ne sont pas fonction de la qualité de l'ADN mais générés par le nombre de cycles de PCR qui est utilisé. En effet, la très grande sensibilité du protocole d'amplification

suivit dans notre étude peut entraîner l'amplification de débris cellulaires ou de chromosomes isolés (Taberlet *et al.*, 1996 ; Gill *et al.*, 2001 ; Kloosterman *et al.*, 2003).

Il semble que ce réactif permette une amélioration de la qualité de l'ADN extrait. Ce résultat pourrait être en relation avec le type de dégradation qui affecte les échantillons préservés dans le permafrost. L'étude de Hansen *et al.* (2006) a en effet montré que l'alkylation, et donc la formation des produits de Maillard, était la principale dégradation rencontrée pour ce type d'échantillons. Le PTB, en dissociant ces complexes sucres-protéines, permettrait de limiter les absences d'amplification provoquées par la formation des liaisons croisées et par conséquent d'augmenter le succès des analyses.

V.1.2.5 Amplification totale du génome

Les tests menés sur nos échantillons avec le kit Genomophi™ ont montré que lors de l'amplification une quantité importante de fragments nucléotidiques aspécifiques pouvaient être générés. La présence de ces produits aspécifiques en grande quantité constitue un inconvénient. En effet, un excès de molécules de petite taille peut provoquer une inhibition de la PCR (Rogan et Salvo 1990 ; Pusch *et al.* 1998 ; Ricaut *et al.*, 2005).

Les quantifications qui ont été réalisées sur les produits d'amplification obtenus avec le kit Genomiphi™ montrent des quantités d'ADN inférieures aux concentrations des extraits de départ. Il semble vraisemblable que ce rendement soit un artéfact lié à l'absence d'amplification du fragment cible du kit Quantifiler™ localisé en position 5p15.33 (Green *et al.*, 2005). En effet, les valeurs de Ct observées pour l'IPC des différents échantillons n'indiquent pas la présence d'inhibiteurs. De plus, la présence sur le gel d'ADN de haut poids moléculaire est patente. Ce biais a été observé dans plusieurs études (Dean *et al.*, 2001 ; Panelli *et al.*, 2002 ; Ballantyne *et al.*, 2005) et est expliqué par l'absence d'amplification des régions télomériques lors de l'utilisation de protocoles d'amplification du génome.

L'ensemble de ces artéfacts d'amplification est lié aux caractéristiques des matrices en faible nombre de copies puisque il semble que l'utilisation d'ADN dégradé mais présent en quantité importante (de 5 à 50ng dans la solution de départ) permette tout de même d'obtenir des résultats satisfaisants (Ballantyne *et al.*, 2006).

De manière générale, il semble que l'utilisation de ce kit ne soit pas adaptée aux spécificités des études sur l'ADN ancien. En effet, la présence de ruptures de brin, même peu fréquentes,

et les concentrations inférieures à $5\text{ng}.\mu\text{l}^{-1}$ rendent nos échantillons impossibles à amplifier à l'aide de la Taq Phi29.

V.1.2.6 Efficacité des kits d'amplification STR

L'évaluation des efficacités de typage respectives des kits AMPflSTR® Profiler Plus™ et AMPflSTR® Identifiler™ montre que, pour l'analyse des échantillons osseux, le comportement des deux multiplexes diffère. Pour le kit Profiler Plus™ on constate, en raison de la corrélation entre l'efficacité de la PCR et la taille des marqueurs amplifiés, une diminution graduelle du nombre de profils complets obtenus avec l'augmentation de poids moléculaire des amplicons. Les artefacts d'amplification et les absences d'amplifications se maintenant à des fréquences faibles (inférieures à 10%) pour l'ensemble des loci de taille inférieure à 250pb.

Pour le kit Identifiler™, les proportions de profils complets sont globalement plus faibles et le nombre d'absences d'amplification beaucoup plus élevé. Il faut toutefois considérer que les analyses pratiquées avec le kit Identifiler™ se sont déroulées ces deux dernières années sur des échantillons différents de ceux testés avec le kit Profiler Plus™. Ceci peut entraîner un biais important attendu que la proportion d'échantillons n'ayant pas permis d'obtenir de résultats était plus importante durant cette période.

La comparaison des résultats des kits Power Plex® Y et AMPflSTR® Y-Filer™ a révélé des divergences importantes entre les efficacités de typage des deux kits pour les loci de haut poids moléculaire notamment. Pour le kit Power Plex® Y la corrélation entre la taille des marqueurs et le succès de l'amplification est clairement visible. La courbe polynomiale des génotypes complets montre une réduction de 80% à 50% alors que les absences d'amplification augmentent de manière linéaire avec la taille des marqueurs (Annexe 17). Seul le locus DYS389II montre une efficacité de typage légèrement supérieure à celle attendue. Cette divergence pourrait s'expliquer par un biais lié au nombre réduit d'amplifications prises en compte pour l'évaluation.

Pour le kit AMPflSTR® Y-Filer™, on observe une efficacité de typage supérieure ou égale à 60% pour l'ensemble des loci sauf pour DYS439 et DYS390. La corrélation entre la taille du marqueur et le succès d'amplification est beaucoup moins marquée. La pente de la courbe

polynomiale pour les génotypes complets est en effet moins forte et la différence observée entre les marqueurs de faible et de fort poids moléculaire est seulement de 10% (85%-75%) (Annexe 17). Trois loci ne s'inscrivent pas dans cette tendance : DYS19, DYS439 et DYS390. Ces marqueurs présentent les efficacités de typages les plus faibles de ce kit alors qu'ils se placent dans une gamme de taille intermédiaire. Il est difficile de déterminer les raisons exactes de cette variation et aucune donnée de la littérature ne fait état d'amplifications différentielles pour ces loci.

L'approche méthodologique des résultats collectés durant ce travail nous a permis de souligner la très grande qualité des échantillons provenant de Sibérie Orientale. Il a ainsi été possible d'étudier des marqueurs rarement analysables dans les études d'anthropologie moléculaire comme les STR autosomaux et du chromosome Y, mais également d'obtenir des résultats uniques sur des substrats difficiles comme les cheveux. Cet ensemble de facteurs a conduit à l'obtention de données dont l'authenticité est manifeste, permettant d'établir des comparaisons fiables dans une optique phylogéographique.

Le nombre important de sujets étudiés et la mise en parallèle, sur plusieurs types d'échantillons, de nouveaux protocoles ont permis d'apporter des informations originales sur les propriétés de ces différents substrats. Ainsi, la méthode de cryobroyage des tissus durs et l'utilisation d'échantillons dentaires apparaissent à privilégier. En outre, cette étude a mis en évidence que certaines méthodes étaient totalement inadaptées à l'étude de l'ADN ancien et, en revanche, a confirmé l'utilité de certains réactifs, comme le PTB. L'ensemble de ces données nous a également conduit à envisager certaines perspectives afin d'optimiser encore les protocoles ainsi que l'analyse des résultats

V.1.3 Perspectives méthodologiques

V.1.3.1 Protocole d'extraction d'ADN

Concernant les protocoles d'extraction employés dans cette étude, plusieurs points sont apparus pour l'amélioration de la qualité des résultats.

Tout d'abord, l'étude de Salamon *et al.*, (2005) met en avant la possibilité d'obtenir de l'ADN de bonne qualité par la dissolution des agrégats de cristaux d'hydroxyapatite présents dans les os. L'utilisation d'un protocole conduisant à une déminéralisation complète des échantillons osseux semble donc judicieuse puisque la dissolution totale de la poudre d'os pourrait permettre l'accès à ces cristaux et ainsi augmenter le rendement de l'extraction. Loreille *et al.*, (2006) ont optimisé un tampon de lyse qui permet de dissoudre entièrement les échantillons osseux grâce à des concentrations élevées d'EDTA (0,5M) et du N-Laurylsarcocinate de sodium. Les autres réactifs entrant dans la composition du tampon ainsi que les étapes d'extraction organique, de purification et de concentration des extraits sont identiques à ceux employés actuellement au laboratoire. Cette étape de déminéralisation complète permettrait d'obtenir des rendements trois fois supérieurs à ceux observés avec les protocoles classiques et ne nécessite qu'un gramme de poudre d'os. Cette méthode représenterait un avantage certain puisqu'il serait possible d'obtenir de l'ADN de meilleure qualité, en quantité plus importante, tout en utilisant moins de matériel. Ce protocole est actuellement en cours d'évaluation au laboratoire.

V.1.3.2 Quantification de l'ADN mitochondrial

Cette étude a permis de montrer l'intérêt de la quantification de l'ADN nucléaire à la fois pour ajuster au mieux les volumes d'extraits et le nombre de cycles à utiliser pour les PCR mais également pour déterminer la présence d'inhibiteurs dans les extraits.

Le rapport de plusieurs centaines, voire milliers, de copies d'ADN mitochondrial par copie d'ADN nucléaire permet de conclure que, dans les gammes de concentrations d'ADN nucléaire amplifiables, suffisamment d'ADNmt doit être présent. Cependant, pour les extraits

non quantifiables il n'est pas possible de déterminer si la quantité d'ADNmt permettra l'amplification des séquences d'intérêt. En outre, la quantification de l'ADN mitochondrial de manière concomitante à celle de l'ADN nucléaire permettrait de déterminer, grâce à la comparaison des données, si une préservation différentielle des deux génomes existe.

Une étude récente (Walker *et al.*, 2005) a présenté une méthode permettant de quantifier en duplex l'ADN nucléaire et mitochondrial par l'utilisation de sondes Taqman pouvant être utilisée sur la plateforme de détection présente au laboratoire. La taille des séquences cibles correspond à celle du kit Quantifiler™ puisque les fragments amplifiés pour l'ADNmt et nucléaire mesurent respectivement 79pb et 71pb. Il serait ainsi possible d'évaluer par cette méthode les gammes de concentrations d'ADNmt autorisant une amplification optimale de la région HV1 et également de déterminer si les sujets pour lesquels nous avons des difficultés à amplifier une des deux demi-régions présentent des quantités trop faibles d'ADNmt.

V.1.3.3 Les STR de faible poids moléculaire

L'évolution des techniques employées pour l'identification des individus en médecine légale tend vers l'analyse d'échantillons de plus en plus dégradés. Les développements méthodologiques dans ce domaine profitent par conséquent à l'anthropologie moléculaire puisque les problèmes de dégradation de l'ADN sont encore plus cruciaux qu'en médecine légale.

Ces dernières années, plusieurs études (Coble et Butler, 2005 ; Dixon *et al.*, 2005 ; Grubwieser *et al.*, 2005) ont été publiées sur l'utilisation de STR dont le poids moléculaire est réduit par rapport à ceux présents dans les kits commerciaux. Les amores sont dessinées afin de se rapprocher le plus possible des séquences cibles et ainsi réduire la taille des amplicons.

Applied Biosystems va prochainement (Janvier 2007) commercialiser un kit comprenant 9 loci. Les résultats préliminaires du kit AMPflSTR® MiniFiler™, présentés au congrès "DNA in Forensic" d'Innsbruck, ont montré que la sensibilité de cette PCR multiplexe permettait l'obtention de profils complets à 90% pour des quantités d'ADN de départ de $60\text{pg}.\mu\text{l}^{-1}$. De plus, les tests réalisés en présence d'inhibiteurs montrent que ce kit est beaucoup moins sensible aux inhibiteurs que les kits standards tels Identifiler™.

Lors de l'étude d'échantillons anciens et des recherches des liens de parentés entre sujets d'un même ensemble funéraire, nous sommes bien souvent confrontés à la conservation

différentielle de l'ADN suivant les sujets en raison des micro-variations taphonomiques, de l'âge des individus ou du type d'échantillon analysé. L'utilisation de ces mini-STRs représente donc une approche prometteuse, applicable au chromosome Y, qui permettra d'accéder à l'information génétique d'un plus grand nombre d'individus.

V.1.3.4 Détermination des haplogroupes mitochondriaux par la technique SNaPshot

L'affiliation des haplotypes mitochondriaux à un haplogroupe se révèle parfois problématique par la seule analyse de la région HV1. Sur les échantillons provenant de populations modernes il est possible d'utiliser des enzymes de restrictions afin d'étudier un ensemble de positions polymorphes présentes dans la région codante de l'ADNmt et qui permettent, en combinaison avec l'analyse des mutations de la région HV1, de déterminer de manière fine l'haplogroupe d'un sujet. La technique des RFLP n'est pas applicable à l'ADN ancien du fait de la quantité importante d'ADN nécessaire à l'ensemble des analyses et des problèmes de digestion partielle qui compliquent l'interprétation des résultats et nécessite la répétition des expériences, augmentant encore les quantités d'ADN utilisées (Salas *et al.*, 2005).

Les succès obtenus au sein du laboratoire avec la technique SNaPshot pour le typage des SNP du chromosome Y permettent de penser que la méthode pourrait être appliquée aux SNP de la région codante de l'ADN mitochondrial. Par ailleurs, cette technique est déjà utilisée avec succès en criminalistique et en génétique des populations (Salas *et al.*, 2005 ; Brandstätter *et al.*, 2006). On peut également penser que le taux de réussite pourrait être encore plus élevé que pour le chromosome Y du fait de la présence en plus grand nombre de copies de l'ADN mitochondrial.

Une étude récente (Lee *et al.*, 2006) a montré que l'application de cette méthode à des échantillons osseux et dentaires provenant de sujets inhumés il y a une cinquantaine d'années permettait de déterminer, de manière fiable, l'appartenance aux différents haplogroupes. Lee et collaborateurs ont mis au point trois réactions multiplexes qui assurent le typage des principaux haplogroupes et sous haplogroupes présent en Asie de l'est. Une approche similaire est en cours pour les populations d'intérêt du laboratoire sur une sélection de SNP informatifs présents dans les haplogroupes les plus fréquemment rencontrés, mais également pour ceux dont l'affiliation est problématique à partir des seules séquences HV1.

V.2 Le peuplement de la Yakoutie Centrale

V.2.1.1 Le recrutement des ensembles funéraires

L'analyse des relations de parenté au sein des sites regroupant plusieurs sujets ne permet pas d'établir un modèle général concernant le recrutement des ensembles funéraires que nous avons fouillé. En effet, il semble que les critères de recrutement puissent être basés sur trois niveaux : l'appartenance à une même famille de sujets partageant des liens de proches parentés ; l'appartenance à un même groupe familial de sujets possédant la même lignée paternelle ou maternelle et enfin l'appartenance à un même clan d'individus ne partageant aucune relation de parenté. Ce dernier mode de recrutement rejoignant les conclusions proposées par Ricaut *et al.* (2006) sur le regroupement d'individus appartenant à une même structure clanique mais ne partageant pas forcément des liens de proche parenté.

Le site d'Arbre Chamanique illustre bien, au sein d'un même site, cette diversité des modes de recrutement. Le regroupement des quatre sujets au sein de la sépulture multiple d'Arbre Chamanique 1 semble avoir été motivé par les relations de proche parenté unissant les différents individus. Le sujet de la tombe Arbre Chamanique 2 semble également partager des liens de parenté étroit avec les autres sujets bien qu'il ne soit pas inhumé dans la même sépulture ce qui pourrait être lié à un décès une date différente, qui n'a pu être précisée. La position de ces tombes devant un *alas* de grande importance et la richesse du mobilier de la tombe n°1 amènent à penser que ces différents individus pouvaient appartenir à un clan d'importance au sein duquel cette lignée maternelle aurait perduré. Enfin, le sujet de la troisième tombe présente une séquence mitochondriale différente de celle observée précédemment. La tombe est par ailleurs plus éloignée de la tombe n°1, localisée plus en lisière de forêt. Cette femme aurait pu appartenir au même clan sans toutefois partager de liens parentaux étroits avec les membres de ce même groupe inhumés à proximité. Cette dernière hypothèse n'est toutefois pas vérifiable et la possibilité que l'inhumation corresponde à celle d'un sujet n'ayant aucune relation avec les autres individus est également à envisager.

V.2.1.2 La période ancienne

Durant notre étude nous avons analysé sept tombes antérieures à la formation des Yakoutes dont les datations, bien que mal assurées pour la plupart, sont distribuées du Néolithique à l'Age du Bronze.

Les sujets YAKa27 et YAKa28 ont tous deux été mis au jour dans la région de la Kolyma dans l'extrême nord-est de la Yakoutie. Nous n'avons pu obtenir des informations exploitables que pour le sujet YAKa27 qui daterait du Néolithique. Les données des STR du chromosome Y étaient trop fragmentaires pour permettre une recherche BLAST dans les bases de données. Toutefois, la comparaison de la séquence HV1 dans notre base de données ainsi que dans les bases de données en ligne a donné plusieurs correspondances avec des populations actuelles. L'haplotype de ce sujet est retrouvé chez un sujet Tchouktche, 2 Koryaks et chez 6 Esquimaux vivants dans l'Arrondissement Autonome Tchouktche. Il semble que cet haplotype soit présent dans ces populations du nord-est sibérien depuis une période très ancienne. La position 16319 rapproche cet haplotype de l'haplogroupe D3 qui est dispersé dans de nombreuses populations de Sibérie mais toujours avec des fréquences réduites (Derbeneva *et al.*, 2002 ; Starikovskaya *et al.*, 2005).

La localisation des restes osseux des sujets collectés dans les musées de Yakoutsk correspond à deux *ulus* distincts : le sujet YAKa47 et du site de Diring Urekh (YAKa48) provenant de Yakoutie Centrale, de l'*ulus* Khangalassky, et les sujets du site d'Allalaïka (YAKa43) et de Tchersinsky (YAKa44) de la région de Vilyuy. Les résultats concernant les lignées paternelles de ces différents sujets n'ont permis la comparaison que pour les données obtenues pour le sujet YAKa47. L'haplotype de ce sujet n'est retrouvé dans aucune base de données et bien que son haplogroupe n'ait pu être déterminé avec précision, il n'appartient pas à l'haplogroupe N3 car il ne présente pas la mutation TAT-C. La séquence mitochondriale de ce sujet est affiliée à l'haplogroupe D et cet haplotype est retrouvé dans de nombreuses populations de l'ensemble de la Sibérie (Mansis, Tuvas, Evenks), d'Asie Centrale (Kazakhs) et de Mongolie (Eygin Gol).

Les données portant sur ces sujets qui occupaient la Yakoutie avant la formation des Yakoutes révèlent des affinités avec différentes populations actuelles considérant leurs lignées maternelles. Les sujets découverts dans la Vilyuy appartiennent à l'haplogroupe C qui est le plus largement retrouvé au sein de l'ensemble des populations sibériennes, en particulier

evenks, (Schurr *et al.*, 2004 ; Starikovskaya *et al.*, 2005) et le sujet YAKa47 a pu être affilié à l'haplogroupe D qui est également très répandu en Sibérie. Ces haplogroupes sont retrouvés à la fois dans les populations d'Asie Centrale mais également dans les populations de Sibérie du sud et de l'ouest (Bouriates, Tuvas, Mansis), les populations d'origine toungousse (Evenks), ou les populations du nord-est (Koryaks, Tchouktches). Cette très vaste répartition des haplogroupes C et D serait liée à l'expansion des premiers groupes humains dans ces régions durant le Paléolithique Supérieur (Schurr *et al.*, 2004) expliquant le nombre et la diversité des populations actuelles dans lesquelles sont retrouvées ces séquences. Les principales rivières de Yakoutie (Lena, Vilyuy et Aldan) étaient en effet utilisées comme voies de dispersion depuis la fin de la période Sartan (Pléistocène Supérieur) (Vasil'ev *et al.*, 2002) permettant une certaine mobilité des populations anciennes. L'haplotype du sujet néolithique YAKa27, retrouvé dans les populations actuelles de l'extrême nord-est et appartenant à un haplogroupe aujourd'hui partagé par la plupart des populations sibériennes, représente bien cette dispersion de lignées maternelles anciennes.

La tombe de Pokrovsk

Au sein de ce groupe de tombes anciennes, le sujet de Pokrovsk occupe une place particulière. Cette tombe, mise au jour à une soixantaine de kilomètres au sud de Yakoutsk, datée entre le III^{ème} et le IV^{ème} siècle avant notre ère, illustre les plus anciens contacts entre la région du lac Baïkal et de Mongolie et la Yakoutie Centrale (cf Article n°1). Le matériel archéologique présent dans cette tombe comprend un armement (cuirasse en os et pointe de flèche en silex) rappelant l'équipement des populations autochtones de Sibérie Centrale. Mais un perçoir en métapode de cheval ainsi qu'une pointe de flèche en fer ont également été découverts et ces objets sont caractéristiques des populations des steppes. Cette variété dans le mobilier archéologique évoque indéniablement des échanges entre les populations de chasseurs de Sibérie Orientale et les populations de pasteurs nomades des steppes du sud.

Les affinités génétiques du sujet de Pokrovsk reflètent également la possibilité de contacts entre ces deux zones géographiques. La séquence HV1 de Pokrovsk a en effet été retrouvée chez deux Bouriates, une femme inhumée dans la nécropole d'Eygin Gol au III^{ème} siècle avant notre ère, 1 Evenk, 2 Mansis, ainsi que chez 2 Yakoutes et 1 Yakoute ancien (YAKa29).

Les données génétiques permettent de préciser que les contacts entre les chasseurs de Sibérie et les éleveurs nomades des steppes du sud dépassaient le cadre des simples échanges commerciaux. La correspondance entre la séquence du sujet de Pokrovsk et des Bouriates

ainsi qu'un individu de la nécropole d'Eygin Gol, localisée plus de 2000km au sud, confirme l'hypothèse archéologique de contacts précoces entre ces régions et indique une origine méridionale de la mère de ce sujet. La correspondance entre un sujet inhumé dans une nécropole Xiongnu et des sujets originaires de Bouriatie s'explique par les expansions dans la région péri-Baïkale des tribus Xiongnu qui eurent lieu au III^{ième} siècle avant JC (Tsydendambaev, 1972 *in* Khitrinskaya *et al.*, 2001).

La proximité génétique entre les Yakoutes et les Evenks a été mentionnée dans de nombreuses études de génétique des populations (Pakendorf *et al.*, 1999; Pakendorf *et a.*, 2003; Derenko *et al.*, 2002; Puzyrev *et al.*, 2003). Le partage de la séquence du sujet de Pokrovsk pourrait être en faveur de contacts génétiques entre les populations des steppes et les éleveurs de rennes toungousses durant le III^{ième} siècle avant JC et, par conséquent, antérieurs à l'émergence et la migration des Kurykans ou des tribus de la culture *Kulun-Atakh* proposés à partir des données archéologiques.

Le fait que cet haplotype soit partagé avec des Yakoutes actuels est en faveur, d'une part, de la participation précoce au pool génétique yakoute de femmes originaires de régions plus méridionales et d'autre part, d'une certaine continuité des lignées maternelles dans les populations sibériennes comme précédemment proposé par Ricaut *et al.* (2005).

La correspondance de cette séquence avec deux Mansis des montagnes de l'Oural et deux sujets issus de populations de Sibérie Occidentale pourrait être en relation avec des flux géniques d'est en ouest le long du Ienisseï (Starikovskaya *et al.*, 2005). De plus, si on considère la proportion importante d'haplogroupes asiatiques dans la population mansi (Derbeneva *et al.*, 2002), cette similarité peut être mise en parallèle des expansions des tribus nomades depuis la Mongolie vers l'ouest durant la période Xiongnu.

V.2.1.3 *La période turco-mongole*

Les tombes les plus anciennes pouvant être rattachées aux débuts de la période médiévale yakoute, sont datées de la fin du XIV^{ième} siècle jusqu'au XVII^{ième} siècle. Les pratiques funéraires de cette période montrent une nette influence turco-mongole (Crubézy, 2006). Le mobilier associé aux défunt comprend différents éléments d'harnachement tels des selles, des étriers ou des mords et également différentes armes dont l'arc qui est fréquemment retrouvé. Tous les sujets présents dans les tombes de ce type étaient de sexe masculin, laissant penser

que les femmes et les enfants devaient être inhumés selon des pratiques différentes, probablement des *arangas* (Crubézy, données non publiées). La conservation plus aléatoire de ces tombes aériennes, associée au fait que lors de l'arrivée des Russes ces tombes furent pratiquement toutes détruites (elles étaient supposées favoriser la diffusion des épidémies) pourrait expliquer le fait que nous n'avons pas retrouvé de sujets féminins de cette période.

La plupart des sujets pour lesquels il a été possible de déterminer l'haplogroupe du chromosome Y appartiennent à l'haplogroupe N3, majoritaire à la fois chez les sujets yakoutes anciens et dans la population actuelle. Par ailleurs, les sujets des tombes de Batta Tcharana (YAKa66) et Bouogaryma (YAKa78) portent un des haplotypes les plus fréquents chez les Yakoutes. De plus, le sujet de Djoussoulen (YAKa40), daté du XV^{ème} siècle, appartient à l'haplotype le plus représenté dans l'échantillon ancien analysé et est également très fréquent dans la populations actuelle. Cet haplotype aurait donc été transmis durant cinq siècles avec des fréquences importantes. En revanche, on constate que ces deux haplotypes ne sont retrouvés chez aucune des populations incluses dans les bases de données, ils apparaissent comme spécifiques des Yakoutes.

L'haplotype du sujet du site de Tyty Bapyt (YAKa86) appartient à l'haplogroupe N3 mais est retrouvé chez 8 Evenks actuels. La fréquence de l'haplogroupe N3 dans ces populations est habituellement beaucoup moins importante que chez les Yakoutes, de l'ordre de 16% (Karafet *et al.*, 1999 ; Derenko *et al.*, 2006). Il est fort probable que cet haplotype ait été transmis par les Yakoutes dans la population Evenks par métissage. En effet, l'étude de Pakendorf *et al.* (2006) permet de se rendre compte que parmi la population parlant aujourd'hui la langue Evenk, de nombreux sujets appartiennent à l'haplogroupe N3 et pourraient être en fait d'ascendance yakoute.

Le sujet de la tombe de Balyktakh (YAKa26), daté entre 1420 et 1470 AD, est le seul n'appartenant pas à l'haplogroupe N3 puisque affilié à l'haplogroupe K. Cet haplogroupe est présent avec une fréquence d'environ 20% chez les Sojots (vivant au sud du lac Baïkal) et les Mongols (Derenko *et al.*, 2006). Ces deux populations présentent la plus forte fréquence d'haplogroupe K en Sibérie (les fréquences les plus élevées étant observées chez les Coréens). Considérant les résultats des STR du chromosome Y, l'haplotype de ce sujet a été retrouvé chez 3 sujets Mongols, 4 sujets vivants dans la vallée d'Eygin Gol, 1 Kazakh et 11 Turcs d'Anatolie. La lignée paternelle de ce sujet semble par conséquent fortement liée aux populations à la fois des steppes du sud et de l'Asie Centrale.

Il est intéressant de constater que l'ensemble des lignées paternelles anciennes appartenant à l'haplogroupe N3 sont aujourd'hui retrouvées dans la population yakoute, contrairement à l'haplotype du sujet YAKa26. La disparition de cette lignée paternelle pourrait être mise en relation avec les importantes réductions d'effectifs au sein de la population masculine. On peut toutefois penser que la transmission à des fréquences plus faibles de cet haplotype n'a pas permis de le détecter dans la population actuelle.

Les haplotypes mitochondriaux des sujets appartenant à cette période ont tous perduré dans la population actuelle. Les séquences des individus YAKa21/YAKa86 et YAKa66 appartiennent à l'haplogroupe C2 et sont présentes dans de nombreuses populations de Sibérie, de Mongolie, de Bouriatie et d'Asie Centrale.

Le sujet YAKa26 appartient également à l'haplogroupe C et on peut noter que sa séquence est retrouvée chez un sujet de la nécropole d'Eygin Gol ainsi qu'un sujet ancien de la région de Krasnoïarsk. La concordance entre les lignées paternelles et maternelles de ce sujet le rattachant à la nécropole d'Eygin Gol évoque des contacts entre cette région de la Mongolie et la Yakoutie Centrale, comme nous l'avons déjà mentionné plus haut.

L'haplotype du sujet de Djoussoulen, tombe datée du XV^{ième} siècle au C14, est retrouvé très fréquemment chez les Yakoutes actuels et chez un Daur (population de langue mongole vivant en Chine) mais dans aucune autre population. Cet individu appartient à l'haplogroupe D5 qui est aujourd'hui présent avec une fréquence très importante dans la population Yakoute. La distribution de ce sous haplogroupe est relativement restreinte. Les Yakoutes présentent la fréquence la plus importante de cette lignée au sein des populations de Sibérie et d'Asie Centrale, les fréquences observées en Asie Centrale et en Sibérie étant très faibles (Derenko *et al.*, 2004). L'étude de Pakendorf *et al.* (2006) a mis en évidence une fréquence importante de cet haplogroupe dans la population yakoute qui pourrait être liée à un effet fondateur survenu au XIII^{ième} siècle. Les lignées maternelles appartenant à cet haplogroupe seraient, selon les auteurs, à rattacher à la population Bouriate. La découverte d'un sujet du XV^{ième} siècle appartenant à cet haplogroupe est par conséquent en accord avec ces conclusions, confirmant la stabilité de cette lignée maternelle au cours du temps.

La séquence du sujet de Bouogaryma 1 (YAKa78), qui est daté de la fin de cette période (*e.g.* XVII^{ième} siècle), a été retrouvée chez un Bouriate actuel et chez un sujet de la vallée d'Eygin Gol. Les étriers déposés avec le corps évoquent la tradition des steppes ce qui est en accord avec une origine méridionale de ce sujet. Cette séquence est affiliée à l'haplogroupe A qui est présent avec des fréquences réduites dans de nombreuses populations de Sibérie Occidentale

et Centrale (Starikovskaya *et al.*, 2005). La fréquence la plus importante de cet haplogroupe est toutefois observée chez les Sojots (Derenko *et al.*, 2003). Les origines de ce sujet évoquent également la région du Baïkal, puisque les Sojots vivent dans la zone péri-Baïkal dans l'actuelle République de Bouriatie. Le fait que cet haplotype, affilié à un haplogroupe relativement rare en Sibérie, soit retrouvé à la fois chez un Yakoute ancien et chez un sujet de la période Xiongnu vient encore renforcer l'hypothèse sur des contacts entre ces deux zones géographiques.

Les résultats des analyses génétiques portant sur les sujets de cette première période nous renvoient clairement à une origine méridionale. Il semble que les premiers contacts établis durant la période Xiongnu, comme l'atteste le sujet de Pokrovsk, aient par la suite donné lieu à la migration de groupes d'éleveurs nomades originaires de la région de la Bouriatie et de Mongolie du nord. Les nouveaux arrivants se seraient installés et auraient donné lieu à l'émergence de la culture yakoute traditionnelle.

V.2.1.4 La période mongole

La deuxième période distinguable par les modes d'inhumation correspond à des tombes plus récentes, XVII^{ième} et XVIII^{ième} siècles. Ces tombes considérées comme de tradition yakoute classique, présentent un mode de construction utilisant des troncs évidés pour les sujets riches qui évoquent les tombes mongoles de la région péri-Baïkale ainsi que les sépultures Xiongnu de la vallée d'Eygin Gol (Crubézy, 2006). Outre le mode de construction des coffres, cette période se distingue surtout par la présence, dans le mobilier associé, d'éléments d'importation telles des perles, des vêtements en soie sans doute d'origine chinoise qui évoquent un changement culturel important.

L'analyse des lignées paternelles des sujets de cette période révèle que l'ensemble des haplotypes appartient à l'haplogroupe N3. Par ailleurs, les différents haplotypes sont répartis dans les deux lignées les plus fréquentes aujourd'hui. L'haplotype du sujet YAKa81, bien que différent de H1 et H2, est également retrouvé de nos jours.

Ces sujets du XVII^{ième} et XVIII^{ième} siècles présentent tous des lignées paternelles présentes dans la population Yakoute actuelle et dont deux déjà en place depuis le XV^{ième} siècle. On

peut considérer que les sujets mis au jour dans les tombes de la période Turco-mongole appartenaient à une certaine élite guerrière considérant les égards avec lesquels ils ont été inhumés et les offrandes présentes dans les tombes. Les sujets de cette seconde période semblent également appartenir à la classe sociale dominante après le passage à une économie basée sur l'élevage et dans laquelle le terroir était une source de richesse. Les sujets inhumés dans la tombe du site d'Arbre Chamanique et de Munur Urekh sont inhumés face à de vastes alas et l'opulence du mobilier associé est notoire dans ces deux cas. La tunique en soie et la chevalière de chef de clan pour la tombe de Munur Urekh et les quatre selles décorées, les manteaux brodés de perles ainsi que le nombre d'offrandes pour la tombe d'Arbre Chamanique attestent de l'importance sociale des sujets. De plus, les hommes de ces deux tombes ont été inhumés avec leurs armes : arc long et *palma* qui confirment un statut de guerrier et pourraient représenter un rappel des traditions antérieures. On peut, par conséquent, envisager que les descendants des personnages influents de la période turco-mongole ont continué à prospérer dans la société yakoute médiévale.

Considérant les lignées maternelles, nous avons observé la présence pour trois sujets de lignées pouvant être attribuées à des haplogroupes présents dans les populations ouest eurasiatiques et qui sont habituellement plus fréquentes dans les populations européennes. Il semble peu probable que ces haplogroupes européens soient en rapport avec un métissage avec des colons russes. En effet, ces tombes sont toutes antérieures à la colonisation effective de la Yakoutie. Le sujet du site de Ouhorai, daté entre 1650 et 1680, semble appartenir à l'haplogroupe JT bien que l'haplotype exact n'ait pu être retrouvé dans les bases de données. Cet haplogroupe n'est présent que dans les populations Bouriates, Tuvianes, Kakhasses et Tofalars (Derenko *et al.*, 2003). La séquence du sujet YAKa33 appartient à l'haplogroupe K qui n'est que très rarement représenté dans les populations sibériennes. L'haplotype de ce sujet n'a été retrouvé que chez un Yakoute. Le sujet YAKa18 appartient à l'haplogroupe H et sa séquence est retrouvée chez 3 Yakoutes, 2 Bouriates et 3 Tofalars. Cet haplogroupe est le plus fréquemment rencontré dans les populations d'Europe de l'ouest. Cependant, cet haplogroupe est présent à des fréquences réduites, 3,5% en moyenne dans de nombreuses populations sibériennes (Derenko *et al.*, 2003) et dans la population mongole à une fréquence de 10% (Starikovskaya *et al.*, 2005).

La présence de ces séquences appartenant à des haplogroupes ouest-eurasiatiques pourrait être le fait d'une dispersion lors du Paléolithique Supérieur de populations du Moyen-Orient et

d'Europe du sud-est qui n'auraient pas été gommées par les migrations et les flux géniques ultérieurs (Starikovskaya *et al.*, 2005).

Les autres sujets se répartissent dans les haplogroupes considérés comme plus spécifiques des populations est-eurasiatiques C et D.

La fréquence importante de l'haplogroupe C dans l'échantillon ancien analysé pourrait être en relation avec des mariages entre les Yakoutes et les populations evenks au sein desquelles l'haplogroupe C est prédominant (Starikovskaya *et al.*, 2005 ; Pakendorf *et al.*, 2006). Comme nous l'avons déjà souligné dans le paragraphe concernant le sujet de Pokrovsk, plusieurs études attestent de la proximité entre les haplotypes mitochondriaux yakoutes et evenks. Ces similarités génétiques évoquent le partage de lignées maternelles communes à l'origine des pools génétiques actuels (Derenko *et al.*, 2002).

Plusieurs haplotypes sont également partagés avec les populations Bouriates, Tuvas et Mongols. Les Tuvas reflètent bien la dualité des influences mongoles et toungousses des populations de Sibérie. En effet, cette population vivant à l'ouest du lac Baïkal, présente des fréquences très importantes de séquences appartenant à l'haplogroupe C et des affinités génétiques marquées avec les Evènes. Cependant, du point de vue culturel l'incidence mongole est très nette (Golubenko *et al.*, 2001).

Cette dispersion importante des haplotypes de l'haplogroupe C dans les populations turciques et une population toungousse pourrait refléter des mélanges ayant eu lieu avant la migration des Yakoutes vers le nord. Ce pool génétique commun partagé par ces populations pourrait être en relation avec les éleveurs de rennes appelés *Uraangkhaj* mentionnés par Ksenofontov et Okladnikov. Ces populations de la région péri-Baïkale et d'ascendance toungousse auraient subis une forte influence des tribus turciques et mongoles avant d'être partiellement assimilés et/ou partiellement repoussés vers le nord par les Kurykans. La dispersion de cette population métissée anciennement pourrait expliquer le partage de ces haplotypes entre ces différentes populations.

Les haplotypes appartenant à l'haplogroupe D des sujets YAKa16 et YAKa46 n'ont pu être retrouvées dans aucune des bases de données disponibles. Concernant les séquences des sujets du site d'Arbre Chamanique (YAKa34, 35, 36, 37, 38, et 41) et d'Okhtobout 2 (YAKa82) nous avons pu déterminer leur affiliation au sous haplogroupe D5. A l'exception de la population Yakoute ces deux haplotypes sont peu partagés.

Ceci évoque, d'une part, que ces lignées maternelles ont perduré durant 5 siècles dans la population yakoute, confirmant la stabilité de la structure génétique des Yakoutes. D'autre part, les séquences affiliées à l'haplogroupe D ne sont que très peu partagées avec d'autres populations contrairement à celles de l'haplogroupe C. Par ailleurs, les fréquences supérieures à 10% rencontrées pour le sous-haplogroupe D5 semblent être relativement spécifiques de la population Yakoute (Pakendorf *et al.*, 2006). A l'instar du chromosome Y, il semble qu'un effet fondateur au cours du XII^{ième} siècle aurait pu conduire à la présence de ces haplogroupes spécifiques dans la population Yakoute (Pakendorf *et al.*, 2006).

V.2.1.5 La période chrétienne

Au sein des sujets mis au jour dans les tombes du XVIII^{ième} et XIX^{ième} siècle, les sujets masculins appartiennent majoritairement à l'haplogroupe N3. Les haplotypes des sujets découverts dans la tombe de Dirilaa Sayliga (YAKa56) et Ken Ebe 4 (YAKa71) n'avaient pas été détectés parmi les sujets des périodes précédentes. Tout comme les haplotypes N3 décrits précédemment, la comparaison avec les bases de données s'est révélée infructueuse confirmant la spécificité des lignées N3 de la population Yakoute ancienne. L'haplotype du sujet YAKa56 n'a pas été retrouvé dans la population moderne, il est possible que cette lignée paternelle soit de nos jours présente dans la population Yakoute mais avec une fréquence faible expliquant ainsi qu'elle n'ait pas été détectée dans notre étude. On peut également considérer que les réductions d'effectifs de la population Yakoute, notamment au cours des déportations survenues sous le régime soviétique ou lors des Première et Deuxième Guerres Mondiales, auraient pu favoriser la disparition de certaines lignées paternelles.

L'haplotype des sujets YAKa17 et YAKa19 n'est pas affilié à l'haplogroupe N3 puisqu'il ne présente pas la mutation TAT-C caractéristique. Cet haplotype est partagé avec 3 sujets Mongols, 19 Bouriates et est particulièrement fréquent dans la population Kalmyk dont les lignées paternelles indiquent une forte ancestralité mongole (Nasidze *et al.*, 2005). La lignée paternelle de ce sujet est, par conséquent, fortement ancrée dans les populations des steppes du sud.

Pour le sujet YAKa57 du site de Bekh Alas, tombe n°1, l'analyse des SNP du chromosome Y a permis l'affiliation à l'haplogroupe P. Cet haplogroupe est observé dans la plupart des populations sibériennes à l'exception des Evenks et des Teleuts (Derenko *et al.*, 2003) mais la

fréquence la plus importante de cet haplogroupe est observée chez les Tuvas (Wells *et al.*, 2002 ; Derenko *et al.*, 2003). Nous avions noté lors de la fouille une robustesse extrême de ce sujet qui présentait un torus mandibulaire et palatin avec des proportions exceptionnelles ainsi qu'un crâne plutôt brachycéphale. On observe donc une concordance entre ces caractéristiques, plus fréquemment rencontrées chez les populations d'origine mongole et le fait que cet haplotype soit très fréquent dans une population ayant subit une forte influence mongole.

Parmi les séquences mitochondrielles présentes durant cette période, de nombreux haplotypes sont partagés avec les sujets des deux époques antérieures. Parmi ces séquences, deux ne sont retrouvées dans aucune des bases de données utilisées. Ces séquences pourraient marquer l'arrivée de nouveaux migrants en Yakoutie Centrale durant cette période. Il n'est toutefois pas possible d'établir de liens préférentiels avec une région donnée. En effet, ces séquences sont retrouvées chez des populations de Mongolie (Mongols, Bourriates) ou d'Asie Centrale (Kazakhs), mais également dans des populations de Sibérie (Mansis, Evenks). Il faut également considérer qu'en raison de l'effectif plus important de ce groupe, nous aurions pu mettre en évidence ces séquences qui n'avaient pas été détectées au préalable.

Sept des dix séquences non décrites auparavant sont retrouvées dans la population yakoute actuelle ce qui, comme précédemment, évoque une stabilité des lignées maternelles yakoutes au cours du temps.

Trois des haplotypes sont classés dans l'haplogroupe C et sont partagés avec de nombreuses populations. Nous avons également retrouvé deux haplotypes appartenant à l'haplogroupe D, dont la séquence du sujet de la tombe d'Arbre Chamanique 3 qui est affiliée au sous haplogroupe D5a. Il faut noter que les cinq autres séquences appartiennent à des haplogroupes qui sont plus rares dans la population Yakoute en particulier des haplogroupes qui sont plus fréquents dans les populations du sud de la Sibérie : B, F et W.

Le sujet YAKa24 présente les sites polymorphes caractéristiques de l'haplogroupe B4 qui est largement répandu dans les populations mongoles et chinoises (Kolman *et al.*, 1996 ; Yao *et al.*, 2000 et 2002) mais également dans la population tuviniane (Starikovskaya *et al.*, 2005). Les origines supposées de cet haplogroupe se retrouvent à la fois en Asie Centrale et en Asie du sud-est (Starikovskaya *et al.*, 2005).

La séquence de l'enfant découvert dans la tombe Jarama 2 (YAKa51) est partagée avec des sujets originaires d'Asie Centrale (Kazakh et Ouigours) et également avec un individu

originaire de Bouriatie. Cet haplotype a été rattaché à l'haplogroupe W qui est peu représenté dans la population Yakoute actuelle.

Enfin, le sujet de la tombe Ken Ebe 7 porte un haplotype retrouvé chez un Mansi et nous avons pu déterminer qu'il appartenait à l'haplogroupe F. Cet haplogroupe est généralement rencontré dans les populations de Sibérie de l'ouest et chez les Tuvas mais également chez les populations Evenks de la région du Ienisseï (Derbeneva *et al.*, 2002 ; Starikovskaya *et al.*, 2005).

Les influences génétiques qui sont détectées dans la population yakoute du XVIII^{ième} et XIX^{ième} siècle, contemporaine de la colonisation russe, ne dénotent pas en comparaison des périodes précédentes. Il semble donc que les caractéristiques génétiques de la population yakoute soit en place dès le XV^{ième} siècle et perdurent jusqu'à nos jours. De plus, la fréquence d'haplogroupes européens n'augmente pas parmi les sujets inhumés durant cette période. L'influence russe sur le pool génétique yakoute apparaît comme relativement limité.

V.2.1.6 Approche globale des lignées paternelles anciennes

Un regard d'ensemble sur les résultats que nous avons pu obtenir pour les lignées paternelles présentes dans les échantillons anciens étudiés met en avant la très forte proportion de l'haplogroupe N3 ainsi qu'une diversité génétique faible en comparaison des populations mongoles ou centrasiatiques, à l'exception de la population bouriate actuelle. Cette tendance rapproche les sujets anciens de la population moderne actuelle, indiquant une stabilité importante des lignées paternelles depuis ce qu'il est permis de considérer comme les premiers Yakoutes.

Cette stabilité est confirmée par la grande homogénéité des différents groupes chronologiques que nous avons distingué. Les analyses multidimensionnelles ont, en effet, mis en évidence l'absence de différenciation entre ces différents sous groupes et cela malgré les modifications dans les pratiques funéraires et donc les changements culturels sous-jacents.

L'analyse network basée sur les STR du chromosome Y révèle que l'ensemble des haplotypes composant cet haplogroupe est très homogène. Les haplotypes ne sont séparés les uns des autres que par un évènement de mutation révélant une expansion rapide à partir d'un nombre réduit de fondateurs. On peut dès lors supposer que la colonisation de la Yakoutie Centrale fut

le fait d'un nombre réduit de sujets dont l'origine commune devait se situer en dans la région du lac Baïkal ou dans le nord de la Mongolie.

Par ailleurs, plus de la moitié des haplotypes anciens du chromosome Y sont regroupés dans deux lignées paternelles qui sont retrouvées de nos jours dans la population actuelle avec des fréquences équivalentes. Ces lignées paternelles présentes dès le XV^{ième} siècle chez des sujets appartenant à une élite guerrière ont été transmises, malgré les changements culturels, aux sujets de la classe sociale dominante chez les Yakoutes dits traditionnels du XVII^{ième} siècle. Les origines de ces lignées paternelles apparaissent chez les cavaliers des steppes du sud, Bouriates de la région du Baïkal mais également chez les nomades de Mongolie.

La prédominance et la pérennité de ces deux lignées paternelles pourraient être mises en relation avec le fait que les *tojons*, chefs de clans de la période yakoute médiévale, qui appartenaient à ces lignées auraient eu une descendance plus importante du fait de la pratique répandue de la polygamie dans cette classe sociale. L'arrivée des Russes et les décrets qui furent édictés renforçèrent encore l'autorité de ces chefs de clans en leur accordant plus de terres et la gestion du yassak ce qui aurait pu favoriser davantage la fixation de ces lignées paternelles dans la population yakoute.

V.2.1.7 Approche globale des lignées maternelles anciennes

La répartition des haplogroupes de l'ADN mitochondrial au sein de l'ensemble des sujets anciens analysés illustre la dualité des origines des lignées maternelles des Yakoutes. La fréquence des haplogroupes D et D5 serait le reflet d'un effet fondateur survenu au cours des migrations depuis les régions du sud probablement de la Bouriatie d'après les données de Pakendorf *et al.* (2006). La présence de nombreuses séquences retrouvées dans les populations d'Asie Centrale et de Mongolie, notamment de la vallée d'Eygin Gol, confirme une origine méridionale d'une partie des lignées maternelles. Par ailleurs, la fréquence élevée d'haplogroupes C et le nombre de lignées partagées avec des populations sibériennes, notamment toungousses, évoquent des échanges génétiques entre les nouveaux arrivants et les populations autochtones.

Ces caractéristiques semblent se mettre en place dès le XV^{ième} siècle puisque l'ensemble des lignées maternelles se révèle très homogène au cours du temps. En effet, aucune différence

significative n'est détectée entre les sous-groupes chronologiques. De plus, les distances génétiques observées pour les lignées maternelles de notre population du XV^{ième} au XVIII^{ième} siècle indiquent des affinités avec de nombreuses populations à la fois de Sibérie (Evenks et Tuvas) mais également d'Asie Centrale (Kazakh, Kirghiz et Ouigours) et de Mongolie/Bouriatie. L'influence des populations de Mongolie et d'Asie Centrale est donc bien présente. Par ailleurs, il est intéressant de constater la proximité avec les populations Evenks indiquant l'assimilation des populations autochtones par les Yakoutes dès les premières étapes de la colonisation. Les indices de diversités calculées pour les sujets anciens montrent d'ailleurs des affinités marquées avec les populations sibériennes.

Cette étude a également permis de mettre en évidence des contacts plus anciens entre les populations nomades des steppes et les chasseurs autochtones de Sibérie. Les lignées maternelles présentes chez les sujets les plus anciens, avant l'individualisation des Yakoutes, montrent des affinités avec de nombreuses populations à la fois de Sibérie (Altaï, Evenks, Koryaks et Tuvas) mais également d'Asie Centrale (Kazakh, Kirghiz et Ouigours) et de Mongolie/Bouriatie. Ces lignées paternelles pourraient représenter un pool génétique commun dont la répartition daterait du Paléolithique Supérieur comme le proposent certains auteurs (Schurr *et al.*, 2004). Cependant, les artefacts retrouvés dans la tombe du sujet de Pokrovsk ainsi que le partage de cette lignée maternelle avec un sujet de la nécropole d'Eygin Gol, évoquent clairement des contacts entre les populations des steppes et les groupes autochtones de Sibérie au moins dès le III^{ième} siècle avant notre ère. De plus, la présence de nombreux haplotypes partagés entre les sujets anciens de ces deux populations, dont un appartenant à un haplogroupe peu fréquent, renforce l'hypothèse de contacts précoce entre ces deux régions probablement durant la période Xiongnu.

V.2.1.8 Des différences régionales au sein de la Yakoutie Centrale

Les analyses multidimensionnelles regroupant les sujets suivant leur origine géographique a fait apparaître des différences entre certains des *ulus* tant pour les lignées paternelles que maternelles. Ce résultat contraste avec les données obtenues lors de l'AMOVA réalisée à partir des données des STR autosomaux qui montrait une absence de différence que l'on considère des critères chronologiques ou géographiques.

Il est important de noter que, pour les STR du chromosome Y et la région HV1 de l'ADN mitochondrial, les différents groupes géographiques ne présentent pas de différence avec la population yakoute actuelle, indiquant que la contribution des sujets des trois *ulus* au pool génétique des Yakoutes modernes a du être similaire.

Pour le chromosome Y, les populations des sous groupes de Churapchinsky et de Khangalassky ne présentent pas de différence significatives entre eux. En revanche l'*ulus* de Tattinsky se distingue par ses distances F_{st} , certes plus réduites qu'avec l'ensemble des autres populations, mais tout de même considérées comme significatives.

Cette divergence peut s'expliquer par le fait que les cinq sujets de cet *ulus* appartiennent tous à l'haplotype H1. En revanche, quatre des haplotypes de la région de Khangalassky sont affiliés à H2, et les deux autres n'appartiennent pas à l'haplogroupe N3. Enfin, quatre des sujets de l'*ulus* de Churapchinsky appartiennent à la lignée H2, un individu porte la mutation TAT-C mais n'appartient pas à H2 et le dernier appartient à l'haplogroupe P. La présence d'un seul haplotype, spécifique à la population yakoute, dans la région de Tattinsky est responsable des distances plus importantes par rapport aux deux autres *ulus* qui présentent tous deux des haplotypes n'appartenant pas uniquement à N3.

On peut voir dans cette distinction entre les *ulus* une différenciation entre clans yakoutes fondés par des sujets appartenant à des lignées masculines différentes. En effet, la famille élargie, reliée par des ancêtres masculins communs, occupait généralement des *alas* voisins favorisant ainsi la prédominance de certaines lignées paternelles à l'échelle micro-locale. Ce découpage clanique du territoire en Yakoutie Centrale pourrait expliquer un isolement des sujets de l'*ulus* de Tatta. Toutefois ce résultat contraste avec les rapports ayant établis que le découpage administratif de la Yakoutie par les Russes ne suivait pas l'ancienne répartition des terres suivant les différentes structures claniques ayant cours pendant la période médiévale (LeBerre-Semenov, 2002). Il faut également tenir compte du fait que la diversité des lignées paternelles de l'*ulus* de Tattinsky pourrait être plus importante mais en raison du nombre réduit de sujets, elle se révèle impossible à détecter dans notre échantillon.

La distinction géographique entre les différentes lignées maternelles semble plus difficile à expliquer. En effet, la variété des haplotypes mitochondriaux présents chez les Yakoutes suggère que les migrants provenant des steppes du sud épousaient des femmes appartenant aux populations autochtones. Les origines des lignées maternelles sont ainsi beaucoup plus diverses dans la population yakoute, regroupant des sujets d'ascendance turco-mongole mais

également toungousse ou présentant des affinités avec des populations de Sibérie de l'ouest et du sud comme les Tuvas. Le fait que les lignées maternelles de l'*ulus* de Churapcha diffèrent légèrement de celles des deux autres régions est par conséquent surprenant.

Une explication potentielle pourrait venir de la distribution des haplogroupes dans ces trois régions. L'*ulus* Churapchinsky présente la fréquence plus importante de l'haplogroupe D et en particulier D5 dans cette région, plus de 50% des lignées maternelles. Par ailleurs, la comparaison avec les autres populations révèle des différences significatives avec l'ensemble des populations d'Asie Centrale et de Sibérie. En revanche, dans l'*ulus* de Tatta plus de 60% des séquences appartiennent à l'haplogroupe C et l'*ulus* de Khangalassky comprend un mélange de ces deux haplogroupes. Ces deux groupes montrent des distances non significatives avec les Bouriates, les Evenks et les Tuvas indiquant une mixité plus importante des lignées maternelles.

L'*ulus* de Churapcha présente une concentration exceptionnelle d'*alas*. Ces vastes dépressions dépourvues d'arbres ont certainement été mises à profit par les premiers arrivants pour adapter l'élevage en Yakoutie Centrale. Par ailleurs, cet *ulus* est proche de la Lena que l'on peut considérer comme la voie d'entrée préférentielle des migrants venus du sud. Il est par conséquent possible qu'un groupe de migrants, accompagné d'un petit nombre de femmes, se soit fixé dans cette région. La fréquence importante du sous haplogroupe D5 dans la population Yakoute ainsi que la proportion réduite d'haplotypes partagés dans l'haplogroupe D évoque en effet une expansion à partir d'un nombre de sujets féminins réduit probablement originaire de Bouriatie si on considère les résultats de Pakendorf *et al.* (2006). La proportion plus importante de ces deux haplogroupes dans la région de Churapcha par rapport aux deux autres *ulus* pourrait ainsi refléter la fixation de ces lignées maternelles venant du sud ainsi qu'une assimilation plus réduite de femmes provenant des populations locales. On peut ainsi émettre l'hypothèse que les clans de l'*ulus* de Churapcha auraient moins eu recours à l'exogamie que les clans des autres régions en raison de la présence d'un nombre plus important de femmes lors de leur migration. En revanche, pour les deux autres *ulus*, et notamment Tattinsky, des mariages plus nombreux avec la population autochtones auraient eu lieu.

V.2.1.9 Conclusions générales

Cette étude a permis de mettre en évidence l'existence de relations anciennes entre les populations du nord de la Mongolie avec les peuples autochtones qui occupaient la Yakoutie Centrale dès le III^{ème} siècle avant notre ère (cf figure 53). L'existence de contacts aussi anciens n'avaient jusqu'alors jamais été proposés. Les relations entre ces deux régions auraient dépassé les simples échanges commerciaux et il est possible que, dès cette période, de petits groupes de nomades se soient déplacés vers le nord, sans toutefois exporter leur mode de vie. Cependant, il conviendra de confirmer ces résultats par une étude plus approfondie de sujets datés de cette période provenant à la fois de Yakoutie Centrale mais également de la région du Baïkal afin de valider si une expansion de tribus nomades durant la période Xiongnu aurait pu conduire à la diffusion de ces lignées maternelles à la fois vers l'ouest et le nord.

L'hypothèse archéologique évoquant la migration des Kurykans depuis le Baïkal semble confirmée par les analyses réalisées sur les sujets anciens de la période médiévale. En effet, de fortes affinités avec les régions des steppes du sud sont retrouvées, notamment avec les Bouriates et d'autres populations vivant dans la région du lac Baïkal (cf figure 53).

Considérant les données du chromosome Y, il semble qu'un nombre réduit de sujets aient entrepris cette migration vers la Yakoutie Centrale. La diversité génétique très réduite des lignées paternelles, ainsi que leur stabilité depuis le XV^{ème} siècle et l'arrivée des premiers migrants serait à mettre en relation avec la pérennité d'une classe sociale dominante. Les individus appartenant à cette élite, guerrière dans les premiers temps puis devenue marchande, auraient engendré une descendance plus importante que l'on retrouve jusqu'à nos jours dans la population yakoute.

Les lignées maternelles attestent également de cet effet fondateur par les fréquences et la spécificité des haplotypes de l'haplogroupe D et en particulier du sous haplogroupe D5. Un nombre réduit de femmes originaires de la région péri-Baïkale, aurait pris part aux premières migrations et leurs séquences se seraient transmises durant cinq siècles jusqu'à nos jours.

Cependant, une plus grande variabilité est observée dans les lignées maternelles. Cette différence entre des haplotypes du chromosome Y, exclusifs à la population yakoutes, et les séquences mitochondrielles dont certaines sont largement retrouvées dans les populations

sibériennes pourrait s'expliquer, d'une part, par le régime patrilocal de la société yakoute. D'autre part, du fait de mariages avec des femmes issues des populations toungousses qui se retrouvent dans la correspondance des séquences anciennes avec des haplotypes retrouvés aujourd'hui chez les Evenks (cf figure 53).

A l'exception de ces mariages, la population yakoute apparaît extrêmement stable au cours du temps. A la fois pour les lignées paternelles et maternelles, la plupart des haplotypes sont en effet retrouvés au cours des siècles depuis les prémisses de la culture yakoute jusqu'à nos jours et cela en dépit des influences culturelles différentes durant la période médiévale et des déprises démographiques liées aux famines et autres épidémies survenues avec l'arrivée des colons russes au XVIII^{ième} siècle.

Cette homogénéité dans la structure génétique durant les époques précédant la colonisation russe ainsi que chez les sujets appartenant à la période suivante montre que l'influence génétique des colons fut limitée. En effet, nous n'avons pas mis en évidence la présence de lignées paternelles d'ascendance russe et les haplotypes présents sont majoritairement spécifiques des Yakoutes. De plus, les lignées maternelles d'origine ouest-eurasiatique précèdent les débuts de la colonisation et leurs fréquences sont similaires à celles rencontrées dans nombre de populations autochtones impliquant une origine ancienne de ces lignées.

La qualité des données qu'il a été possible d'obtenir ainsi que les hypothèses originales qui ont pu être proposées concernant la formation et l'évolution de la population de Yakoutie Centrale confirment la pertinence, dans l'étude des populations du passé, d'une approche pluridisciplinaire incluant une analyse moléculaire. La poursuite des recherches dans cette même optique sur des sujets provenant de différentes régions permettra, sans doute, d'enrichir les connaissances sur l'ethnogenèse yakoute et de préciser les modalités du peuplement de la Sibérie Orientale.

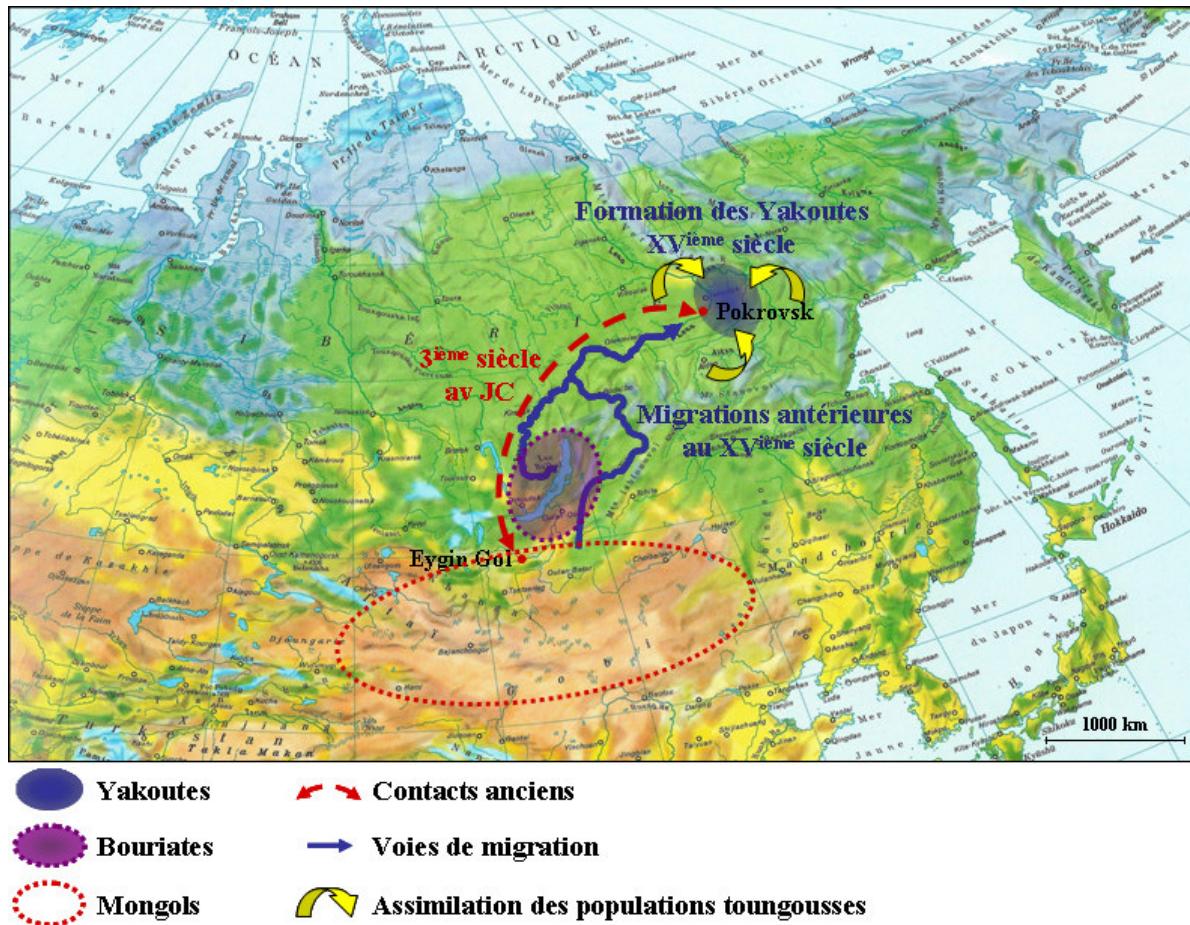


Figure 53 : Contacts entre populations et schémas de migration proposés.

VI REFERENCES BIBLIOGRAPHIQUES

A

- Abbot A. 2001. Earliest malaria DNA found in Roman baby graveyard. *Nature* 412 : 847.
- Ackerman A.B, de Viragh P.A, Chongchitnant N. 1993. Anatomic, histologic and biological aspects, dans: Ackerman A.B, de Viragh P.A, Chongchitnant N. (Eds.), *Neoplasms with follicular differentiation (Ackerman's Histologic Diagnosis of Neoplastic Skin Diseases: A Method by Pattern Analysis)*, Lea and Febiger, Philadelphia, pp. 29–90.
- Adachi N, Umetsu K, Takigawa W, Sakaue K. 2004. Phylogenetic analysis of the human ancient mitochondrial DNA. *J Archaeol Sci* 31:1339–1348.
- Akane, A., K. Matsubara, H. Nakamura, S. Takahashi, and K. Kimura. 1994. Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from blood strains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J. Forensic Sci.* 39:362–372.
- Ammerman, A. J. et Cavalli-Sforza, L. L. 1984 The Neolithic Transition and the Genetics of Populations in Europe, Princeton Univ. Press, Princeton.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–65.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Arctander P. 1999. Mitochondrial recombination? *Science* 284:2090–91.
- Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J, Guhl F, Arriaza B, Renier C, Wittmers LE Jr, Fornaciari G, Allison M. 2004. A 9,000-year record of Chagas' disease. *Proc Natl Acad Sci U S A.* ; 101 (7): 2034-9.
- Awadalla P, Eyre-Walker A, Maynard Smith J. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286: 2524–25.
- Awadalla P, Eyre-Walker A, Smith JM. 2000. Linkage disequilibrium and recombination in hominid mitochondrial DNA: Comments. *Science* 288:1931a

B

- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics*. 141(2): 743-53.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16(1): 37-48.
- Bandelt HJ, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115 :64–69.

- Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet.* 71(5): 1150-60.
- Bandelt HJ. Mosaics of ancient mitochondrial DNA: positive indicators of nonauthenticity. 2005. *Eur J Hum Genet.* 13 (10): 1106-12.
- Barnes I, Matheus P, Shapiro B, Jensen D et Cooper A. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295, 2267–2270.
- Beauval C, Maureille B, Lacrampe-Cuyaubère 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, Pääbo S, Trinkaus E. 2005. A late Neandertal femur from Iles Rochers-de-Villeneuve, France. *PNAS* 102 (20) 7085-7090.
- Bednarik RG et Devlet K. 1992. Rock art conservation in Siberia. *Purakala* 3(1-2): 3-11.
- Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, Casoli A, Goyache F, Royo LJ, Conti S, Lari M, Martini A, Ouragh L, Magid A, Atash A, Zsolnai A, Boscato P, Triantaphylidis C, Ploumi K, Sineo L, Mallegni F, Taberlet P, Erhardt G, Sampietro L, Bertranpetti J, Barbujani G, Luikart G, Bertorelle G. 2006. The origin of European cattle: Evidence from modern and ancient DNA. *Proc Natl Acad Sci U S A*, 103(21):8113-8.
- Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* 57:248–256.
- Beutler E, Gelbart T, Kuhl W. 1990. Interference of heparin with the polymerase chain reaction. *Biotechniques.* 9 (2): 166.
- Boeva SB. 1988. Anthropogenetic study of the indigenous population of the Yakut ASSR in connection with the ethnogenesis of the Yakuts. Ph.D. thesis, Moscow State University, Moscow.
- Bonnichsen R, Hodges L, Ream W, Field K, Kirner DL, Selsor K et Taylor RE. 2001. Methods for the study of ancient hair: Radiocarbon dates and gene sequences from individual hairs. *J. Archeol. Sci.* 28, 775–785.
- Bouwman AS, Chilvers ER, Brown KA, Brown TA. 2006. Brief communication: identification of the authentic ancient DNA sequence in a human bone contaminated with modern DNA. *Am J Phys Anthropol.* 131(3): 428-31.
- Bower MA, Spencer M, Matsumura S, Nisbet RE, Howe CJ. 2005. How many clones need to be sequenced from a single forensic or ancient DNA sample in order to determine a reliable consensus sequence? *Nucleic Acids Res.* 33(8): 2549-56.
- Brandstätter A, Parson W. 2003. Mitochondrial DNA heteroplasmy or artefacts a matter of the amplification strategy? *Int J Legal Med* 117 : 180–184.
- Brandstätter A, Niederstätter H, Parson W. 2004. Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. *Int J Legal Med* 118:47–54.

- Brandstatter A, Sanger T, Lutz-Bonengel S, Parson W, Beraud-Colomb E, Wen B, Kong QP, Bravi CM, Bandelt HJ. 2005. Phantom mutation hotspots in human mitochondrial DNA. Electrophoresis. 26(18): 3414-29.
- Brinkmann B, Klantschar M, Neuhuber F, Huhne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet. 62(6): 1408-15.
- 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, Morling N. 2005. Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis. 26(23): 4411-20.
- Brosch R, Gordon V, Marmiesse M, et al. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci USA; 99: 3684-89.
- Brouchkov A, Fukuda M, Fedorov A, Konstantinov P, Iwahana G. 2004. Thermokarst as a short-term permafrost disturbance, Central Yakutia. Permafrost and Periglacial Processes 15, (1): 81-87.
- Brown DT, Samuels DC, Michael EM, Turnbull DM, Chinnery PF. 2001. Random genetic drift determines the level of mutant mtDNA in human primary oocytes. Am. J. Hum. Genet. 68:533-36.
- Bruzek J. 2002. A method for visual determination of sex, using human hip bone, *American Journal of Physical Anthropology* 117: 157-168.
- Buckleton J, Triggs CM, Walsh SJ. 2005. Forensic DNA interpretation. CRC Press, Boca Raton. pp:128-129.
- Budowle B, Maysibay A, Anderson SJ, Barna C et al. 2001. STR primer concordance study. Forensic Sc Int. 124(1): 47-54.
- Budowle B, Sprecher CJ. 2001. Concordance study on population database samples using the PowerPlex 16 kit and AmplFISTR Profiler Plus kit and AmpFISTR CoFiler kit. J Forensic Sc. 46(3): 637-41.
- Budowle B, Allard MW, Wilson MR, Chakraborty R. 2003. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet. 4: 119-41.

C

- Cadenas AM, Requeiro M, Gayden T, Singh N, Zhivotovsky LA, Underhill PA, Herrera RJ. 2006. Male amelogenin dropouts: phylogenetic context, origins and implications. Forensic Sc Int. Epub.
- Cano RJ, Poinar HN et Poinar Jr GO. 1992a. Isolation and partial characterisation of DNA from the bee *Proplebeia dominicana* (Apidae: Hymenoptera) in 25-40 million year old amber. Med. Sci. Res. 20, 249-251.

- Cano RJ, Poinar HN, Roublik DW et Poinar Jr GO. 1992b Enzymatic amplification and nucleotide sequencing of portions of the 18S rRNA gene of the bee *Proplebeia dominicana* (Apidae: Hymenoptera) isolated from 25–40 million year old Dominican amber. *Med. Sci. Res.* 20, 619–622.
- Cano RJ, Poinar HN, Pieniezak NS et Poinar Jr GO. 1993 Enzymatic amplification and nucleotide sequencing of DNA from 120–135 million year old weevil. *Nature* 363, 536–538.
- Cakır AH, Celebioglu, Yardımcı E. 2004. Y-STR haplotypes in Central Anatolia region of Turkey *Forensic Sci Int.* 144(1):59-64.
- Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetti J, Barbujani G, Bertorelle G. 2003. Evidence for a genetic discontinuity between Neandertals and 24000-year-old anatomically modern Europeans. *PNAS.* 100: 6593-6597.
- Castella V, Dimo-Simonin N, Brandt-Casadevall C, Mangin P. 2006. Forensic evaluation of the QIAshredder/QIAamp DNA extraction procedure. *Forensic Sci Int.* 156(1): 70-73.
- Chomette G, Auriol M. 1992. Histologie du complexe dentino-pulpaire. Paris: Editions Techniques -EMC, Stomatologie et odontologie, 13p.
- Cipollaro M, Di Bernardo G, Galano G, Galderisi U, Guarino F, Angelini F, Cascino A. 1998. Ancient DNA in human bone remains from Pompeii archaeological site. *Biochem Biophys Res Commun.* 247 (3):901-4.
- Collins MJ, Waite ER, van Duin AC. 1999. Predicting protein decomposition: the case of aspartic-acid racemization kinetics. *Philos Trans R Soc Lond B Biol Sci.* 354(1379): 51-64.
- Collins MJ, Nielsen-Marsh C, Hiller J, Smith CI, Roberts JP, Prigodich RV, Wess TJ, Csapó J, Millard AR, Turner-Walker, G. 2002, The survival of organic matter in bone: a review, *Archaeometry*, 44, 383–94.
- Colson IB, Richards MB, Bailey JF, Sykes BC. 1997. DNA analysis of seven human skeletons excavated from the terp of wijnaldum. *J Archaeol Sci* 24:911-917.
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, et al. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.* 63: 1824–38.
- Comas D, Calafell F, Bendukidze N, Fananas L, Bertranpetti J. 2000. Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol.* 112(1): 5-16.
- Comas, D., S. Plaza, R. S. Wells, N. Yuldaseva, O. Lao, F. Calafell, and J. Bertranpetti. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* 12:495–504.

Cooper A, Mourer-Chauviré C, Chamber GK, Haeseler A, Wilson AC, Pääbo S. 1992. Independent origins of new Zealand moas and kiwis. *Proc Natl Acad Sci USA* 89 :8741-8744.

Crubézy E, Ludes B, Poveda JD, Clayton J, Crouau-Roy B, Montagnon D. 1998. Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III* 321 (11): 941-51.

Cooper A et Poinar H. 2000. Ancient DNA. Do it right or not at all, *Science* 289 1139.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*. 409 (6821): 704-707.

Crubézy E, Braga J, Larrouy G. 2002. *Anthropobiologie*. Masson, Paris.

Crubézy E, Legal L, Fabas G, Dabernat H, Ludes B. 2006. Pathogeny of archaic mycobacteria at the emergence of urban life in Egypt (3400 BC). *Infect Genet Evol*; 6 (1): 13-21.

D

De Knijff P. 2000. Y chromosomes shared by descent or by state. In: Renfrew C, Boyle K (eds) *Archaeogenetics: DNA and the population prehistory of Europe*. The McDonald Institute, Cambridge, pp 301–304.

Derenko MV, Shields GF. 1997. Mitochondrial DNA sequence diversity in three north Asian aboriginal populations groups. *Molecular Biology* 31(5):665-669.

Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM, Nimaev DD, Zakharov AIA. 2000. Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of north Asia. *Human biology* 72(6):945-973.

Derenko MV, Maliarchuk BA, Denisova GA, Dambueva IK, Kakpakov VT, Dorzhu ChM, Luzina FA, Lotosh EA, Ondar UN, Kaplina MI, Zakharov IA. 2002. Molecular genetic differentiation of ethnic populations in Southern and Eastern Siberia based on mitochondrial DNA polymorphism. *Russian Journal of Genetics*. 38 (10): 1196–1202.

Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, Zakharov IA. 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet*. 67(Pt 5): 391-411.

Derenko MV, Lunkina AV, Maliarchuk BA, Zakharov IA, Tsedev T, Park KS, Cho YM, Lee NK, Chu CH. 2004. Restriction polymorphism of mitochondrial DNA in Koreans and Mongolians. *Russian Journal of Genetics*. 40(11): 1292–1299.

Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I, Dorzhu C, Luzina F, Miscicka-Sliwka D, Zakharov I. 2006. Contrasting patterns of Y-chromosome variation in

- South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet.* 118(5):591-604.
- Derbeneva OA, Starikovskaya EB, Wallace DC, Subernik RI. 2002. Traces of early eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet.* 70:1009-1014.
- DeSalle, R., Gatesy, J., Wheeler, W. & Grimaldi, D. 1992 DNA Sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257, 1933–1936.
- DeSalle, R., Barcia, M. & Wray, C. 1993 PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*). *Experientia* 49, 906–909.
- DeSalle, R. 1994 Implications of ancient DNA for phylogenetic studies. *Experientia* 50, 543–550.
- Di Benedetto G, Ayse E, Stenico M, Castri L, Bertorelle G, Togan I, Barbujani G. 2001. DNA diversity and population admixture in Anatolia. *Am J Phys Anthropol* 115:144-156.
- Donoghue HD, Spigelman M, Zias J, Gernaey-Child AM, Minnikin DE. 1998. Mycobacterium tuberculosis complex DNA in calcified pleura from remains 1400 years old. *Lett Appl Microbiol* ; 27 (5): 265-9.
- Donoghue HD, Marcsik A, Matheson C, Vernon K, Nuorala E, Molto JE, Greenblatt CL, Spigelman M. 2005. Co-infection of Mycobacterium tuberculosis and Mycobacterium leprae in human archaeological samples: a possible explanation for the historical decline of leprosy. *Proc Biol Sci.* 22 ; 272 (1561): 389-94.
- Drancourt M, Aboudharam G, Signoli M, Dutour O, 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 (21): 12637-40.
- Drancourt M, Raoult D. Molecular insights into the history of plague. 2002. *Microbes Infect*; 4(1): 105-9.
- Drancourt M, Roux V, Dang LV, Tran-Hung L, Castex D, Chenal-Francisque V, Ogata H, Fournier PE, Crubézy E, Raoult D. 2004. Genotyping, Orientalis-like *Yersinia pestis*, and plague pandemics. *Emerg Infect Dis* ; 10 (9): 1585-92.
- Dupuy BM, Stenersen M, Egeland T, Olaisen B. 2004. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. *Hum Mutat.* 23(2):117-24.
- Ebling F.J. 1980. The physiology of hair growth, *Cosmet. Sci.* 2 181–232.

E

- Eglinton G, Logan GA. 1991. Molecular preservation. *Philos Trans R Soc Lond B Biol Sci.* 333 (1268): 315-27.
- Enriquez JA, Cabezas-Herrera J, Bayona-Bafaluy MP, Attardi G. 2000. Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. *J. Biol. Chem.* 275:11207–15
- Eyre-Walker A, Smith NH, Smith JM. 1999. How clonal are human mitochondria? *Proc. R. Soc. London B Biol. Sci.* 266:477–83.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 131(2): 479-91.
- Excoffier L, Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol* 16:1357–1368
- F**
- Faerman M, Filon D, Kahila G, Greenblatt CL, Smith P, Oppenheim A. 1995. Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. *Gene.* 167 (1-2):327-32.
- Faerman M et Bar-Gal G K. 1998. Determining the Sex of Infanticide Victims from the Late Roman Era through Ancient DNA Analysis. *Journal of Archaeological Science* 25, 861–865.
- Ferreira LF, Britto C, Cardoso MA, Fernandes O, Reinhard K, Araujo A. 2000. Paleo parasitology of Chagas disease revealed by infected tissues from Chilean mummies. *Acta Trop.* 25; 75 (1): 79-84.
- Ferembach, D., Schwiedetsky, I., Stloukal, M., 1979, Recommandations pour déterminer l'âge et le sexe sur le squelette. *Bull Mém Soc Anthropol Paris, XIIIème série*, 7-45.
- Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *Am J Hum Genet.*; 67(1) : 182-96
- Forster P. 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc Lond B Biol Sci.* 29; 359 (1442): 255-64.
- Fotiev S M, 1997. Permafrost ground water Russian Literature Review. In: Haldorsen S, Liebman M, Nelson G, van Everdingen R O, Boike J, 1997: State-of-the-art Report on saturated water movement in permafrost areas. Report No 4/97 (Inr 54). Agricultural University of Norway. ISSN 0805-7214.
- Forsyth J. 1992. History of the people of Siberia. Cambridge University Press.

Fraser R.B., MacRae T.P., Rogers G.E. 1972. Keratins: Their Composition, Structure and Biosynthesis, Charles E. Thomas, Springfield.

G

Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. 2001. Improved mtDNA sequence analysis of forensic remains using "mini-primer set" amplification strategy. *J Forensic Sci* 46:247-253.

Gaytmenn R, Sweet D. 2003. Quantification of forensic DNA from various regions of human teeth. *J Forensic Sci*. 48 (3): 622-5.

Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A. 2003a. Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *Am. J. Hum. Genet.* 72: 32-47.

Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A. 2003b. Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. *Am. J. Hum. Genet.* 72: 48-61.

Gilbert MT, Barnes I, Collins J, Smith C, Eklund J, Goudsmit J, Poinar H, Cooper A. 2005a. Long-Term Survival of Ancient DNA in Egypt: Response to Zink and Nerlich (2003). *Am J. Phys Anthropol.* 121(2): 109-11.

Gilbert MT, Bandelt HJ, Hofreiter M, Barnes I. 2005b. Assessing ancient DNA studies. *Trends Ecol Evol*, 20 (10) : 541-4.

Gilbert MT, Janaway RC, Tobin DJ, Cooper A, Wilson AS. 2006a. Histological correlates of post mortem mitochondrial DNA damage in degraded hair. *Forensic Sci Int.* 156: 201-207.

Gilbert MT, Menez L, Janaway RC, Tobin DJ, Cooper A, Wilson AS. 2006b. Resistance of degraded hair shafts to contaminant DNA. *Forensic Sci Int.* 156: 208-212.

Gilbert MT, Hansen AJ, Willerslev E, Turner-Walker G, Collins M. 2006c. Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. *Int. J. Osteoarchaeol.* 16: 156-164.

Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 77:6715-19.

Gilichinsky, D. 2002. Permafrost as a microbial habitat. In Encyclopedia of Environmental Microbiology (eds), pp. 932-956, John Wiley& Sons.

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet.* 6 (2):130-5.

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. 2000. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int* 112: 17-40.

Gogolev AI. 1993. The Yakuts. Problems of their ethnogenesis and formation of their culture (en russe). Yakutsk: Izdatel'stvo JaGU. In:

Gogolev AI. 2000. The history of Yakutia (en russe). Yakutsk: Izdatel'stvo Jakutskogo Gosudarstvennogo Universiteta. In: Le Berre-Semenov. 2002. Renaissance et renaissance des peuples du nord: évolution de la question autochtone en République Sakha dans le contexte des mutations post soviétiques. Thèse de Doctorat. Institut National des Langues et Civilisations orientales.

Golenberg, E. M., Giannassi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D. & Zurawski, G. 1990 Chloroplast DNA from a Miocene Magnolia species. *Nature* 344, 656–658.

Götherstörn A, Collins M, Angerbjörn A, Liden K. 2002. Bone preservation and DNA preservation. *Archaeometry*. 3: 395-404.

Gordon, CC. et Buikstra, JE. 1981. Soil pH, bone preservation, and sampling bias at mortuary sites, *American Antiquity*, 46, 566–71.

Gray H. 1918. Anatomy of the Human Body. Philadelphia: Lea & Febiger.

Greenberg BD, Newbold JE, Sugino A. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21:33–49.

Green R, Krause J, Ptak S, Briggs A, Ronan M, Simons J, Du L, Egholm M, Rothberg J, Paunovic M, Pääbo S. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 444(16): 330-336.

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

Grimoud AM, Boulbet Mauger M, Lodter JP. 2004. Critères de sélection d'échantillons dentaires pour l'étude de l'ADN ancien. *Antropo*, 6, 43-51.

Grzybowski T 2000. Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis* 21: 548–553.

Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, Kotzbach R. 2003. High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision. *Electrophoresis*. 24(7-8):1159-65.

H

Haas CJ, Zink A, Palfi G, Szeimies U, Nerlich AG. 2000. Detection of leprosy in ancient human skeletal remains by molecular identification of *Mycobacterium leprae*. *Am J Clin Pathol* ; 114 (3): 428-36.

Hacker G. 2000. The morphology of apoptosis, *Cell. Tissue Res.* 301 5–17.

- Handt O, Krings M, Ward RH, Paabo S. 1996. The retrieval of ancient human DNA sequences. *Am J Hum Genet.* 59(2): 368-76.
- Hänni C, Begue A, Laudet V, Stehelin D. 1995. Isopropanol precipitation removes PCR from ancient bones extracts. *Nucleic Acids Res* 23 :881-882.
- Hauswirth WW, Dickel CD, Rowold DJ, Hauswirth MA. 1994. Inter- and intrapopulation studies of ancient humans. *Experientia* 50:585-591.
- Hagelberg E, Sykes B, Hedges R. 1989. Ancient bone DNA amplified. *Nature*. 342 (6249): 485.
- Hagelberg E, Bell LS, Allen T, Boyde A, Jones SJ, Clegg JB. 1991. Analysis of ancient bone DNA: techniques and applications. *Phil Trans R Soc Lond B* 333 :399-407.
- Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhövel W, et al. 1999. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc. R. Soc. London B Biol. Sci.* 266:485-92
- Hall A, Ballantyne J. 2004. Characterization of UVC-induced DNA damage in bloodstains: forensic implications. *Anal Bioanal Chem.* 380 (1): 72-83.
- Handt O, Krings M, Ward RH, Pääbo S. 1996. The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.* 59: 376–386.
- Harada K, Wada K, Sueyoshi T, Fukuda M. 2006. Resistivity structures in alas areas in Central Yakutia, Siberia, and the interpretation of permafrost history. *Permafrost and Periglacial Processes*: 17, (2): 105-118.
- Harding H., Rogers G. 1999. Physiology and growth of human hair, in: J. Robertson (Ed.), *Forensic Examination of Human Hair*, Taylor and Francis, London, pp. 1-77.
- Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37:347-354.
- Hawkins CL, Davies MJ. 2002. Hypochlorite-induced damage to DNA, RNA, and polynucleotides: formation of chloramines and nitrogen-centered radicals. *Chem Res Toxicol.* 15 (1): 83-92.
- Hawkins CL, Pattison DI, Davies MJ. 2003. Hypochlorite-induced oxidation of amino acids, peptides and proteins. *Amino Acids*. 25 (3-4): 259-74.
- Hedges R. 2002. Bone diagenesis: an overview of processes. *Archaeometry*. 3: 319-328.
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am. J. Hum. Genet.* 69:1113–26.
- Higuchi R, Bowman B, Freiburger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*. 312 (5991): 282-4.

- Hofreiter M, Serre D, Poinar H.N, Kuch M, Pääbo S. 2001. Ancient DNA. *Nature Rev. Genet.* 2: 353-360.
- Hoss M, Pääbo S. 1993. DNA extraction from pleistocene bones by a silica based purification method. *Nucleic Acids Res* 21(16): 3913-3914.
- Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 24 (7): 1304-7.
- Howell N et Smejkal CB. 2000. Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. *Am. J. Hum. Genet.* 66:1589–98.
- Hummel S, Herrmann B. 1991. Y-chromosome-specific DNA amplified in ancient human bone. *Naturwissenschaften.* 78(6): 266-7.
- Hummel S. 2003. Ancient DNA typing, Methods, Strategies and Applications. Springer, 298p.
- Hussein MR. 2005. Ultraviolet radiation and skin cancer: molecular mechanisms. *J. Cutan. Pathol.* 32(3): 191-205.

I

- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–13.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409: 860-921.
- Ishida H, and Dodo Y. 1996. Cranial morphology of the Siberians and East Asians. In T Akazawa and EJE Szathmary (eds.): *Prehistoric Dispersal of Mongoloids*. Oxford: Oxford University Press, pp. 113–124.
- Ivanov P, Wadhams M, Roby R, Holland M, Weedn V, Parson T. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12 :417-420.

J

- Janczewski DN, Yunki N, Gilbert DA, Jefferson GT, O'Brien SJ. 1992. Molecular phylogenetic inference from saber-toothed cat fossils of rancho la Brea. *Proc Natl Acad Sci USA* 89: 9769-9773.
- Jans E, Kars H, M Nielsen-Marsh C, Smith C, Nord A, Arthur P, Earl N. 2002. In situ preservation of archaeological bones: a histological study within a multidisciplinary approach. *Archaeometry* 44, 3: 343–352.

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat. Genet.* 18:109–10.

Jobling MA, Tyler-Smith C. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* (11): 449-56.

Jobling MA, Tyler Smith C. 2003. The Human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet.* 4(8): 598-612.

Jorde LB, Bamshad M. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a.

K

Kalmar T, Bachrati CZ, Marcsik A, Rasko I. 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res.* 28(12): e67.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831.

:Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet.* 69(3): 615-28.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A. 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet.* 66(5): 1580-8.

Kemp BM, Smith DG. 2005. Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int.* 154(1): 53-61.

Keyser-Tracqui C, Crubezy E, 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(2):247-60.

Keyser-Tracqui C, Crubezy E, Pamzsav H, Varga T, Ludes B. 2006. Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. *Am J Phys Anthropol.* 131(2):272-81.

Khitinskaya IY, Stepanov VA, Puzyrev VP. 2001. Analysis of the Alu Polymorphism in the Buryat Populations. *Russian Journal of Genetics* 37, 11: 1306-1311.

Khitinskaya IY, Stepanov VA, Puzyrev VP, Spiridonova MG, Puzyrev KV, Maksimova NR, Nogovitsyna AN. 2003. Genetic Peculiarity of the Yakut Population as Inferred from Autosomal Loci. *Mol. biol.* 37 (2): 205-209.

- Kitabata H, Nishizawa K, Yoshida Y, Maruyama K. 2006. Permafrost Thawing in Circum-Arctic and Highlands under Climatic Change Scenario Projected by Community Climate System Model (CCSM3). SOLA, Vol. 2, 053-056.
- Kloosterman A.D, Kersbergen P. 2003. Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. J Soc Biol.; 197 (4) 351-359.
- Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of new world founders. Genetics 142:1321-1334.
- Kolman CJ, Tuross N. 2000. Ancient DNA analysis of human populations. Am J Phys Anthropol. 111 (1): 5-23.
- Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP. 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum. Genet. 113:391.
- Konstantinov IV. 1975. The origins of the Yakut people and their culture. Yakutia and her neighbors in Antiquity. *Publications of the Prilenskaya Archaeological expedition*. Yakutsk :Yakutskiy.
- Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Paabo S, Hofreiter M. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature. 9; 439 (7077): 724-7.
- Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, et al. 2004. Recombination of human mitochondrial DNA. Science 304: 981.
- Kreader CA. 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl Environ Microbiol. 62(3): 1102-1106.
- Krings M, Stone A, Schmits RW, Krainitzki H, Stoneking M, Pääbo S. 1997. Neanderthal DNA sequences and the origin of modern humans. Cell 90 :19-30.
- Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M, Paabo S. 2000. A view of Neandertal genetic diversity. Nat Genet. 26(2):144-6.
- Kumar S, Hedrick P, Dowling T, Stoneking M. 2000. Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931.
- KuzminYV, Orlova L.A, Stuart A.J, Zenin V.N, Dementiev N. 2001. Human-mammoth (Mammuthus primigenius Blumenbach) inter-relationships in the Palaeolithic of Siberia: a review of current knowledge. The World of Elephants - International Congress, Rome.

L

- Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, Aula P, Sajantila A. 1999 Y-chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur J Hum Genet* 7:447–458.
- Lalueza-Fox C, Gilbert MTP, Martinez-Fuentes AJ, Calafell F, Bertranpetti J. 2003. Mitochondrial DNA from pre-columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. *Am J Phys Anthropol* 121 :97-108.
- Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martinez-Maza C, Bastir M, Fortea J, de la Rasilla M, Bertranpetti J, Rosas A. 2005. Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula.
- Lambert DM, Ritchie PA, Millar CD, Holland BJ, Drummond A et Baroni C. 2002. Rates of evolution in ancient DNA from Adélie penguins. *Science* 295, 2270–2273.
- Lakes R. 1993. Materials with structural hierarchy. *Nature* 361, 511-515.
- Lassen C, Hummel S, Herrmann B. 2000. Molecular sex identification of stillborn and neonate individuals ("Traufkinder") from the burial site Aegerten. *Anthropol Anz.* 58 (1):1-8.
- LeBerre-Semenov M. 2002. Renaissance et renaissance des peuples du nord: évolution de la question autochtone en République Sakha dans le contexte des mutations post soviétiques. Thèse de Doctorat. Institut National des Langues et Civilisations orientales.
- Legros F, Lombes A, Frachon P, Rojo M. 2002. Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusins. *Mol. Biol. Cell* 13:4343–54
- Lee HY, Yoo JE, Park MJ, Chung U, Kim CY, Shin KJ. 2006. East Asian mtDNA haplogroup determination in Koreans: Haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis*. Epub.
- Lees, S. 1989. Some characteristics of mineralised collagen in Calcified tissue: topics in molecular and structural biology. Ed: D. W. Hukins. 153–73, Macmillan, London.
- Levin MG, Potapov LP. 1964. Peoples of Siberia. Univ Chicago Press.
- Limpens, J., and A. Kijlstra. 1993. Human vitreous fluid contains a potent inhibitor of the polymerase chain reaction, abstr. 1736. In Abstracts of the Annual Meeting of the Association for Research in Vision and Ophthalmology. ARVO Inc., Bethesda, Md.
- Linch C.A, Smith S.L, Prahlow J.A. 1998. Evaluation of the human hair root for DNA typing subsequent to microscopic comparison. *J. Forensic Sci.* 43 (2) 305–314.
- Linch CA, Whiting DA, Holland MM. 2000. Human hair histogenesis for the mitochondrial DNA forensic scientist. *J Forensic Sci.* 46(4): 844-53.

Lin Z, Kondo T, Minamino T, Ohtsuji M, Nishigami J, Takayasu T, Sun R, Ohshima T. 1995. Sex determination by polymerase chain reaction on mummies discovered at Taklamakan desert in 1912. *Forensic Sci Int.* 75(2-3): 197-205.

Loreille O, Diegoli T.M, Irwin J.A, Coble M.D and Parsons T.J. 2006. High Efficiency DNA Extraction from Bone by Total Demineralization. Submitted in FSI.

Lutz S, Weisser HJ, Heizmann J, Pollak S. 2000. Mitochondrial heteroplasmy among maternally related individuals. *Int J Legal Med* 113:155–161.

M

Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, et al. 1997. mtDNA mutation rates—no need to panic. *Am. J. Hum. Genet.* 61:983–86

Manfredi G, Thyagarajan D, Papadopoulou LC, Pallotti F, Schon EA. 1997. The fate of human sperm-derived mtDNA in somatic cells. *Am. J. Hum. Genet.* 61:953–60

Marshall R.C. 1985. Hair comparison by protein analysis, in: Proceedings of the International Symposium on Forensic Hair Comparisons, FBI, Quantico, Viginia, pp. 71–87.

Merriwether DA, Kaestle FA. 1999. Mitochondrial recombination? (continued). *Science* 285:837

Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152: 1103–10.

Meyer E, Wiese M, Bruchhaus H, Claussen M, Klein A. 2000. Extraction and amplification of authentic DNA from ancient human remains. *Forensic Sci Int* 113:87-90.

Miller DN, Bryant JE, Madsen EL, Ghiorse WC. 1999. Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. *Applied and Environmental Microbiol.* 65 (11): 4715–4724.

Mitchell D, Willerslev E, Hansen A. 2005. Damage and repair of ancient DNA. *Mutation Research* 571: 265-276.

Mochanov Y. A. 1977. Early Stages of Human Occupation in NE Asia. Nauka, Novosibirsk, Russia.

Montiel R, Garcia C, Canadas MP, Isidro A, Guijo JM, Malgosa A. 2003. DNA sequences of *Mycobacterium leprae* recovered from ancient bones. *FEMS Microbiol Lett.* 26 ; 226(2): 413-4.

Mooder KP, Schurr TG, Bamforth FJ, Bazaliiski VI, Savel'ev NA. 2006. Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. *Am J Phys Anthropol.* 129(3):349-61.

N

- Nachman MW et Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156, 297–304.
- Nakahori Y, Takenaka O, Nakagome Y. 1991a. A human X-Y homologous region encodes "amelogenin". *Genomics*; 9(2):264-9.
- Nakahori Y, Hamano K, Iwaya M, Nakagome Y. 1991b. Sex identification by polymerase chain reaction using X-Y homologous primer. *Am. J. Med. Genet.*; 39(4):472-3.
- Nasidze I et Stoneking M. 2001. Mitochondrial DNA variation and language replacements in the Caucasus. *Proc Biol Sci.* 268(1472): 1197-206.
- Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M. 2004. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Annals of Human Genetics* 68, 205-221.
- Nicholson R A. 1998. Bone degradation in a compost heap, *Journal of Archaeological Science*, 25, 393–403.
- Nielsen-Marsh C et Hedges REM. 2000. Patterns of diagenesis in bone II: effects of acetic acid treatment and removal of diagenetic CO_3^{2-} . *Journal of Archaeological Science*, 27 : 1151–9.
- Noonan J, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard K, Rubin E. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 314 : 1113-1118.
- Nordborg M. 1998. On the probability of Neanderthal ancestry. *Am J Hum Genet* 63: 1237–1240.
- Nossintchouk RM. 1985. *Manuel d'odontologie médico-légale* (Paris : Masson), pp 213.

O

- Okazaki M, Yoshida Y, Yamaguchi S, Kaneno M, Elliot JC. 2001. Affinity binding phenomena of DNA onto apatite crystals. *Biomaterials*, 22 : 2459-2464.
- Okladnikov AP. 1955. The history of the Yakut ASSR. Volume 1 : Yakutia before its incorporation into the Russian state. Moscow : Izdatel'stvo Akademii Nauk SSSR.
- Okladnikov AP. 1959. Ancient population of Siberia and its cultures. Cambridge, MA: Peabody Museum.
- Okladnikov AP. 1964. Ancient population of Siberia and its cultures. In: Levin MG, Potapov LP, editors. *The peoples of Siberia*. Chicago: University of Chicago Press. p 13–98.

Okladnikov AP. 1972. Yakutia before its incorporation to the Russian state. Montreal, McGill University Press pp: 191-194.

Ono T, Isobe K, Nakada K, Hayashi JI. 2001. Human cells are protected from mitochondrial dysfunction by complementation of DNA products in fused mitochondria. *Nat. Genet.* 28:272–75.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* 29:20–21.

Orlando L, Leonard JA, Thenot A, Laudet V, Guerin C, Hanni C. 2003. Ancient DNA analysis reveals woolly rhino evolutionary relationships. *Mol Phylogenetic Evol.* 28 (3): 485-99.

O'Rourke DH, Carlyle SW, Parr RL. 1996. Ancient DNA: methods, progress, and perspectives. *Am. J. Hum. Biol.* 8:557–71

O'Rourke D, Hayes G, Carlyle W. 2000. Ancient DNA studies in physical anthropology. *Annu rev Anthropol.* 29: 217-242.

Ovchinnikov IV, Ovtchinnikova OI, Druzina EB, Buzhilova AP, Makarov NA. 1998. Molecular genetic sex determination of Medieval human remains from north Russia: comparison with archaeological and anthropological criteria. *Anthropol Anz.* 56 (1):7-15.

P

Pääbo S. 1985a. Molecular cloning of Ancient Egyptian mummy DNA. *Nature.* 314 (6012): 644-5.

Pääbo S. 1985b. Preservation of DNA in ancient Egyptian mummies. *Journal of Archaeological Sc.* 12: 411-417.

Pääbo S. et Wilson, A. C. 1988 Polymerase chain reaction reveals cloning artefacts. *Nature* 334, 387–388.

Pääbo S. 1989a Ancient DNA; extraction, characterization, molecular cloning and enzymatic amplification. *Proc. Natl Acad. Sci. USA* 86, 1939–1943.

Pääbo S., Higuchi, R. G. et Wilson, A. C. 1989b Ancient DNA and the polymerase chain reaction. *J. Biol. Chem.* 264, 9709–9712.

Pääbo S, Irwin DM, Wilson AC. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem.* 265 (8): 4718-21.

Pääbo S. 1996. Mutational hot spots in the mitochondrial microcosmos. *Am. J. Hum. Genet.* 59:493–96.

- Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic analysis from ancient DNA. *Annu Rev Genet* 38: 645-679.
- Panaccio M et Lew A. 1991. PCR based diagnosis in the presence of 8% (v/v) blood. *Nucleic Acids Res.* 19:1151.
- Panelli S, Damiani G, Espen L, Sgaramella V. 2005. Ligation overcomes terminal underrepresentation in multiple displacement amplification of linear DNA. *Biotechniques.* 39 (2): 174-178.
- Pakendorf B, Morar B, Tarskaia LA, Kayser M, Soodyall H, Rodewald A, Stoneking M. 2002. Y-chromosomal evidence for a strong reduction in male population size of Yakuts. *Hum Genet* 110 :198-200.
- Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. *Am J Phys Anthropol.* 120(3):211-24.
- Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet.* 6:165-83.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M. 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet.* 120(3):334-53. Epub 2006 Jul 15.
- Parr RL, Carlyle SW, O'Rourke DH. 1996. Ancient DNA analysis of Fremont Amerindians of the Great Salt LakeWetlands. *Am. J. Phys. Anthropol.* 99:507–18.
- Parson W, Parsons TJ, Scheithauer R, Holland MM. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 111:124–132.
- Parsons TJ, Irwin JA. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a
- Paulson TG, Galipeau PC, Reid BJ. 1999. Loss of heterozygosity analysis using whole genome amplification, cell sorting, and fluorescence-based PCR. *Genome Res.* 9 (5): 482-91.
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* 65: 439–58.
- Pereira L, Richards M, Goios A, Alonso A, Albaran C, Garcia O, Behar DM, Golge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15(1): 19-24.

Pfeiffer H, Lutz-Bonengel S, Pollak S, Fimmers R, Baur MP, Brinkmann B. 2004. Mitochondrial DNA control region diversity in hairs and body fluids of monozygotic triplets. *Int J Legal Med* 118:71–74.

Pitulko VV, Nikolsky PA, Giryva E Yu, Basilyan AE, Tumskoy VE, Koulakov SA, Astakhov SN, Pavlova E Yu, Anisimov MA. 2004. The Yana RHS site: Humans in the Arctic before the Last Glacial Maximum. *Science*; 303: 52-56.

Poinar Jr, G. O. et Cano, R. J. 1993 DNA from an extinct plant. *Nature* 363, 677.

Poinar HN, Hoss M, Bada JL, Paabo S. 1996. Amino acid racemization and the preservation of ancient DNA. *Science*. 272 (5263): 864-6.

Poinar H.N, Hofreiter M, Spaulding W.G, Martin P.S, Stankiewicz B.A, Bland H, Evershed R.P, Possnert G, Pääbo S. 1998. Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth *Nothrotheriops shastensis*. *Science* 281: 402-406.

Poinar HN, Stankiewicz BA. 1999. Protein preservation and DNA retrieval from ancient tissues. *Proc Natl Acad Sci USA* 96:8426-8431.

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC. .2006 Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*; 311 (5759): 392-4.

Poulton J et Marchington DR. 2002. Segregation of mitochondrial DNA (mtDNA) in human oocytes and in animal models of mtDNA disease: clinical implications. *Reproduction* 123:751–55.

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

Q

Quintans, lvarez-Iglesias, Salas, Phillips, Lareu, Carracedo A. 2004. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Science International* 140 251–257.

R

Rasmussen EM, Sorensen E, Eriksen B, Larsen HJ, Morling N. 2002. Sequencing strategy of mitochondrial HV1 and HV2 DNA with length heteroplasmy. *Forensic Sci Int* 129:209–213.

Raymond M. et Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86:248-249.

Relethford JH .2001. Ancient DNA and the origin of modern humans. *Proc Natl Acad Sci. USA* 98 :390-391.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779.

Ricaut FX, Kolodesnikov S, Keyser-Tracqui C, Alekseev AN, Crubezy E, Ludes B. 2004. Genetic analysis of human remains found in two eighteenth century Yakut graves at At-Dabaan. *Int J Legal Med.* 118(1): 24-31.

Ricaut FX, Keyser-Tracqui C, Crubezy E, Ludes B. 2005. STR-genotyping from human medieval tooth and bone samples. *Forensic Sci Int.* 151(1): 31-5.

Ricaut FX, Kolodesnikov S, Keyser-Tracqui C, Alekseev AN, Crubezy E, Ludes B. 2006. Molecular genetic analysis of 400-year-old human remains found in two Yakut burial sites. *Am J Phys Anthropol.* 129(1): 55-63.

Richards M.B., Sykes B.C., Hedges R.E.M. 1995. Authenticating DNA extracted from ancient skeletal remains, *J. Archaeol. Sci.* 22: 291–299.

Richter C, Park JW, Ames BN. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci USA* 85 :6465-6467.

Robbins CR. 1985. The morphology and chemistry of human hair, in: Proceedings of the International Symposium on Forensic Hair Comparisons, FBI, Quantico, Viginia, pp. 3–21.

S

Saiki, R., Scharf, S., Falloona, F., Mullis, K., Horn, G., and Erlich, H. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-54.

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjærg L, Gedde-Dahl T, Issel-Tarver L, DiRienzo A, Pääbo S. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Gen Res* 5:42-52.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, et al. 2004. The African diaspora: mitochondrialDNAand the Atlantic slave trade. *Am. J. Hum. Genet.* 74: 454–65.

Salas A, Quintans, Alvarez-Iglesias. 2005. SNaPshot typing of mtDNA coding region variants. *Methods Mol Biol.* 297:197-208.

Sampietro ML, Gilbert MT, Lao O, Caramelli D, Lari M, Bertranpetti J, Lalueza-Fox C. 2006. Tracking down human contamination in ancient human teeth. *Mol Biol Evol.* 23(9): 1801-1807.

Santos FR, Pandya A, Tyler-Smith C. 1998. Reliability of DNA-based sex tests. *Nat Genet.* 18(2): 103.

Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ. 1999. The central Siberian origin for Native American Y chromosome. *Am J Hum Genet* 64:619-628.

- Savage RE, Plaut W. 1958. The effect of HCl hydrolysis on the retention of thymidine in DNA. *J Biophys Biochem Cytol.* 4(6): 701-6.
- Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki, Pääbo S, Smith FH. 2002. The Neandertal type site revisited : interdisciplinary investigations of skeletal remains from the Neander valley, Germany. *Proc Natl Acad Sci USA* 99(20) :13342-13347.
- Schneider P, Bender K, Mayr W, Parson W et al. 2004. STR analysis of artificially degraded DNA—results of a collaborative European exercise. *Forensic Science International.* 139: 123–134.
- Schneider S, Roessli D, ExcoYer L. 2000. Arlequin version 2.000: a software for population genetics analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva
- Scholz M, Giddings I, Pusch CM. 1998. A Polymerase Chain Reaction Inhibitor of Ancient Hard and Soft Tissue DNA Extracts Is Determined as Human Collagen Type I. *Analytical Biochem.* 259: 283-286
- Schultes T, Hummel S, Herrmann B. 2000. Ancient DNA-typing approaches for the determination of kinship in a disturbed collective burial site. *Anthropol Anz.* 58 (1): 37-44.
- Schurr TG, Subernik RI, Starikovskaya YB, Wallace DC. 1999. Mitochondrial DNA variation in koryaks and itel'men:population replacement in the okhotsk sea-bering sea region during the neolithic. *Am J Phys Anthropol* 108:1-39.
- Schurr TG. 2004. Molecular Genetic Diversity of Indigenous Siberians: Implications for Ancient DNA Studies of Cis-Baikal Archeological Populations. Northern Hunter-Gatherers: State of the Knowledge, eds. Hugh McKenzie and Andrzej Weber. Canadian Circumpolar Institute: Edmonton
- Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am. J. Hum. Biol.* 16:420–39.
- Schwartz M, Vissing J. 2002. Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med.* 347:576–80.
- Schwartz M, Vissing J. 2003. New patterns of inheritance in mitochondrial disease. *Biochem. Biophys. Res. Commun.* 310:247–51.
- Sekiguchi K, Sato H, Kasai K. 2004. Mitochondrial DNA heteroplasmy among hairs from single individuals. *J Forensic Sci.* 49(5): 986-91.
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G, Paabo S. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol.* 2(3): E57.
- Shewale, JG, Richey SL, Sudhir, SK. 2000. Anomalous Amplification of the Amelogenin Locus Typed by AmpFLSTR Profiler Plus Amplification Kit. *Forensic Science Communications.* Vol 2 (4).

- Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet.* 53(3): 549-62.
- Shimada MK, Kim C, Takahashi A, Spitsyn VA, Ikeo K, Gojorobi T, Saitou N. 2001. Mitochondrial DNA control region sequences for a Buryats population in Russia Unpublished
- Shitara H, Hayashi JI, Takahama S, Kaneda H, Yonekawa H. 1998. Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics* 148:851–57.
- Siéroszewski W. 1993. Jakuty. Opyt ètnografi eskogo issledovanija (Les Yakoutes. Expérience d'une recherche ethnographique) rééd., Moscou 736 p.
- Simon MC., Gray DI., Cook N. 1996. DNA extraction and PCR methods for the detection of Listeria monocytogenes in cold-smoked salmon. *Appl. Environ. Microbiol.* 62:822–824.
- Smith BC., Fischer DL., Weedn, VW. 1993. A systematic approach to the sampling of dental DNA. *Journal of Forensic Sciences*, 38, 1194-1209.
- Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, Collins M. J. 2001 Neanderthal DNA: not just old but old and cold? *Nature* 10, 771–772.
- Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *J Hum Evol* 45: 203-207.
- Soltis PS., Soltis DE, Smiley CJ. 1992. An rbcL sequence from a Miocene Taxodium (bald cypress). *Proc. Natl Acad. Sci. USA* 89, 449–451.
- Spotheim-Maurizot M, Begulosa M, Charlier M. 2003. Dégradation de l'ADN par les rayonnements ionisants. L'actualité Chimique, nov-déc : 97-102.
- Stephan E. 1997. Patterns of chemical change in fossil bones and various states of bone preservation associated with soil conditions, *Anthropozoologica*, 25–6, 173–80.
- Starikovskaya YB, Subernik RI, Schurr TG, Kogelnik AM, Wallace DC. 1998. mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient beringia and the peopling of the new world. *Am J Hum Genet* 63:1473-1491.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, 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(Pt 1): 67-89.
- Stiller M., Green R, Ronan M, Simons J, Du L, He W, Egholm M, Rothberg J, Keates S, Ovodov N, Antipina E, Baryshnikov G, Kuzmin Y, Vasilevski A, Wuenschell G, Termini J, Hofreiter M, Jaenicke-Despre V, Pääbo S. Patterns of nucleotide misincorporations

during enzymatic amplification and direct large-scale sequencing of ancient DNA. PNAS. 103 (37): 13578–13584.

Stone AC, Milner GR, Paabo S, Stoneking M. 1996. Sex determination of ancient human skeletons using DNA. Am J Phys Anthropol. 99 (2):231-8.

Stone AC, Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the NewWorld. Am. J. Hum. Genet. 62:1153–70.

Suenaga E, Nakamura H. 2005. Evaluation of three methods for effective extraction of DNA from human hair. J Chromatogr B Analyt Technol Biomed Life Sci. 820(1): 137-41.

Sullivan KM, Manucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test : fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15 : 636-641.

Szibor R, Michael M. 1999. Correct mitochondrial L-strand sequencing after C-stretches. Int J Legal Med 112:348–349.

Sweet D, Sweet C, 1995, DNA Analysis of dental pulp to link incinerated remains of homicid victim to crime scene. Journal of Forensic Sciences, 40, 310-314.

T

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res. 24 (16) 3189–3194.

Taylor, G.M., Rutland, P. and Molleson, T. 1997. A sensitive polymerase chain reaction method for the detection of Plasmodium species DNA in ancient human remains. Anc. Biomol. 1: 193-203.

Taylor GM, Young DB, Mays SA. 2005. Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. J Clin Microbiol ; 43 (5): 2236-40.

Tebbe CC et Vahjen W. 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. W.Appl Environ Microbiol. 59 (8): 2657-65.

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics. 13:718–25.

Templeton AR. 2005. Haplotype trees and modern human origins. Am J Phys Anthropol. 41 : 33-59.

Thalmann O, Hebler J, Poinar HN, Paabo S, Vigilant L. 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol Ecol 13: 321-335.

- Thèves C, Keyser-Tracqui C, Crubézy E, Salles JP, Ludes B, Thelmon N. 2006. Detection and quantification of the age-related point mutation A189G in the human mitochondrial DNA. *J Forensic Sc.* 51(4) : 865-873.
- Thomas RH, Schaffner W, Wilson AC, Pääbo S. 1989 DNA phylogeny of the extinct marsupial wolf. *Nature* 340, 465–467.
- Thyagarajan B, Padua RA, Campbell C. 1996. Mammalian mitochondria possess homologous DNA recombination activity. *J. Biol. Chem.* 271:27536–43
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5 (2): 123-35.
- Tolk HV, Barac L, Pericic M, Klaric IM, Janicijevic B, et al. 2001. The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. *Eur. J. Hum. Genet* 9:717–23.
- Torroni A, Subernik RI, Schurr TG, Starikovskaya YB, Cabell MA, Crawford MH, Comuzzie AG, Wallace DC. 1993. mtDNA variation of aboriginal siberians reveals distinct genetic affinities with native Americans. *Am J Hum Genet* 53:591-608.
- Triller, M., 1986, Histologie dentaire. (Paris : Masson), pp.210.
- Trinkaus E. 2001. The Neandertal paradox. In: Finlayson C, editor. Neanderthals and modern humans in late Pleistocene Eurasia. Gibraltar: The Gibraltar Museum. 73–74.
- Trueman CN, Martill DM. 2002. The long term survival of bone: the role of bioerosion. *Archaeometry* 44, 3: 371–382 .
- Tsai, YL. et Olson BH. 1991. Rapid method for direct extraction of DNA from soil and sediments. *Appl. Environ. Microbiol.* 57:1070-1074.
- Tsydendambaev, Ts.B. 1972. Buryatskie istoricheskie khroniki i rodoslovnye (Historical Chronicles and Pedigrees of Buryats), Ulan-Ude pp. 272–275.
- Turner CG II. 1987. Late Pleistocene and Holocene population history of East Asia based on dental variation. *Am J Phys Anthropol* 73:305–321.
- Tuross N. 1994. The biochemistry of ancient DNA in bone. *Experientia*. 50 (6): 530-5.

U

- Uchihi R, Tamaki K, Kojima T, Yamamoto T, Katsumata Y. 1992. Deoxyribonucleic acid (DNA) typing of human leukocyte antigen (HLA)-DQA1 from single hairs in Japanese, *J. Forensic Sci.* 37 (3): 853–859.
- Umetsu K, Tanaka M, Yuasa I, Saitou N, Takeyasu T, Fuku N, Naito E, Ago K, Nakayashiki N, Miyoshi A, Kashimura S, Watanabe G, Osawa M. 2001. Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. *Electrophoresis* 22 : 3533–3538.

Underhill PA. 2003. Inferring human history: clue from Y-chromosome haplotypes. *Cold Spring Harb Symp Quant Biol.* 68 : 487-93.

V

Valkovic V. 1988. Human Hair: Volume I, Fundamentals and Methods for Measurement of Elemental Composition, CRC Press, Boca Raton.

van der Kuyl, A.C., Kuiken, C.L., Dekker, J.T., Perizonius, W.R. and Goudsmith, J. 1995. Nuclear counterparts of the cytoplasmatic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies. *J. Mol. Evol.* 40: 652-657.

Vasan S, Zhang X, Zhang X, Kapurniotu A, Bernhagen J, Teichberg S, Basgen J, Wagle D, Shih D, Terlecky I, Bucala R, Cerami A, Egan J, Ulrich P. 1999. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature.* 382 (6588): 275-8.

Voronkin MS (1999) Dialektnaja sistema jazyka Saxa. [The dialect system of the Saxa language]. 'Nauka'. Sibirskaja izdatel'skaja Wrma RAN, Novosibirsk. In: Pakendorf

W

Wallace DC. 1989. Mitochondrial DNA mutations and neuromuscular disease. *Trends Genet.* 5 (1): 9-13.

Walsh PS, Metzger DA, Higuchi R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10 :506-513.

Wakeley J. 1993. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* 37:613–623.

Ward RH, Redd A, Valencia D, Frazier B, Pääbo S. 1993. Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90:10663-10667.

Watson RJ et Blackwell B. 2000. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Can. J. Microbiol.* 46: 633–642.

Watt K.E. 2005. Decontamination techniques in ancient DNA analysis. Thèse : 115p

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, Jamarashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsot I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244– 10249.

Whitaker JP, Cotton EA, Gill P. 2001. A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci Int.* 123(2-3): 215-23.

- Whiteman M, Jenner A, Halliwell B. 1997. Hypochlorous acid-induced base modifications in isolated calf thymus DNA. *Chem Res Toxicol.* 10(11): 1240-6.
- Whiteman M, Hong HS, Jenner A, Halliwell B. 2002. Loss of oxidized and chlorinated bases in DNA treated with reactive oxygen species: implications for assessment of oxidative damage in vivo. *Biochem Biophys Res Commun.* 296(4): 883-9.
- Wiedbrauk DL, Werner JC, Drevon AM. 1995. Inhibition of PCR by aqueous and vitreous fluids. *J Clin Microbiol.* 33(10) : 2643-2646.
- Widjojoatmodjo, M. N., A. C. Fluit, R. Torensma, G. P. H. T. Verdonk, and J. Verhoef. 1992. The magnetic immuno polymerase chain reaction assay for direct detection of salmonellae in fecal samples. *J. Clin. Microbiol.* 30:3195–3199.
- Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P, 1999. Diversity of Holocene life froms in fossil glacier ice. *Proc Natl Acad Sci* 96 : 8017-8021.
- Willerslev E., Hansen A. J., Brand T. B., Rønn R., Barnes I., *et al.*, 2004a Long-term persistence of bacterial DNA. *Curr. Biol.* 14: R9-R10.
- Willerslev E., Hansen, A. J., Poinar, H. N. 2004b Isolation of nucleic acids and cultures from ice and permafrost. *Trends Ecol. Evol.* 19, 141–147.
- Willerslev E et Cooper A. 2005. Ancient DNA. *Proc R Soc B* 272, 3-16.
- Whiting D et Howsden F. 1996. Color Atlas of Differential Diagnosis of Hair Loss. Canfield publishing Inc.
- Woodward, S. R., Weyand, N. J. & Bunell, M. 1994 DNA sequence from Cretaceous period bone fragments. *Science* 266, 1229–1232.

X

- Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, Xu J, Du R, Fu S, Li P, Yang H, Tyler-Smith C. 2005. Recent spread of a Y-chromosomal lineage in northern China and Mongolia. *Am J Hum Genet.* 77(6): 1112-6.

Y

- Yang DY, Eng B, Dudar JC, Saunders SR, Waye JS. 1997. Removal of PCR inhibitors using silica-based spin columns : application to ancient bones. *Can Soc Forens Sci* 30(1) :1-5.
- Yang D, Kane D L., Hinzman L D., Zhang X, Zhang T, Ye H. 2002. Siberian Lena River hydrologic regime and recent change. *J Geophys Res*, vol. 107, no. D23, 4694.
- Yang DY, Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. *Journal of Archaeological Science.* 32: 331–336.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet.* 70(3): 635-51.

Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP. 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol.* 20(2): 214-9.

Young CC, Burghoff RL, Keim LG, Minak-Bernero V, Lute JR, Hinton SM. 1993. Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. *Appl Environ Microbiol.* 59(6): 1972-1974.

Z

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174–1183.

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet.* 71(3): 466-82.

Zink A, Haas CJ, Reischl U, Szeimies U, Nerlich AG. 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol* ; 50 (4): 355-66.

Zink AR, Grabner W, 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 (1): 32-47.

Zischler H, Höss M, Handt O, von Haeseler A, van der Kuyl AC, Goudsmit J, Pääbo S. 1995 Detecting dinosaur DNA. *Science* 268, 1192–1193.

VII ANNEXES

VII.1 Annexe 1

Localisation géographique et classification chronologique des sujets analysés

n° Labo	Nom	Localisation	Période
YAKa25	Pokrovsk	Khangalassky	Tombes anciennes
YAKa27	3, squelette néolithique	Kolyma	
YAKa28	squelette néolithique, n°2	Kolyma	
YAKa43	Fedosseva, Allalaika	Vilyuysky	
YAKa44	Fedosseva, Tchersinsky	Vilyuysky	
YAKa47	Musée d'ethnologie	Khangalassky	
YAKa48	Fedosseva, Diring Urekh	Khangalassky	
YAKa21	Tabalakh (TA1)	Khangalassky	
YAKa26	Balyktakh (BKH)	Khangalassky	Période 1
YAKa40	Djoussou Len 1	Churapchinsky	
YAKa49	Jardin Botanique	Khangalassky	
YAKa65	Oulakhans Alas 1	Tattinsky	
YAKa66	Batta Tcharana	Tattinsky	
YAKa78	Bouogaryma 1	Tattinsky	
YAKa79	Bouogaryma 2	Tattinsky	
YAKa86	Tyyt Bappyt	Churapchinsky	Intermédiaire
YAKa69	Kous Tcharbyt	Tattinsky	
YAKa80	Nelegher	Tattinsky	
YAKa15	Ouhoraï	Khangalassky	
YAKa16	TD1	Khangalassky	
YAKa18	Kemus Bulunse 1 (KM1)	Khangalassky	
YAKa31	AE2	Khangalassky	
YAKa33	Kemus Bulunse 2 (KM2)	Khangalassky	
YAKa34	Arbre Chamanique 1 sujet 1	Churapchinsky	Période 2
YAKa35	Arbre Chamanique 1 sujet 5	Churapchinsky	
YAKa36	Arbre Chamanique 1 sujet 2	Churapchinsky	
YAKa37	Arbre Chamanique 1 sujet 3	Churapchinsky	
YAKa38	Arbre Chamanique 1 sujet 4	Churapchinsky	
YAKa39	Munur Urekh 1	Churapchinsky	
YAKa46	Munur Urekh 1bis	Churapchinsky	
YAKa50	Jarama 1	Churapchinsky	
YAKa52	Jarama 3	Churapchinsky	
YAKa55	Koulousoun Nakh 1	Churapchinsky	
YAKa64	Sette Toumoul	Tattinsky	
YAKa77	Ken Ebe 1	Tattinsky	
YAKa81	Orto Aryy	Tattinsky	
YAKa82	Okhtobout 2	Churapchinsky	
YAKa83	Kyys Ounouoga	Churapchinsky	

Localisation géographique et classification chronologique des sujets analysés (suite).

n° Labo	Nom	Localisation	Période
YAKa17	n°10	Khangalassky	
YAKa19	n°9	Khangalassky	
YAKa20	N°11	Khangalassky	
YAKa22	12	Khangalassky	
YAKa23	n°5	Khangalassky	
YAKa24	n°6	Khangalassky	
YAKa29	Rivière Tandy 1a (STR1a)	Khangalassky	
YAKa30	Rivière Tandy 1b (SRT1b)	Khangalassky	
YAKa32	Rivière Tandy 2 (SRT2)	Khangalassky	
YAKa41	Arbre Chamanique 2	Churapchinsky	
YAKa42	Koulousoun Nakh 2	Churapchinsky	
YAKa45	Arbre Chamanique 3	Churapchinsky	
YAKa51	Jarama 2	Churapchinsky	
YAKa53	Jarama 4	Churapchinsky	
YAKa54	Toumousaktakh 1	Churapchinsky	
YAKa56	Dirilaa Sayliga 1	Churapchinsky	
YAKa57	Bekh Alaas 1	Churapchinsky	
YAKa58	Bekh Alaas 2	Churapchinsky	
YAKa59	Bekh Alaas 3	Churapchinsky	
YAKa60	Bekh Alaas 4	Churapchinsky	
YAKa67	Ken Ebe 3	Tattinsky	
YAKa68	Ken Ebe 2	Tattinsky	
YAKa70	Seden	Tattinsky	
YAKa71	Ken Ebe 4	Tattinsky	
YAKa72	Ken Ebe 5	Tattinsky	
YAKa73	Ken Ebe 6	Tattinsky	
YAKa74	Ken Ebe 7	Tattinsky	
YAKa75	Ken Ebe 8	Tattinsky	
YAKa76	Ken Ebe 9	Tattinsky	

Période 3

VII.2 Annexe 2

Résultats des datations au radiocarbone obtenus pour 4 des sujets analysés.

n° Laboratoire	Nom	n° d'analyse	Datation
YAKa15	Ouhoraï	193454	1650 à 1680
YAKa26	Balyktakh	193453	1420 à 1470
YAKa25	Pokrovsk	198197	390 à 190 avant JC
YAKa40	Djoussoulen	198198	1440 à 1660

VII.3 Annexe 3

+ Models

ARTICLE IN PRESS



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Forensic Science International xxx (2006) xxx-xxx

Forensic
Science
International

www.elsevier.com/locate/forsciint

STR typing of ancient DNA extracted from hair shafts of Siberian mummies

S. Amory^{a,b,*}, C. Keyser^{a,b}, E. Crubézy^b, B. Ludes^a

^aLaboratory of Molecular Anthropology, Institute of Legal Medicine of Strasbourg, 11 rue Humann, 67085 Strasbourg Cedex, France

^bLaboratory of Anthropo-biology, UMR 8555, CNRS, 39 allées Jules Guesde, 31400 Toulouse, France

Received 5 May 2006; received in revised form 22 May 2006; accepted 24 May 2006

Abstract

The aim of this study was to determine if ancient hair shafts could be suitable for nuclear DNA analysis and to develop an efficient and straightforward protocol for DNA extraction and STR typing of ancient specimens. The developed method was validated on modern and forensic samples and then successfully applied on ancient hairs collected from Siberian mummies dating from the 16th to the early 19th centuries. In parallel extractions including or excluding a washing step were performed at least two times for each sample in order to evaluate the influence on the quantity of nuclear DNA yielded and on the typing efficiency. Twelve ancient individuals were analyzed through our approach and full and reliable profiles were obtained for four of them. These profiles were validated by comparison with those obtained from bone and teeth DNA extracted from the same ancient specimens. The present study demonstrates that the washing step cannot be considered as deleterious for DNA retrieval since the same results were obtained by the two approaches. This finding challenges the hypothesis that recoverable nuclear DNA is only found on the outer surface of hair shafts and provides evidence that nuclear DNA can be successfully extracted from ancient hair shafts. The method described here constitutes a promising way for non-invasive investigations in ancient DNA analysis for precious or historical samples as well as forensic casework analyses.

© 2006 Elsevier Ireland Ltd. All rights reserved.

Keywords: Ancient DNA; STR; Y chromosome; Hair shaft; Siberia

1. Introduction

Hairs as a potential source of DNA are considered as samples of interest (i) in forensic investigations because shed hairs are most often the evidentiary source on crime scenes and (ii) in archaeological research as a non-invasive method of analysis compared to bone or teeth DNA extraction which implies a destruction of precious remains [1]. Indeed, hair sampling is less destructive than the standard requirements necessary for bone or teeth analysis. Thus, nuclear DNA testing of ancient specimens could become a tremendous tool to complement the mitochondrial DNA data obtained for Amerindian mummies [2], Pazyryk horses and ancient bison [3] or extinct animal species such as cave bears [4]. The resistance against degradation and exogenous contamination

due to the histological and chemical properties of the cuticle are also valuable specificities [5].

Forensic and ancient DNA researchers have investigated the possibility of extracting DNA from hairs since 1988 [6]. The properties of mtDNA extracted from hair shafts have been described in depth by several authors [1,5,7–10] and a few research articles on the possibility to use STR typing on DNA extracts have also been published in recent last years [11]. Furthermore, McNevin et al. [12] recently published an optimized protocol for nuclear DNA extraction from modern hair shafts.

The aim of this study was to propose a straightforward and efficient protocol for hair lysis, DNA extraction and STR amplification applicable to ancient material as well as to forensic samples. STR genotyping is a powerful tool to investigate parental relationships among burial sites and the study of Y chromosome STR allows the determination of paternal lineages useful for population studies and settlement hypotheses.

* Corresponding author. Tel.: +33 390 243348; fax: +33 390 243362.
E-mail address: sylvain.amory@iml-up.u-strabg.fr (S. Amory).

The extraction process was validated on modern and forensic samples and the optimized method was tested on 12 hair samples collected during excavations of Siberian graves dating from the 16th to the 19th century.

Nuclear DNA is assumed to be contained in nuclei of epithelial cells adhering to the outer surface of the shaft [12] rather than in the shaft itself. We have thus investigated if a washing treatment prior to extraction could be deleterious for the subsequent nuclear DNA extraction. Hence, for each sample we have tested two parallel extractions, with or without washing the hairs. The genotypes obtained from hairs were then compared with those obtained for bone and teeth DNA extracts of the same individuals. This comparative analysis also allowed to test for the occurrence of amplification artifacts such as allelic dropout or spurious alleles and to evaluate if hair samples are more prone to biasing factors than bones and teeth.

To the best of our knowledge, this study is the first presenting successful amplification of nuclear DNA and STR typing of ancient DNA extracted from hair shafts.

2. Material and methods

2.1. Samples

Twelve individuals dating from the 16th to the early 19th century were analyzed in this study (see Table 1); these samples were collected during excavations performed by the MAFSO (French Archaeological Mission in Oriental Siberia) team in the years 2004 and 2005 in Central Yakutia. Yakutia is an Autonomous Republic of the Russian Federation located in the north-eastern part of Siberia. The extremely cold and dry climate of this region combined with the inhumations realized on the permafrost generally induces a good preservation of the bodies.

Moreover, the majority of the studied graves is characterized by a covering of several layers of birch bark inside and outside the coffin and thus constitutes a very efficient insulation from oxygen and humidity.

Hair samples were plucked directly from the scalp of the exhumed individuals and put in decontaminated plastic bags.

The length and the abundance of the collected samples varied between graves. All the specimens presented a hair pigmentation characteristic of the Asian population except YAKa 68 and YAKa 70. For these two individuals the absence of pigmentation might be due to a well-known postmortem phenomenon implying the loss of coloration after inhumation [13].

2.2. Measures against contamination

A number of measures were put into practice in order to prevent possible contamination. Excavations were performed wearing facial masks and latex gloves. All people handling the material or working in the laboratory were genotyped and their profiles were compared to those of the samples. All the steps of the analysis were performed in a laboratory dedicated to ancient DNA studies.

Open tube experimentations were made in a room separated from the rest of the laboratory. Laboratories, laboratory ware, coats, plastic ware, reagents, pipettes, benches and equipment were irradiated under ultraviolet light (250 nm for 45 min) after each experiment. Samples were handled wearing sterile gloves, face masks, hair nets and laboratory coats dedicated to ancient DNA. Reagents were prepared in small volumes to avoid multiple uses. Aerosol-resistant filter tips were used for all reactions with ancient DNA. Extraction and negative blanks were included in all PCR assays.

For each specimen, multiple, time-spaced extractions were performed and PCR amplifications were made at least two times to ensure the accuracy of the results.

2.3. DNA extraction

2.3.1. Preparation of the samples

Considering that the DNA in the proximal part is supposed to be less subjected to alteration and thus to be more suitable for DNA analysis, we decided to restrict our sampling to these parts. Furthermore, the progressive degradation of DNA along the shaft could even be more critical for nuclear DNA than for mtDNA [14]. Thus, 3 cm sections of the proximal part of hair shafts were selected. For samples shorter than 3 cm, the full

Table 1
Specimens chosen for hair samples analyzed in this study

Lab number	Site name	Morphological sex	Relative dating	Color/note
YAKa 34	Shamanic Tree 1	M	End 17th/18th century	Black/clean
YAKa 37	Shamanic Tree 1	F	End 17th/18th century	Brown/dirty
YAKa 39	Munur Urek 1	M	End 17th/18th century	Brown/dirty
YAKa 45	Shamanic Tree 3	F	18th century	Black
YAKa 46	Munur Urek 1'	F	18th century	Brown/clean
YAKa 55	Koulousoun Nakh 1	F	Early 19th century	Black/clean
YAKa 66	Batta Tcharana	M	16th century	Brown
YAKa 67	Ken Ebe 3	-	End 18th/early 19th century	Brown
YAKa 68	Ken Ebe 2	-	End 18th/19th century	Clear/dirty
YAKa 69	Kous Tcharbyt	M	16th/early 17th century	Brown/dirty
YAKa 70	Seden	M	Middle 19th century	Clear/clean
YAKa 79	Bouogaryma 2	F	Late 16th/early 17th century	Black/clean

The relative dating is based on the archaeological material associated with the body. “-” indicates that no morphological sex determination was possible on these immature individuals.

length shafts were used. For each sample, between five and seven hairs were chosen and cut into fragments of 0.5 cm. Hair fragments were then placed in 1.5 ml tubes in order to obtain a total length limited to 15/21 cm to reduce the possibility of a melanin inhibition [15,16].

2.3.2. Washing protocol

We have tested for each sample two parallel extractions with (W) and without (nW) a washing step. Hair samples were washed two times with 500 µl of 1% filtered sodium dodecyl sulfate (SDS) and then the samples were rinsed two times with DNase/RNase-free, sterile water and finally fixed with absolute ethanol. The samples were left to air dry under a hood dedicated to ancient DNA work. For dirty samples, the not washed hairs (nW) were rinsed once with DNase/RNase-free and sterile water.

Following the addition of the solution each step comprises brief vortexing, brief centrifugation and withdrawal of liquid with a pipette.

2.3.3. Digestion and extraction protocol

The samples were digested in 500 µl of lysis buffer overnight at 55 °C under vertical rotation. The lysis buffer contained 0.005 M of EDTA, 2% of SDS, 0.01 M of Tris-HCl (pH 8), 0.3 M of sodium acetate, 10 mg/ml of dithiothreitol (DTT) and 0.001 M of *N*-phenacylthiazolium bromide (PTB) and 0.5 mg of proteinase K. PTB is supposed to increase the yield of DNA retrieved after DNA extraction due to its property to cleave glucose-protein derived cross-links [17].

After the lysis phase hair samples were extracted using a phenol-chlorophorm protocol. In each tube 500 µl of phenol/chloroform/isoamyl alcohol (25/24/1, v/v) was added and then centrifuged for 5 min at 1000 × g. The supernatant was collected and transferred to a new 1.5 ml tube. The previous step was repeated a second time and the supernatant was then purified.

2.3.4. Purification and concentration protocol

The aqueous phase was purified with the CleanMix kit (Talent, Trieste, Italy) and each sample was eluted in 400 µl of DNase/RNase-free, sterile water. The samples were then concentrated on Microcon YM30 (Millipore, Billerica, MA) up to 30 µl.

2.3.5. Reference samples

Bone samples were available for all the samples and we were able to collect teeth for samples YAKa 34, YAKa 37, YAKa 39, YAKa 66, YAKa 67, YAKa 69 and YAKa 79. Extraction and amplification protocols for bone and teeth are fully described in Keyser and Ludes [18]. The results obtained from bone and teeth DNA extracts were validated using at least two extractions and two different PCR amplifications.

2.4. DNA quantitation

Quantitation was performed with the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol and carried out on an ABI Prism 7000 SDS (Applied Biosystems, Foster City, CA).

2.5. PCR amplification

2.5.1. Autosomal STRs

Autosomal STRs were amplified using the AmpF/STR® Identifier™ kit (Applied Biosystems, Foster City, CA). Enhanced PCR conditions were applied to all the samples using 34 cycles as recommended by Gill et al. [19]. Depending on DNA quantitation results the volume was adapted to fit the manufacturer's recommendation on the amount of DNA. PCR amplifications were performed according to manufacturer's protocol on a T3 Biometra thermal-cycler.

For subsequent GeneScan analysis, 1.5 µl of PCR product was added to 9 µl of Hi-Di Formamide (Applied Biosystems, Foster City, CA) and 0.5 µl of Genescan GS 500 LIZ (Applied Biosystems, Foster City, CA). Electrophoresis was carried out on an ABI Prism® 3100 Genetic Analyzer and electropherograms were analyzed using the GeneMapper® Software v1.02.

2.5.2. Y chromosome STRs

Y chromosome genotyping was performed using the AmpF/STR® Y-filer™ kit (Applied Biosystems, Foster City, CA). As for autosomal STR, the number of amplification cycles was increased up to 34. The manufacturer's recommendations were applied for the PCR mix preparation and amplifications were carried out on a T3 Biometra thermal-cycler. GeneScan analyses were performed with 1 µl of PCR product, 9.7 µl of Hi-Di Formamide and 0.3 µl of GS 500 LIZ on a ABI Prism® 3100 Genetic Analyzer and electropherograms were analyzed using the GeneMapper® ID software.

3. Results

3.1. DNA quantitation

In addition to the role of validation criterion [20] the DNA quantitation enabled us to adjust the DNA amount necessary for the different analyses and to determine the best PCR cycling conditions. Since an internal PCR control (IPC) is included in each reaction, it allowed us to monitor the presence of PCR inhibitors through the threshold cycle (C_t) value. For all DNA extracts (data available upon request) these values indicated the absence of PCR inhibitors as no delay was observed compared to the C_t values of the standard dilution series.

The DNA concentration varied greatly between the samples from 0 (undetermined value) for samples YAKa 45, YAKa 46, YAKa 55, and YAKa 68 to 0.945 ng/µl for the second extraction of sample YAKa 66 nW. For the majority of the samples, the DNA concentration was reproducible between two different extractions for the same individual. Furthermore, DNA quantities were relatively low since none of the samples showed a DNA concentration greater than 1 ng/µl.

No correlation was observed between the DNA conservation in bone and teeth and the yield of the hair DNA extracts of the same individual (see Fig. 1). For instance, the highest concentration was observed for the bone DNA extract of YAKa 37. Nevertheless, this specimen failed to give any results

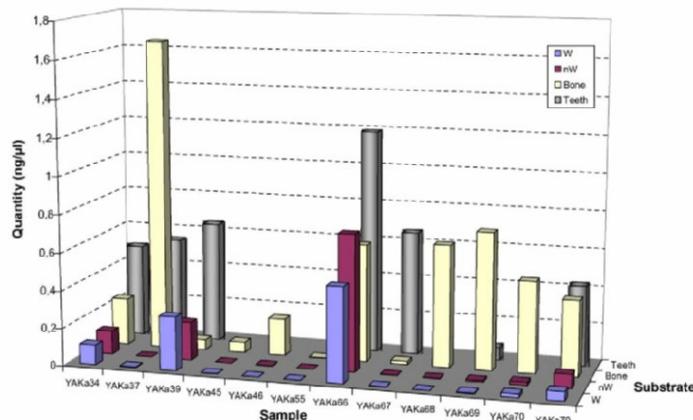


Fig. 1. Mean value of the real-time PCR DNA quantitation performed on the different substrates available for each sample. Note that teeth samples were only available for YAKa 34, YAKa 37, YAKa 39, YAKa 66, YAKa 67, YAKa 69 and YAKa 79. "W": washed samples, "nW": unwashed samples.

for STR typing except for the amelogenin locus (see Table S1, Supplementary data). Conversely, for YAKa 39 we observed that DNA was poorly preserved in bone while hair samples yielded enough DNA to allow full profile amplification. These results suggest that specimens failing to give any results with bone samples could be successfully typed for nuclear markers through hair samples, if available.

3.2. STR results

3.2.1. Autosomal STRs

For all the mature specimens, both morphological and molecular sex determinations were in accordance (see Table 2). Concerning the two immature individuals, the amplification of the amelogenin locus was successful for the YAKa 68 specimen revealing a female genotype. An allelic dropout occurred for YAKa 67 as only the X marker was detected and the comparison with profiles obtained from other substrates revealed a male genotype (see Table S1, Supplementary data).

For all the samples we performed two extractions separated in time and two different amplifications for each DNA extract (see Table 2). Problematic samples such as YAKa 37 were extracted up to four times. Thus, for each individual, we compared at least eight allelic profiles in order to determine the consensus genotype of an individual.

Among the 12 tested specimens we obtained complete allelic profiles for four individuals (YAKa 34, YAKa 39, YAKa 66 and YAKa 70) and for two of them (YAKa 69 and YAKa 79) we determined a consensus profiles composed of 9 and 12 loci out of a total of 16, respectively. The six remaining samples (YAKa 37, YAKa 45, YAKa 46, YAKa 55, YAKa 67 and YAKa 68) gave only few or no amplified alleles.

Considering the occurrence of amplification artifacts, no major differences were observed between allelic profiles of washed and unwashed hairs. The quality of STR genotyping

was in accordance with the results of the Quantifiler™ kit except for samples YAK 70 and YAK 79. Indeed, YAK 70 gave almost complete profiles for each amplification, whereas YAK 79, showing a higher concentration, yielded only partial profiles (see Fig. 1).

The typing efficiency for the different loci was evaluated and reported in Fig. 2. For each locus, we compared the profile of the sample and the genotype of the corresponding reference (i.e. bone/teeth). The different kinds of typing events were numbered, then a percentage was calculated for washed and unwashed DNA extracts of the same specimen.

Bone and teeth DNA extracts were chosen as references in spite of the possibility of allelic dropouts. In order to estimate the potential influence of this bias, we evaluated the frequency of homozygous genotypes in a sample of 97 contemporary, unrelated Yakut individuals. All these genotypes have been retrieved in our modern samples except the 18/18 for D2S1338 found for YAKa 66. Thus we assessed that the proportion of homozygotes in our reference samples was not superior to that observed in the modern sample.

The proportion of correct alleles observed during our experiments was high and wrong typing occurrences (for DSS818, D2S111, D16S539 and FGA) remained scarce. Peak imbalance and allelic dropouts were observed for all loci with the proportion increasing in relation to the size of the studied marker. This pattern for the typing efficiency was observed for both protocols. The highest ratio of dropouts was observed for the CSF1PO locus in both washed and unwashed extracts. The marker presenting the poorest results was D2S1338 for the two protocols. For these two loci, amplification products are longer than 300 bp. Therefore, a clear correlation was observed between the fragment length of the marker and the success of amplification.

Extractions blanks were exempt of contamination since no genotypes were amplified, nevertheless low intensity peaks

Table 2
Results of STR genotyping using the AmpFISTR® Identifier™

Lab number	Ext.	Amel	CSF1PO	D13S317	D16S539	D18S51	D19S433	D21S11	D2S1338	D3S1358	D5S818	D7S820	D8S1179	FGA	TH01	TPOX	vWA	
YAKa 34 W	1	XY		10 11		15 15		29 29	15 16	10 12	8 8	13 16	25 25			17 19		
		XY		10 11		(10) 12		14 17.2	29 29	23 23	15 16	10 12	13 16	21 25	8 8	17 17		
	2	XY		10 11	13 13		14 14			15 16	10 12	13 13			8 8	11 11	17 17	
	3	XY		11 11		13 13	14 17.2	29 29	17 17	15 16	10 12	8 8	13 16	(21) 25	8 9.3	11 11	17 19	
		XY	9 9	10 11	13 13	13 13	14 17.2	29 29	17 17	15 16	10 12		13 16	21 25	8 9.3	11 11	17 19	
YAKa 34 nW	1	XY		10 11				29 29	15 16	10 12	8 8	13 16	21 25			17 19		
		XY	9 9	10 11	9 9		14 17.2	29 29	15 16	10 12	8 8	13 16	21 21	8 9.3	11 11	19 19		
	2	XY			13 13		14 17.2	29 29	15 16	10 12		13 16	21 25	8 9.3	11 11	17 18 19		
	3	XY	(14 14)	10 11	9 13	13 13	14 17.2	29 29	17 17	15 16	10 12	8 8	13 16	25 25	8 9.3	11 11	17 19	
		XY	10 10	9 13	13 13	14 17.2	29 29		15 16	10 12	8 8	13 16	21 21	8 9.3	11 11	17 19		
Consensus Bone/teeth	XY	9 9	10 11	9 13	13 13	14 17.2	29 29	17 17	15 16	10 12	8 8	13 16	21 25	8 9.3	11 11	17 19		
	XY	9 14	10 11	9 13	13 15	14 17.2	29 29	17 23	15 16	10 12	8 8	13 16	21 25	8 9.3	12 11	17 19		
YAKa 39 W	1	XY		8 13				29 30	15 16	11 12	8 8	13 14	19 24			14 16		
		XY	(8 8)			(13 13)			(15 16)	(11 12)	(13 14)		(7 9.3)			(14 16)		
	2	XY	8 13	9 11	13 13	13 13	29 30	20 20	15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16		
	XY	10 10	8 8	9 11		13 13	29 30		15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16		
YAKa 39 nW	1	XY	8 13					30 30	15 16	11 12	8 8	13 14	19 24			14 16		
		XY	(10 10)	(8 13)		13 13	29 30		15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16		
	2	XY	12 12	8 13	9 11	13 17	13 13	29 30	19 20	15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16	
	XY	12 12	8 13	9 11		17 17	13 13	29 30	19 20	15 16	11 12	8 8	13 14	22 24	(7 9.3)	8 11	14 16	
Consensus Bone/teeth	XY	12 12	8 13	9 11	13 13	13 13	29 30	19 20	15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16		
	XY	10 12	8 13	9 11	13 17	13 17	29 30	19 20	15 16	11 12	8 8	13 14	19 24	7 9.3	8 11	14 16		
YAKa 66 W	1	XY		11 11		13 13		29 31	16 17	12 13	10 11	11 16	20 25			17 19		
		XY	11 11	9 9	13 13	13 2 15	29 31	18 18	16 17	12 13	10 10	11 16	20 20	9 3 9.3	8 8	17 19		
	2	XY	11 11	11 11	9 9	13 13	13 2 15	29 31	18 18	16 17	12 13	10 11	11 16	20 25	9 3 9.3	8 8	17 19	
	XY	11 12	11 11	9 9		13 13	13 2 15	29 31	18 18	16 17	12 13	10 11	11 16	20 25	9 3 9.3	8 8	17 19	
YAKa 66 nW	1	XY		11 11		13 13		29 31	16 17	12 13	10 11	11 16	20 25			17 19		
		XY				13 2 15			16 16			11 16		9 3 9.3	8 8			
	2	XY	11 12	11 11	9 9	13 13	13 2 15	29 31	18 18	16 17	12 13	10 11	11 16	20 25	9 3 9.3	8 8	17 19	
	XY	11 11	11 11	9 9		13 13	13 2 15	29 31	18 18	16 17	12 13	10 11	11 16	20 25	9 3 9.3	8 8	17 19	
Consensus Bone/teeth	XY	11 12	11 11	9 9	13 13	13 2 15	29 31	18 18	16 17	12 13	10 11	11 16	20 25	9 3 9.3	8 8	17 19		
	XY	11 12	8 10	9 9	14 20	13 13.2	30 30	17 19	15 17	10 13	8 10	13 16	22 24	9 9	8 9	14 16		
YAKa 69 W	1	YY						17 17								14 14		
	2	YY						13 13.2			11 11		13 13		8 8			
YAKa 69 nW	1	XY																
		XY						13.2 13.2			15 15		10 10		13 13			
	2	XY						13.2 13.2			15 15		10 13		10 10			
	XY							13.2 13.2			13 16		24 24				14 14	
Consensus Bone/teeth	XY	—	10 10	—	—	13 13.2	—	—	15 15	—	10 10	13 16	24 24	—	8 8	14 14		
	XY	11 12	8 10	9 9	14 20	13 13.2	30 30	17 19	15 17	10 13	8 10	13 16	22 24	9 9	8 9	14 16		
Table 2 (Continued)	1	XX																
	2	XX																
	3	XX																
	XX	10 10	8 8															
YAKa 79 W	1	XX																
	2	XX																
	3	XX																
	XX	10 10	8 8															
YAKa 79 nW	1	XX																
	2	XX																
	3	XX																
	XX	10 10	8 8															
Consensus Bone/teeth	XX	10 10	8 8	—	—	14 14.2	29 30	—	16 17	12 12	8 8	10 10	20 20	9 3 9.3	—	16 18		
	XX	10 10	8 11	9 10	13 14	14 14.2	29 30	20 20	16 17	12 12	8 11	10 10	20 24	6 9.3	8 11	16 18		
Team 1	XY	9 10	9 11	10 13	13 16	13 15	33.2 33.2	23 25	17 18	12 13	9 10	12 13	20 20.2	6 7	8 11	16 18		
Team 2	XY	11 12	12 12	9 11	14 15	13 14	27 31.2	22 24	14 18	9 13	8 10	8 15	20 20	6 9.3	8 8	17 19		
Team 3	XY	10 10	12 13	11 11	14 14	12 13	31 32.2	17 20	16 17	11 12	9 12	13 13	22 25	7 9.3	8 8	17 19		
Team 4	XX	11 11	11 14	9 11	13 13	15 16	27 31	16 22	16 17	12 12	7 11	13 13	20 23	6 6	8 8	17 19		

This table shows only samples yielding more than 3 amplified loci. The bold typed genotype is the consensus profile determined for the considered individual. The bone/teeth genotype corresponds with the consensus profile determined after comparison of bone and teeth DNA extracts amplifications. Alleles into brackets are inferior to 150 RFUs.

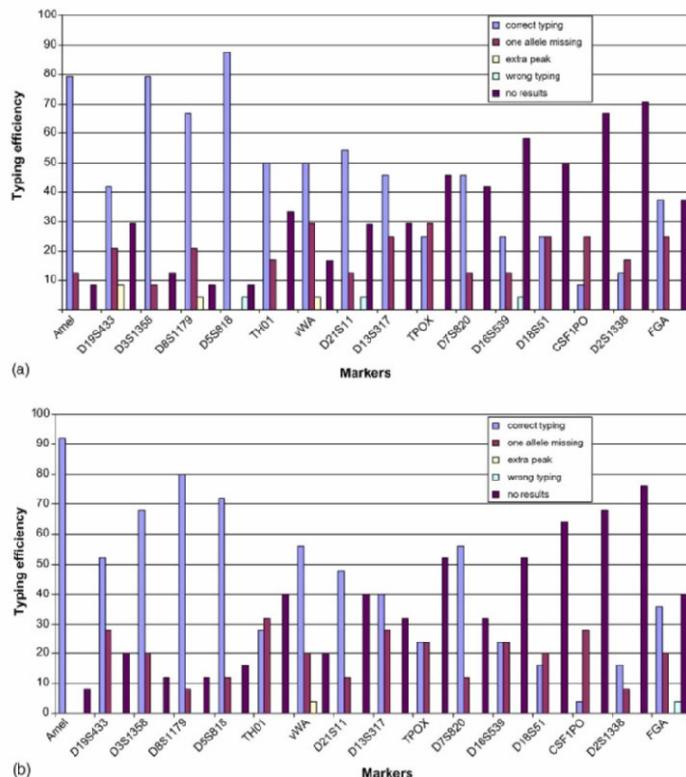


Fig. 2. Percentage of typing efficiency for autosomal STR genotyping for hair samples (a) submitted to the washing protocol and (b) without a washing treatment. Markers are ordered from the smallest to most important molecular weight. Samples with less than three amplified loci were not considered for calculation since the poor results come from the DNA conservation rather than the PCR amplification process.

were noted for some of them (see Table 3). These peaks were not reflected in the electropherogram of the corresponding sample. The presence of such peaks may come from the amplification conditions specific to LCN analyses [19]. As the

enhanced PCR cycles allowed the detection of only one copy of DNA, we assumed that these spurious alleles are due to single chromosome or cell debris contamination [19,21,22]. PCR reagents contaminations were excluded since no detectable alleles were found in any of the PCR negative controls (data not shown).

3.2.2. Y chromosome STRs

Male specimens were tested for Y chromosome analysis. As expected, the individuals yielding a complete autosomal profile gave complete Y haplotypes (YAKa 34, YAKa 39, YAKa 66 and YAKa 70) whereas YAKa 69 yielded incomplete profiles (see Table 4). The Y haplotypes of samples YAKa 34 and YAKa 39 were identical and concord with the most common haplotype observed in both modern and ancient Yakuts. The predominance of this Y haplotype in the Yakut population was explained by a significant reduction of Y chromosomal genetic variability through time [23].

Table 3
Results of autosomal STR typing performed on the extraction blanks

Lab Number	Ext.	Amel	D13S317	D18S51	D2S1338	FGA	TH01
YAKa 34	3	XY					
YAKa 37	1			12	17		
	2		12				
YAKa 46	1	Y					
YAKa 55	1				21		
YAKa 67	1					9	
YAKa 79	1	XY					
	2		12				

Only blanks yielding amplified loci are presented.

Table 4
Results of Y chromosome STR genotyping performed with the AmpF/STR[®] Y-filerTM

Lab Number	Ext.	DYS456	DYS389I	DYS389II	DYS390	DYS458	DYS19	DYS385ab	DYS393	DYS391	DYS399	DYS635	DYS439	Y GATA H4	DYS437	DYS438	DYS439	
YAKa 34 W	1	14	14	23	23	16	14	11/13	14	11	10	16	12	14	11	11	11	
	3	14	14	23	16	14	14	11/11	14	11	10	16	12	14	11	11	19	
	3	14	14	23	16	14	11/11	14	11	10	22	16	12	14	11	11	19	
YAKa 34 nW	1	14	14	32	23	16	14	11/11	14	11	10	22	16	12	14	11	11	
	3	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	
	3	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	
Consensus	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
Bone/teeth	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
YAKa 39 W	1	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19
	2	14	14	32	23	16	14	11/11	14	11	10	22	16	12	14	11	11	19
	2	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19
YAKa 39 nW	1	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19
	2	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19
	2	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19
Consensus	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	19	
YAKa 66 W	1	14	31	23	16	14	13/13	14	11	10	22	16	12	14	11	11	20	
	14	14	(31)	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	2	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	20	
Consensus	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	20	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	20	
	14	14	32	23	16	14	11/13	14	11	10	22	16	12	14	11	11	20	
Bone/teeth	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
YAKa 66 nW	1	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
Consensus	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
Bone/teeth	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
YAKa 69 W	1	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
YAKa 69 nW	1	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20
Consensus	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
	14	14	31	23	16	14	11/13	14	11	10	22	15	12	14	11	11	20	
YAKa 70 W	1	14	14	31	23	16	14	11/11	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/11	14	11	10	22	15	12	14	11	11	20
	2	14	14	31	23	16	14	11/11	14	11	10	22	15	12	14	11	11	20

		DYS456	DYS389I	DYS389II	DYS390	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	Y GATA H4	DYS437	DYS438	DYS448
YAKa 70 nW	1	14	14	23	14	11	11	14	11	10	10	15	12	12	14	14	20
	2	14	14	23	16	14	11	11	14	11	10	22	12	12	11		
Consensus		14	14	31	23	16	14	14	14	11	10	22	15	12	14	11	20
Bone/teeth		14	14	31	23	16	14	11	13	14	11	10	22	15	12	14	11
Team 1		14	14	31	23	16	14	11	13	14	11	10	22	16	12	14	19
Team 2		15	12	28	24	17	14	11	14	14	11	13	23	13	11	14	12
Team 3		14	12	28	22	15	14	13	16	14	10	11	21	11	11	16	10

The bold typed genotype is the consensus profile determined for the considered individual. The bone/teeth genotype is the consensus profile determined after comparison of bone and teeth DNA extracts amplifications.

The typing efficiency for these multiplexes reactions was also evaluated (see Fig. 3). As already highlighted for autosomal STRs, the loss of typing efficiency was closely related to the increase of the marker's fragment length. This correlation between the size range of the marker and the typing efficiency was even clearer for Y chromosome STR. Allelic dropout cannot be tested since most of the loci of the Y-Filer® kit are mono-allelic. However, for the bi-allelic marker, DYS385a/b, allelic dropout was frequently observed. One wrong allele was detected for the DYS635 locus and no evidence was found for an extra peak event.

Negative controls were exempt from any peak but two and one alleles were observed for extraction blanks of YAKa 34 and YAKa 39, respectively. These alleles were not found in the corresponding sample DNA extracts. The observed intensity of these alleles was between 40 and 150 RFUs. This low level of contamination could be attributed to the analyst since they concord with markers found in the team members' profile. Nevertheless, this hypothesis was rejected since the samples showed haplotypes corresponding neither with the analyst's profiles nor with the archaeologists' haplotypes.

4. Discussion

Numerous studies have investigated the possibility to analyze mitochondrial DNA extracted from ancient hair shafts. More recently nuclear DNA analysis has been conducted but so far, STR typing of hair samples has been realized in the forensic field only on fresh or recent hairs.

In the present study a straightforward protocol for extraction and amplification was successfully applied to ancient hair shafts. In order to evaluate the way in which the washing step reduces the chance of achieving successful nuclear DNA amplification, we performed all experiments on washed and unwashed samples. Contrary to results observed by McNevin et al. [12], we found no differences on our samples for DNA quantitation and for STR typing between cleaned and not cleaned hairs. The washing step did not induce a deleterious effect on the quantity of extracted DNA. The variations between the two extraction protocols on the final DNA concentration seem to be more related to stochastic variations than to a clear pattern since some samples gave better results with the washing protocol whereas other samples showed the contrary. For STR typing, the amplification results did not differ between the two protocols and the proportion of amplification artifacts was equal for the two types of DNA extracts. This finding revealed that our samples contained intact or, at least, well preserved nuclei in the shaft itself and not only trapped in cuticle scales [12]. Hence, our data showed that the nuclei present in ancient hairs 15–20 cm long seem to be sufficient to obtain a complete profile. These results could also depend on the phase of hairs studied and on the taphonomic conditions. As the hair samples were plucked directly from the scalp we estimated that they were in an anagen phase at the time of the death and that the keratinisation process responsible for the DNA degradation had stopped. The taphonomic conditions were certainly variable between the studied burial sites and these variations could, in

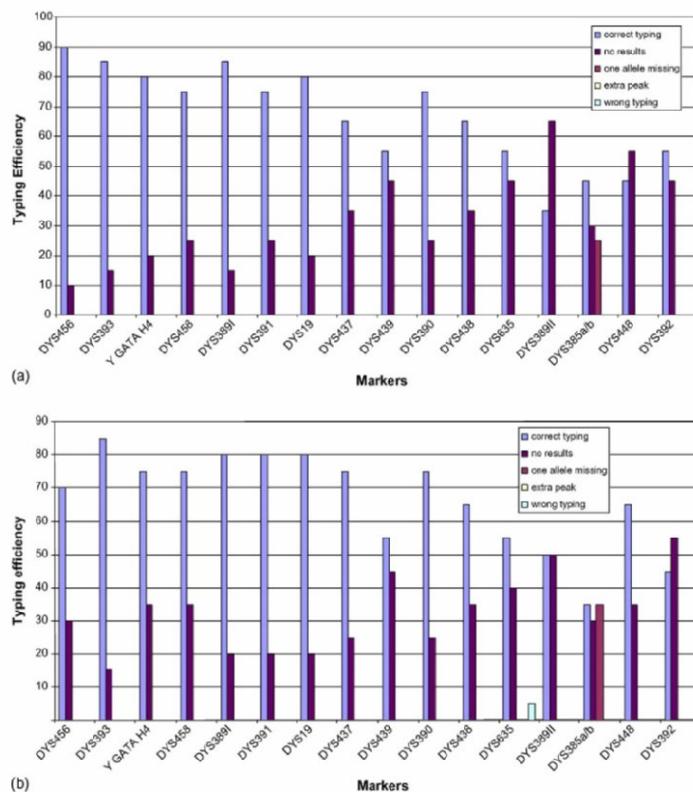


Fig. 3. Percentage of typing efficiency for Y chromosome genotyping for hair samples (a) submitted to the washing protocol and (b) without a washing treatment.

some way, explain the differences of the amplification success observed among our samples. The absence of results for sample YAKa 45, YAKa 46, YAKa 55, YAKa 67 and YAKa 68 could, therefore, be explained by the taphonomic factors such as environmental exposure due to the absence of birch bark cover, inclusion of sediment due to a break in the coffin, inhumation in a zone of shallow permafrost, etc. Nevertheless, two cases remained problematic: YAKa 37 and YAKa 46 were respectively buried in the same tomb as YAKa 34 and in a grave located few meters away from YAKa 39. The discrepancies observed in the results obtained for these individuals are unexpected but could be explained by micro-variations of the taphonomic conditions since the suitability for genetic analysis of closely located samples could be completely different, even for the same bone [24].

The percentage of correct typing was evaluated for each marker and we noted that the typing efficiency was closely related to the size of the locus and varied from 4% for CSF1PO to more than 92% for the amelogenin locus. Allelic dropouts were detected for all the loci but the occurrence of

such amplification artifacts was clearly worsened by the increase of the marker size. Wrong typing events (for the D5S818, D21S11, D16S539 and FGA) and the detection of extra peaks (for the D19S433, D8S1179 and vWA) were rare. Therefore, our approach proved to be efficient since we experienced low occurrences of such amplification artifacts compared to data published on modern samples [11,12]. The presence of extra peaks could result from slippage of the polymerase [11] or to a low level of contamination because the enhanced PCR protocol is sufficiently sensitive to detect single DNA molecules [19]. Furthermore, the melanin might decrease the PCR efficiency for high molecular weight markers. Indeed, this protein causes a reduction of the processivity of Taq polymerase and of the elongation rate which is more critical for large fragments [16]. However, the C_t values of the IPC included in the Quantifiler kit did not show any delay. Thus we concluded that the use of a reduced quantity of hair and the successive organic extraction procedure combined with the purification steps may have prevented an inhibition by melanin.

In order to minimize the risk of potential contamination, extensive precautions were put into practice (see Section 2). Despite the fact that not all reported criteria of authenticity [20] could be met, the possibility that our data came from exogenous DNA contaminations was unlikely for several reasons.

- (i) Even though the results obtained via real-time PCR quantitation indicated relatively low concentrations, the DNA quantities observed were sufficient for STR typing. Moreover, for most of the samples, positive amplification results were in accordance with the results of the real-time PCR assay. Nevertheless, the size of the quantitation assay target (62 bp) allows the detection of highly fragmented DNA molecules that are not suitable for STR typing. This bias was observed for YAKa 70 and YAKa 79 and the relationship between a high concentration revealed by RT-PCR and the success of the subsequent STR typing was sometimes unreliable.
- (ii) For all the mature individuals the results of morphological and amelogenin-based sex determination were in accordance. This finding is considered to be an indication for the authenticity of the results.
- (iii) As mentioned previously, the typing efficiency of the two studied multiplex systems revealed a clear correlation between the amplification success and the fragment length. This molecular behavior is a well-known property of LCN and degraded DNA molecules [25]; moreover, this feature has been validated as an authenticity criterion [20].
- (iv) The blanks of each extraction were analyzed for both molecular markers. Low intensity peaks were sporadically detected, nevertheless, these alleles were not reflected in the corresponding samples. Negative controls were performed for each PCR amplification and no reportable allele was detected.
- (v) The reproducibility of the results was thoroughly evaluated since at least four different extracts were generated per specimen and two different PCR amplifications performed. The eight allelic profiles generated on each extract were similar and the comparison with results obtained for bone and teeth samples indicated a full concordance for all the specimens.

In conclusion, the present study shows for the first time that nuclear DNA could be successfully extracted from ancient hair shafts. In spite of their rarity, hairs could be considered as a new substrate in ancient nuclear DNA studies. In addition, our data demonstrated that hair samples constitute an alternative source of nuclear DNA for individuals who failed to give results from bone samples. Furthermore, this approach could be added to the list of authentication criteria as a multi-substrate analysis. Indeed, the comparison of data from three different substrates (bones, teeth and hairs) could be considered as an indication for reliable results, as the probability of a simultaneous contamination of the different substrates and of the numerous extractions separated in time is very low.

Acknowledgments

This research was supported by a grant of the French Ministry of Foreign Affairs in the frame of the ACI "le complexe Altaï-Baïkal, Plaque tournante des flux géniques en Haute Asie, de la période protohistorique à l'époque moderne". The authors thank all the archaeologists and anthropologists for the quality of the excavations.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [10.1016/j.forsciint.2006.05.042](https://doi.org/10.1016/j.forsciint.2006.05.042).

References

- [1] M.T. Gilbert, R.C. Janaway, D.J. Tobin, A. Cooper, A.S. Wilson, Histological correlates of post mortem mitochondrial DNA damage in degraded hair, *Forensic Sci. Int.* 156 (2006) 201–207.
- [2] L.E. Baker, Mitochondrial DNA haplotype and sequence analysis of historic chocaw and menominee hair shaft samples, PhD Thesis, University of Tennessee, Knoxville, 2001.
- [3] M.T.P. Gilbert, A.S. Wilson, A. Michael Bunce, J. Hansen, E. Willerslev, B. Shapiro, T.F.G. Higham, M.P. Richards, T.C. O'Connell, D.J. Tobin, R.C. Janaway, A. Cooper, Ancient mitochondrial DNA from hairs, *Curr. Biol.* 14 (12) (2004) 463–464.
- [4] I. Barnes, P. Matheus, B. Shapiro, D. Jensen, A. Cooper, Dynamics of mammal population extinctions in eastern Beringia during the last glaciation, *Science* 295 (2002) 2267–2270.
- [5] M.T. Gilbert, L. Menez, R.C. Janaway, D.J. Tobin, A. Cooper, A.S. Wilson, Resistance of degraded hair shafts to contaminant DNA, *Forensic Sci. Int.* 156 (2–3) (2006) 208–212.
- [6] R. Higuchi, C.H. von Beroldingen, G.F. Sensabaugh, H.A. Erlich, DNA typing from single hairs, *Nature* 332 (1988) 543–546.
- [7] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Repple, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, *Int. J. Legal Med.* 108 (1995) 68–74.
- [8] B. Budowle, M.W. Allard, M.R. Wilson, Critique of interpretation of high levels of heteroplasmacy in the human mitochondrial DNA hypervariable region I from hair, *Forensic Sci. Int.* 126 (1) (2002) 30–33.
- [9] A. Brandstätter, W. Parson, Mitochondrial DNA heteroplasmacy or artifacts a matter of the amplification strategy? *Int. J. Legal Med.* 117 (3) (2003) 180–184.
- [10] T. Grzybowski, B.A. Malyarchuk, J. Czarny, D. Miscicka-Sliwka, R. Kotzbach, High levels of mitochondrial DNA heteroplasmacy in single hair roots: reanalysis and revision, *Electrophoresis* 24 (2003) 1159–1165.
- [11] A. Hellmann, U. Rohleder, H. Schmitter, M. Wittig, STR typing of human telogen hairs a new approach, *Int. J. Legal Med.* 114 (4–5) (2001) 269–273.
- [12] D. McNevin, L. Wilson-Wilde, J. Robertson, J. Kyd, C. Lennard, Short tandem repeat (STR) genotyping of keratinised hair. Part 2. An optimised genomic DNA extraction procedure reveals donor dependence of STR profiles, *Forensic Sci. Int.* 153 (2–3) (2005) 247–259.
- [13] L. Dérobert, Médecine Légale, Flammarion, Paris, 1980, pp. 871.
- [14] C.A. Linch, D.A. Whiting, M.M. Holland, Human hair histogenesis for the mitochondrial DNA forensic scientist, *J. Forensic Sci.* 46 (4) (2001) 844–853.
- [15] R. Uchihi, K. Tamaki, T. Kojima, T. Yamamoto, Y. Katsumata, Deoxyribonucleic acid (DNA) typing of human leukocyte antigen (HLA)-DQ α 1 from single hairs in Japanese, *J. Forensic Sci.* 37 (3) (1992) 853–859.
- [16] L. Eckhart, J. Bach, J. Ban, E. Tschächler, Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity, *Biochem. Biophys. Res. Commun.* 271 (2000) 726–730.
- [17] H.N. Poinar, M. Hofreiter, W.G. Spaulding, P.S. Martin, B.A. Stankiewicz, H. Bland, R.P. Evershed, G. Possnert, S. Pääbo, Molecular coproscopy:

- dung and diet of the extinct ground sloth *Nothotheriops shastensis*, *Science* 281 (1998) 402–406.
- [18] C. Keyser, B. Ludes, Methods for the study of ancient DNA, *Methods of Molecular Biology*, vol. 297, Humana Press, Totowa, 2004, pp. 253–264.
- [19] P. Gill, J. Whitaker, C. Flaxman, N. Brown, J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Sci. Int.* 112 (2000) 17–40.
- [20] A. Cooper, H. Poinar, Ancient DNA: do it right or not at all, *Science* 289 (2000) 1139.
- [21] P. Taberlet, S. Griffin, B. Goossens, S. Questiaux, V. Manceau, N. Escaravage, L.P. Waitis, J. Bouvet, Reliable genotyping of samples with very low DNA quantities using PCR, *Nucleic Acids Res.* 24 (16) (1996) 3189–3194.
- [22] A.D. Kloosterman, P. Kersbergen, Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci, *J. Soc. Biol.* 197 (4) (2003) 351–359.
- [23] B. Pakendorf, B. Morar, L.A. Tarskaia, M. Kayser, H. Soodyall, A. Rödewald, M. Stoneking, Y-chromosomal evidence for a strong reduction in male population size of Yakuts, *Hum. Genet.* 110 (2) (2002) 198–200.
- [24] T. Schultes, S. Hummel, B. Herrmann, Recognizing and overcoming inconsistencies in microsatellites typing of ancient DNA samples, *Ancient Biomol.* 1 (1997) 227–233.
- [25] P.M. Schneider, K. Bender, W.R. Mayr, W. Parson, B. Hostic, R. Decorte, J. Cordonnier, et al., STR analysis of artificially degraded DNA—results of a collaborative European exercise, *Forensic Sci. Int.* 139 (2–3) (2004) 123–134.

VII.4 Annexe 3

First successful assay of Y-SNP typing

by SNaPshot minisequencing on ancient DNA

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

Article accepté sous réserve de modifications par l'*International Journal of Legal Medicine*.

Abstract In the present study, a set of 13 Y-SNPs selected for the identification of the most frequent Asian Y-haplogroups were included in an allele specific primer extension assay. The SNP genotyping was accomplished by co-amplification of these 13 DNA fragments within two multiplex PCRs followed by detection with one minisequencing reaction using the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems) and analysis of extension products by capillary electrophoresis. First developed on modern samples, the assay was optimized for the analysis of 11 ancient DNA samples from the Krasnoyarsk region (Southern Siberia) that were dated from 5500 to 1800 YBP. The SNP typing was successful for most of them, which were all assigned to Y-haplogroup R1a1 except one. These results show that SNPs are well suited for the analysis of aged and degraded DNA samples. Moreover, we found that the SNaPshot minisequencing methodology is a convenient, robust and efficient method for SNP typing. To our knowledge, this study reports the first successful investigation of Y-SNPs on ancient DNA samples. The potential use of Y-SNPs in both evolutionary and forensic fields is also discussed.

Keywords Y-chromosome, SNPs, SNaPshot, single base extension, ancient DNA, Southern Siberia, R1a1 Y-haplogroup

³ Corresponding author

e-mail: caroline.bouakaze@ulp.u-strasbg.fr, Fax: +33-(0)390-243-348

Introduction

In the last years, single nucleotide polymorphisms (SNPs) located within the non-recombining part of the Y-chromosome (NRY) have been of particular interest for evolutionary and forensic purposes (Jobling and Tyler-Smith 2003; Jobling 2001). Due to their paternal inheritance, lack of recombination and very low mutation rate, their typing allows the

identification of stable male lineages over large periods of time. Moreover, these simple and frequent markers define a highly resolved tree of binary Y-chromosomal haplogroups composed of 18 major clades (A-R) that show a geographical distribution (YCC 2002; Jobling and Tyler-Smith 2003).

In evolutionary and anthropological studies, these markers have been commonly used to infer the origin, evolution and history of humans by tracing back male patterns of migration from modern human populations (Jobling and Tyler-Smith 2003). However, better insights into this issue should be provided by comparing these modern genetic data to those obtained by studying past human populations.

In the forensic field, Y-chromosomal markers have become recently a powerful tool for the screening of samples from rape cases (male and female DNA admixture) and for paternity testing where DNA from the alleged father is not available, since they are confined to males (Jobling et al. 1997). Y-chromosomal SNPs (Y-SNPs) may be of particular interest for difficult identification since they provide information on the geographical origin of an unknown or tracing sample, however not in France (Jobling 2001).

The overlap between both fields is that the biological material is often highly degraded and only available in low quantities. The analysis of such samples using conventional STR typing kit for multiplex amplifications often fails to resolve informative profiles. This is assumed to be due to the large size of amplicons that are generated (150-400 bp). Since the amplification of small amplicons was reported to increase the chance of successful typing of degraded DNA (Wiegand and Kleiber 2001) and Y-SNPs can be analyzed in amplicons shorter than 150 bp, they seem to be a tool of choice for the analysis of highly degraded samples.

Among the various SNP typing methodologies available today (reviewed in Budowle, 2004 and Sobrino et al. 2005), we selected the SNaPshot minisequencing based approach, which consists in the single base extension (SBE) of an unlabeled primer that anneals one base

upstream to the relevant SNP with a fluorochrome labeled dideoxynucleotide. The determination of the SNP allelic state is then possible by separating extended products and detecting the fluorescence by capillary electrophoresis. Currently, various multiplex minisequencing based assays were successfully validated for the analysis of mitochondrial DNA (Brandstätter et al. 2003, 2006; Grignani et al. 2006), autosomes (Dixon et al. 2005; Sanchez et al. 2006), the Y-chromosome (Brion et al. 2005a, 2005b; Onofri et al. 2006; Sanchez et al. 2003, 2005), Duffy and ABO group systems (Doi et al. 2004; Ferri et al. 2006) and the melanocortin 1 receptor gene as indicator of red hair phenotype (Grimes et al. 2001). All these studies were consistent in the fact that this SNP typing methodology is robust, reliable and extremely sensitive. Moreover, the only equipment needed to perform this SNaPhot assay is an automated DNA sequencer that is readily available in most laboratories performing DNA analyses.

The aim of the present study was to evaluate the effectiveness of a minisequencing based assay for the typing of Y-SNPs with low amounts of aged and degraded DNA recovered from human bone remains. For this purpose, we first selected Y-SNPs relevant for the past populations currently under study in our ancient DNA (aDNA) laboratory, namely Mongolian, Yakut, Siberian and Amerindian populations. Second, we developed on modern samples an efficient multiplex assay working with tiny amounts of DNA. Finally, we optimized and validated the whole strategy by analyzing a set of 11 aDNA samples.

Materials and Methods

Y-SNP selection

A set of 13 biallelic markers (M3, M9, M17, M45, M89, M173, M175, M216, M217, M242, 92R7, RPS4Y₇₁₁ and Tat) characterizing Asian and Amerindian haplogroups was chosen from the literature. The sequences of regions spanning the relevant SNP loci were retrieved either from the National Center for Biotechnology Information (NCBI) Single Nucleotide Polymorphism database (dbSNP) or

from Genbank using the nucleotide basic local alignment tool nBLAST (<http://www.ncbi.nih.gov/BLAST/>) with published PCR primers as query. The sequence variation for each SNP and the Y-haplogroup tree defined by them are represented in Fig. 1.

Development of the strategy: modern samples

Modern DNA samples

Modern DNA samples from 5 male individuals of different geographical origin (Europe, Yakutia, Mongolia) and one female individual, as negative control, were analyzed. Appropriate informed consent was obtained from these 6 subjects. Genomic DNA was extracted from blood or buccal swabs using a phenol-chloroform method, purified using the CleanMix Kit (Talent), concentrated using YM-30 Microcons (Millipore) and finally, quantified by real-time PCR using the QuantifilerTM Human DNA Quantification Kit (Applied Biosystems) with the ABI PRISM[®] 7000 Sequence Detection System (Applied Biosystems) following manufacturer's recommendations.

Design and validation of PCR primers

Despite the fact that primers for the amplification of all SNP loci were already described in the literature, most of them were redesigned in order to reduce the amplicon length using the following 4-step procedure: (i) DNA sequences spanning the SNP loci were first screened for interspersed repeats and low complexity regions using the RepeatMasker program (<http://www.repeatmasker.org>), (ii) primer sequences were designed using the Primer3 software (http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi), (iii) selected sequences were compared to those of the NCBI database using nBLAST to avoid spurious matches with other sequences of the human genome and (iv) were finally checked for primer dimer and secondary structures using the AutoDimer program (<http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm>). The size of the PCR products ranged from 81 to 155 bp to enable the amplification of short DNA fragments. PCR primer pair sequences are shown in Table 1.

Lyophilized and desalted PCR primers (Invitrogen) were dissolved at reception in 1X Tris-EDTA buffer, pH7, to a stock concentration of 100 µM and stored at -20°C.

Each primer pair was tested in a singleplex amplification reaction of 1 ng standard male template using conditions outlined by Sanchez and co-workers (Sanchez et al. 2004). The thermal cycling was carried out in a T3 Thermocycler (Biometra) and consisted of a first denaturation step at 94°C for 5 min followed by 33 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s and extension at 65°C for 30 s and a final extension at 65°C for 7 min. Singleplex amplification products were purified using Microcon PCR concentrators (Millipore) following manufacturer's recommendations. Sequencing of singleplex amplification products was done using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) according to the manufacturer's protocol in order (i) to check the specific amplification of the region surrounding the SNP locus, (ii) to verify that this one is free from other polymorphisms and finally, (iii) to confirm the correct polymorphic site. Sequencing products were purified by ethanol precipitation and separated by capillary electrophoresis on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) using the Performance Optimum Polymer 4 (POP-4™, Applied Biosystems). Data were analyzed using Sequencing Analysis v.3.7 software (Applied Biosystems) and sequences were manually verified and compared to expected sequences using ChromasPro software v.1.32 (Technelysium).

Multiplex PCR amplifications

The 13 target sequences were amplified within two multiplex reactions (Table 1), a 7-plex (Multiplex I) and a 6-plex (Multiplex II), in 50 µL final volume composed of 1X PCR Gold Buffer (Applied Biosystems), 8 mM MgCl₂, 400 µM of each dNTP, 2 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems) and 1 ng of male template DNA. Concentrations of PCR primers in each reaction mix are specified in Table 1. Thermal cycling conditions were the same as for singleplex amplifications.

Removal of excess primers and dNTPs

Three PCR clean-up methods were tested on pooled PCR products. (i) Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (ExoI) treatment: 15 µL pooled PCR products were treated with 5 U of SAP (Amersham Biosciences) and 2 U of ExoI (Amersham Biosciences) by incubation at 37°C for 1 h followed by enzyme inactivation by heating at 75°C for 15 min. (ii) UltraClean PCR Clean-up Kit (MoBio): 50 µl of pooled PCR products were purified following manufacturer's recommendations and

were finally eluted from filters in 50 µl of sterile water. (iii) Genopure dsTM (Bruker Daltonics): 30 µL pooled PCR products were purified following manufacturer's recommendations and were finally eluted from magnetic beads in 10 µL sterile water.

Design and validation of minisequencing primers

Each minisequencing primer was designed so that its 3' end annealed one base upstream to the relevant SNP locus using the same 4-step procedure than for PCR primer design. Some of them were 5' tailed with a non-homologous sequence (5'-AACTGACTAAACTAGGTGCCACGTCGTGA AGTCTGACAA-3', Lindblad-Toh et al. 2000) and poly-C if necessary to produce extension products ranging from 21 to 91 nt and deferring in length from each other by 6 nt to be enough separated by capillary electrophoresis. Minisequencing primer sequences are shown in Table 2.

Lyophilized and desalted primers (Invitrogen) were purchased either HPLC or PAGE-purified to remove incomplete primer synthesis products. At reception, primers were dissolved in 1X Tris-EDTA buffer at a stock concentration of 100 µM and were kept at -20°C.

To ensure the correct design and their spatial position, each minisequencing primer was first tested in a 10 µL final volume single reaction containing 4 µL SNaPshotTM Multiplex Ready Mix (Applied Biosystems), 3 µL cleaned pooled PCR products and 0.2 µM minisequencing primers. Extension was performed for 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s and extension at 60°C for 30 s.

Multiplex single base extension reaction (Minisequencing)

The multiplex minisequencing reaction was carried out in a 10 µL final volume containing 4 µL of SNaPshotTM Multiplex Ready Mix (Applied Biosystems), 20 mM ammonium sulfate, 3 µL cleaned pooled PCR products. Concentrations of primers in the reaction mix are specified in Table 2. Thermal cycling conditions were the same as for single SBE reactions.

Removal of unincorporated ddNTPs

Extension products were treated with 1 U of SAP (Amersham Biosciences) by incubation at 37°C for 1 h followed by enzyme inactivation by heating at 75°C for 15 min.

Capillary electrophoresis and data analysis

For preparation of samples to capillary electrophoresis, 2 µL purified extension products were mixed to 17.6 µL Hi-Di™ Formamide (Applied Biosystems) and 0.4 µL of internal size standard GeneScan-120 LIZ™ (Applied Biosystems). After denaturation, samples were run on an ABI PRISM 3100® Genetic Analyzer (Applied Biosystems) using POP-4® (Applied Biosystems). Data were analyzed using the GeneScan™ v.3.7 software (Applied Biosystems) and automated allele call was done by creating an optimized spectral matrix on the Genotyper™ v.3.7 software (Applied Biosystems) with peak threshold set to a minimum of 100 relative fluorescence units (RFUs). Haplogroup assignment was performed according to the YCC2003 tree (Jobling and Tyler-Smith 2003).

Male-Female DNA mixture assay

The specificity of amplifications and the influence of contaminant female DNA were checked by analyzing mixtures of a fixed amount of male DNA template (1 ng) with a dominant amount of female DNA (6, 12, 24, 48 and 190 ng). The experiment was done with two different male and female samples.

Sensitivity test

The performance of the assay was tested by analyzing serial dilutions of male template DNA at concentrations of 10 ng/ μ L, 1 000, 800, 600, 400, 200, 100, 50 and 1 pg/ μ L. The same experiment was performed with PCR thermal cycling conditions as previously detailed and with increasing the cycles from 33 to 37.

Analysis of aDNA samples

Ancient samples

Archaeological samples were composed of 11 bone fragments from unrelated male individuals. These samples originated from the Krasnoyarsk region, Southern Siberia. One of them was attributed to the Afanassievo culture (3500 to 2500 BC), 3 to the Andronovo culture (2500 to 1500 BC), 6 to the Tagar culture (800 to 200 BC) and the remaining one to the Tachtyk culture (200 BC to 200 AC). Removal of surface contamination and genomic DNA extraction were carefully done according to a published protocol (Keyser-Tracqui and Ludes 2005).

Y-SNP typing

The same protocol than for modern DNA was applied to aDNA samples with the two following modifications. (i) Multiplex amplifications were performed on 15 μ L aDNA extract for 37 thermal cycles instead of 33. PCR reaction clean-up was carried out using the Genopure dsTM kit (Bruker Daltonics). (ii) The multiplex SBE reaction was performed on 3 μ L purified PCR products and extension was performed for 30 cycles instead of 25.

Precautions against contaminations

To ensure the accuracy and reliability of results obtained from ancient samples, all amplifications were performed at least two times from two independent DNA extracts. Additional measures taken to avoid contamination with modern DNA were previously published (Keyser-Tracqui et al. 2003).

Results

Modern DNA samples

The 13 Y-SNPs were analyzed by simultaneous amplification into two multiplex PCRs followed by allelic discrimination within one multiplex minisequencing reaction using the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems) and finally by capillary electrophoresis detection.

As recommended in the SNaPshot manufacturer's protocol, we first used the ExoI and SAP treatment as PCR clean-up method. All modern samples investigated were successfully typed. Signals assigning the SNPs of interest were distinct, well spaced and thus accurately identifiable except for marker M217 because of high background noise between 25 and 32 nt (Fig. 2a). Assuming that these extraneous signals were due to an insufficient PCR clean-up, we tested two other purification methods. The use of the UltraClean PCR Clean-up Kit (MoBio), which is based on the high affinity of DNA to silica filters, did not allow the elimination of background noise (Fig. 2b). In addition, the overall intensity of signal was reduced from two to three folds with some peaks below 50 RFUs. This was certainly a consequence of the very high elution volume, 50 µL, that is needed and that cannot be decreased. We found that Genopure ds™ (Bruker Daltonics), which is based on the reversible immobilization of DNA to the surface of magnetic beads, was the most efficient purification method since signal assigned to marker M217 was clearly distinguishable (Fig. 2c).

Moreover, the overall signal intensity was considerably increased up to 6000 RFUs probably due to the low elution volume, 15 µL, which enabled sample concentration at the same time. The specificity of Y-SNP amplification was tested using a negative control that contained female DNA. No peaks were observed on the resulted electropherogram except at Tat locus, where a green signal was clearly visible. This suggested the existence of a homologous sequence with ancestral Tat allele. Previously, by sequencing Tat singleplex amplification products, we observed a weak amplification for a female sample. However, we did not take care of it since PCR primers were picked from the literature, where they were reported to be specific of the Y-chromosome. The sequence of the region surrounding Tat locus was aligned to those of the GenBank database using nBLAST on the NCBI website and we found that this SNP locus is located within the ubiquitin-specific protease 9 gene (USP9Y). Since this gene has a homolog on the X-chromosome, USP9X, we supposed that the X-chromosome linked homologous sequence was co-amplified during the PCR. Thus, sequences of PCR and minisequencing primers were also BLASTed against the Genbank database and, indeed, a part of each sequence matched to the X-chromosome. To determine whether this X-aspecific amplification did interfere with the Y-specific amplification, we genotyped a Yakut male individual. This individual bore the derived Tat allele but no amplification of the X-chromosome homologous sequence was observed.

In addition, male-female DNA mixtures were analyzed to test the influence of female DNA. The same results were obtained with male template DNA only and male-female mixture (1:200). This indicated that female DNA did not interfere with the amplification of Y-SNPs even for Tat marker.

Both multiplex PCRs were designed to produce amplification products with length varying from 81 to 155 bp to allow the analysis of highly degraded DNA. To test the sensitivity of the assay, experiments with decreasing amounts of DNA template were carried out with 33 and

37 amplification cycles. When using 33 PCR cycles, experiments showed that our strategy works very well with DNA amounts ranging from 1 ng to 400 pg. A lot of artifact pull-up peaks were observed with 10 ng DNA rendering the electropherogram interpretation very difficult. Below 400 pg template DNA, either only few or no peaks were distinguishable. By increasing amplification cycles from 33 to 37, the lower limit of detection was pushed back to 50 pg.

aDNA samples

To evaluate the effectiveness of this SNP typing assay with aged and degraded DNA, we tested it with minor modifications on 11 DNA samples recovered from ancient distinct human specimens found in Siberia. The typing was carried out at least twice for each sample and all replicates gave consistent results that always correlated with the YCC2003 tree (Jobling and Tyler-Smith 2003). Moreover, all extractions and PCR blanks were consistently negative throughout the study indicating that our results were unlikely to derive from contaminants in extraction and PCR processes. The typing always failed for two samples, the oldest sample belonging to the Afanasievo culture (3500 to 2500 BC) and one belonging to the Tagar culture (800 to 200 BC). We were able to determine the Y-haplogroup of the 9 remaining samples. Among them, 8 were assigned to Y-haplogroup R1a1 and one to Y-haplogroup C(xC3). An example of electropherogram obtained from an ancient sample belonging to Y-haplogroup R1a1 is shown in Fig. 3.

In nearly all cases, the M242 locus could not be amplified. This could be a consequence of the large size of the amplification product, namely 155 bp.

Moreover, we also observed an extraneous green peak sized approximately 82 nt, a few bases greater than the peak assigned to marker M9 (Fig. 3). We did not identify the causes of this extraneous peak but it did not affect the SNP interrogation process. Indeed, no peak was

expected to appear at this size and color since marker M9 is reported as a C to G transversion and the corresponding minisequencing primer is a reverse strand.

Although a single peak was expected for each SNP marker, we observed a double signal for the ancestral state of markers 92R7 and M17 (Comparison of Fig. 2c and Fig. 3). Since marker 92R7 was reported to be a paralogous sequence variant that originated from a duplication event, it was not surprising for this locus (Sanchez et al. 2004). Conversely, it was never reported for marker M17 although PCR and minisequencing primers were picked from the literature. This double signal showed a similar pattern than for marker 92R7, i.e. the ancestral allele 4G was always associated with a 3G variant and in the remaining samples only a 3G allele was observed. Causes are unknown but may be explained by the location of marker M17 in a duplicated region of the Y-chromosome.

As illustrated on Fig. 3, we observed a difference of peak height between the four dye lanes. According to the literature, this phenomenon is attributed to the ratios of the fluorophore emissions of dR110 (blue), dR6G (green), TAMRA (yellow) and dROX (red) dyes, which is approximately 4:2:1:1 (Sanchez et al. 2006).

Moreover, we also observed a difference of migration according to the size of extension products. The observed size of small extension products was always greater than the expected one, on average 3-4 nt more. However, no difference was observed between observed and expected sizes if above 50 nt. This phenomenon can be explained by the fact that migration depends on the secondary structure assumed by extension fragments in capillary electrophoresis and by the fact that the different masses of fluorochromes influence the mobility of shorter DNA molecules (Onofri et al. 2006).

Discussion

Technical point of view

The main challenge of this study was probably to multiplex a high number of primers in the amplification reaction so that the whole procedure was capable of working with minimal amounts of template DNA as available in forensic and aDNA laboratories. This assay was first conceived with the intention to amplify all amplicons in a unique 13-plex PCR and to detect the SNPs in a unique 13-plex minisequencing reaction. Although it successfully worked with modern DNA template, no peaks were observed on electropherograms obtained with aDNA. This may only be explained by the degraded nature of aDNA. That is why we needed to divide our 13-plex PCR into one 7-plex and one 6-plex amplification reactions. It should be kept in mind that an increase in multiplexing degree always implies a decrease in amplification efficiency.

It was also difficult to set up homogenous multiplex reactions. For this purpose, the concentrations of amplification and minisequencing primers were adjusted until we could obtain the same signal height for all extension products detected in the same dye lane. However, some products still presented either stronger, e.g. marker M3, or lower signal, e.g. marker RPS4Y₇₁₁, than others. Generally, we observed that the yield of products decreased with increasing PCR product size since it was generally the larger amplicon (marker M242) that was either weakly amplified or totally failed to amplify. This is also supposed to be due to the degraded nature of aDNA molecules because the amplification was always successful with the modern DNA samples.

Finally, considerable time was spent to find an efficient PCR cleaning method that enabled the complete removal of residual primers and excess dNTPs. As suggested by the ABI PRISM® SNaPshot™ Multiplex Kit's manufacturer and by many published studies, we first

used the straightforward and fast ExoI and SAP treatment but strong background signal between 25 and 32 nt did interfere with the SNP interrogation process for marker M217. Since this size range corresponded to that of PCR primer + 1 nt, we assumed that these background noise resulted from the extension of unremoved primers that formed dimers. Thus, the PCR clean-up step is of critical importance for correct typing using our strategy. As suggested by Onofri and colleagues, this phenomenon can be avoided by designing minisequencing primers with length above 30 nt, out of the range of PCR primers, so that result interpretation is unambiguous even in case of insufficient purification (Onofri et al. 2006). Nevertheless, the magnetic bead DNA purification system Genopure dsTM (Brucker Daltonics) was found to be very efficient.

SNaPshot minisequencing based genotyping of Y-SNPs with aged DNA samples

The aim of this study was to evaluate the effectiveness of a SNP typing based on SNaPshot minisequencing with low amounts of aged and degraded DNA samples.

This SNP typing assay was chosen because no investment in a new equipment is needed since it works with a thermalcycler and an automated DNA sequencer, both of which being readily available in laboratories performing DNA analyses. Moreover, it is a rapid and straightforward assay because of the kit format and the existence of numerous detailed technical published papers.

The typing of the 13 selected Y-chromosomal binary markers gave successful results for all aDNA samples investigated except two, which can be explained by the fact that the DNA content of these two bone samples was too scarce according to real-time PCR quantitation data. As concluded by many other studies, we found this methodology very robust because reproducible and reliable results were obtained from both modern and ancient samples that contained different quantities and qualities of DNA. The high rate of success with aDNA samples can be explained in part by the cold and dry environmental conditions where the

ancient human remains were collected, namely the Krasnoyarsk region, Southern Siberia, which may favor DNA preservation. Moreover, it is important to note that such results would probably never be obtained without increasing the number of amplification cycles from 33 to 37 and extension cycles from 25 to 30, which was done to improve the chance of amplifying the few molecules present in aDNA extracts.

To date, experiments to explore the suitability of a SNaPshot minisequencing based SNP typing with low molecular weight DNA were either done with artificially degraded DNA (e.g. sonicated DNA) or with forensic degraded samples with indication about neither the age of the sample nor the extent of DNA degradation. Since it can be predicted that, without repair mechanisms of living cells, DNA is spontaneously degraded in short DNA fragments over a time of period over thousands years (Lindahl 1993) and since the bone samples from which DNA was recovered in this study were approximately dated from 5500 to 1800 YBP, we assume these aDNA samples were degraded and fragmented. However, complete or almost complete autosomal- and Y-STR profiles were previously obtained from these aDNA samples suggesting that recovered DNA was not highly damaged but present in minute amount according to quantitation by real-time PCR. Thus, the present study reports for the first time the efficiency of a Y-SNP typing using the SNaPshot minisequencing methodology with low amounts of aged and degraded DNA.

Nevertheless, this SNP typing assay presents two disadvantages. First, it requires an increased amount of initial DNA since the 13 SNP loci are first amplified within two multiplex PCRs. Second, this strategy consists in “multi-step” procedures. This is usually not recommended in both anthropological and forensic fields because it may lead to loss of material and to an increase of sample contamination. Moreover, further optimization is still needed. First, it is necessary to further validate this strategy on a larger ancient DNA sample set. Second, it

would be interesting to include additional SNP markers that should provide a fine Y-haplogroup characterization.

Concluding remarks

The simple, robust and efficient SNP typing assay based on SNaPshot minisequencing should found an application in both evolutionary and forensic studies.

Today, evolutionary studies are conducted to infer the origin of humans and subsequent migration patterns and timing by extrapolating data obtained from the typing of Y-SNPs on modern human populations. Direct typing of these biallelic markers on ancient human populations could lead to a better understanding of human history by validating or invalidating these hypotheses. For example, the present work shows that Y-haplogroup R1a1 was widely distributed among the Krasnoyarsk region, Southern Siberia, since the Bronze and Iron Age. This result will be further interpreted with regards to data obtained from typing of other genetic markers, namely STRs and mitochondrial DNA, on the same aDNA samples (article in preparation).

In general, Y-SNPs may also be of special interest for forensic purposes. Autosomal SNPs have already proven their usefulness for identification of victims of mass disasters where conventional STR typing failed to resolve any informative profile (Biesecker et al. 2005). Today, forensic scientists have to face to an increase of modern disaster scenarios, i.e. genocides or mass disasters like airplane crashes, tsunamis or terrorist attacks, where human bodies are severely decomposed or fragmented and thus contain only low amounts of highly degraded DNA and where people from various geographical areas are involved. When used in supplement to other genetic markers, Y-SNPs may offer a promising tool for the rapid identification of geographical origin of victims from such large mass disasters.

Acknowledgements We gratefully thank Sarah Romac for her valuable technical help.

References

- Alessandrini F, Turchi C, Onofri V, Buscemi L, Pesaresi M, Tagliabracci A (2005) Multiplex PCR development of Y-chromosomal biallelic polymorphisms for forensic application. *J Forensic Sci* 50:519-525.
- Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B, Eisenberg A, Forman L, Kidd KK, Leclair B, Niezgoda S, Parsons TJ, Pugh E, Shaler R, Sherry ST, Sozer A, Walsh A. (2005) Epidemiology. DNA identifications after the 9/11 World Trade Center attack. *Science*. 310:1122-1123
- Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117:291-298
- Brandstätter A, Salas A, Niederstatter H, Gassner C, Carracedo A, Parson W (2006) Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27:2541-2550
- Brion M, Sobrino B, Blanco-Verea A, Lareu MV, Carracedo A. (2005a) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int J Legal Med* 119:10-15
- 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, Morling N (2005b) Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26:4411-4420
- Budowle B (2004) SNP typing strategies. *Forensic Sci Int* 146 Suppl:S139-142
- Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci Int* 154:62-77

- Doi Y, Yamamoto Y, Inagaki S, Shigeta Y, Miyaishi S, Ishizu H (2004) A new method for ABO genotyping using a multiplex single-base primer extension reaction and its application to forensic casework samples. *Leg Med (Tokyo)* 6:213-223
- Ferri G, Bini C, Ceccardi S, Ingravallo F, Lugaresi F, Pelotti S (2006) Minisequencing-based genotyping of Duffy and ABO blood groups for forensic purposes. *J Forensic Sci* 51:357-60
- Grignani P, Peloso G, Achilli A, Turchi C, Tagliabracci A, Alu M, Beduschi G, Ricci U, Giunti L, Robino C, Gino S, Previdere C (2006) Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. *Int J Legal Med* 120:151-15
- Grimes EA, Noake PJ, Dixon L, Urquhart A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 122:124-129
- Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118-124 Review
- Jobling MA. (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci Int* 118:158-162
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612 Review
- Keyser-Tracqui C, Crubezy E, 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: 247-260
- Keyser-Tracqui C, Ludes B (2005) Methods for the study of ancient DNA. *Methods Mol Biol* 297:253-264
- Lindahl T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362:709-715 Review
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24: 381-386
- Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int* 157:23-35

- Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, 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, Carracedo A (2004) Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications. *Forensic Sci Int* 140:241-250
- Sanchez JJ, Borsting C, Morling N (2005) Typing of Y chromosome SNPs with multiplex PCR methods. *Methods Mol Biol* 297:209-228
- 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, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713-1724
- Sobrino B, Brion M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181-194 Review
- Vallone PM, Butler JM (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J Forensic Sci* 49:723-732
- Wiegand P, Kleiber M (2001) Less is more--length reduction of STR amplicons using redesigned primers. *Int J Legal Med* 114:285-287
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339-348
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and Northern European

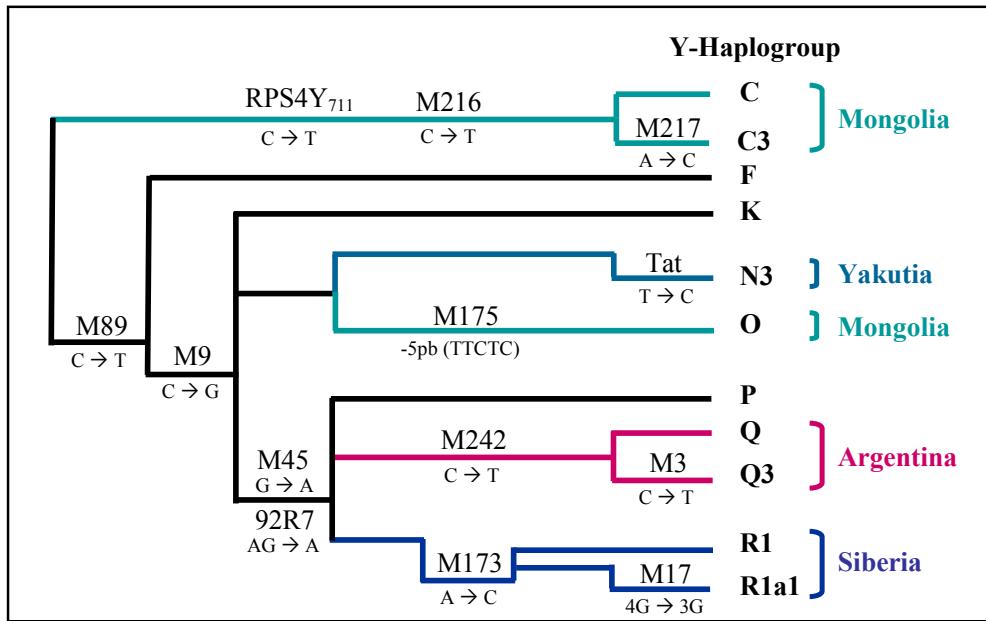


Fig. 1 Phylogenetic tree of the 13 Y-chromosomal binary polymorphisms analyzed in this study.

Marker names are indicated above the lines, sequence variation under the marker name and corresponding lineage names are shown at the end of each branch. The length of each one has no significance in term of phylogenetic distance. Colors indicate the geographical region where the haplogroup frequency is predominant (blue panel for Asian populations and red for Amerindian populations).

Table 1 PCR primers for the two multiplex PCR amplifications of the 13 Y-SNPs used in this study

Marker	PCR primer sequence (5'→3')		Amplicon size (bp)	Concentration (μM)
	Forward	Reverse		
<i>Multiplex I</i>				
M173	TTTCTTACAATTCAAGGGCATTTAG ²	CTGAAAACAAACACTGGCTTATCA ²	81	0.3
M175	AGTACCCAAATCAACTCAACTCCAGTG	CTGATACCTTGTTCAGTCTGTCATTCTGA	100	0.2
M217	GATTCTTAATTGTGAAGGAGAATGAA	CGTAAGCATTGATAAAAGCTGTGTG	115	0.3
M45	GGTGTGGACTTACGAACCAACCT ¹	TATCTCTGGCCTGGACCTCAGAAG	119	0.4
RPS4Y ₇₁	TGGTGGGATGTTGTTTCTCTCCT	CAACAGTAAGTCGAATGCCCTTCC	123	0.35
M17	ATTGGGAATACCTGGTCATAAC ³	AATAGTTGGCCACTTAACAAACCC ³	124	0.4
M216	TCCCAACCAGTTTATGAAGCTAGA	GCAAAGATAATTGTCAGGGTAAGC	145	0.2
<i>Multiplex II</i>				
M9	TCAGGACCCGAAATACAGAACTGC	TTGAACGTTGAACATGTCTAAATTAAAG	93	0.3
92R7	TTAAATCCCTCTATTGTGCTAACCA ²	CACTCTTCAGAAAAATGCATGAAC	107	0.35
M3	AATGTGCCAAGTTTATCTGCTG	GGCATTTCTATTTAGGTACAGCTC	111	0.15
TAT	GACTCTGAGTGTAGACTGTGA ⁴	GAAGGTGCCGTAAAGTGTGA ⁴	112	0.2
M89	TGGATTCACTCTCTTAAGGTTAT ²	CTGCTCAGGTACACACAGAGTATCA ²	135	0.2
M242	GCATAGAAAGTTGTGCAAAAGGTGAC	GGGCTTCAGCATAATACCTTACCTAGAA	155	0.25

study

Primers picked from the literature are labeled with a number that indicates the publication of reference.

¹ Alessandrini et al. 2005; ² Sanchez et al. 2005; ³ Onofri et al. 2006; ⁴ Zerjal et al. 1997.

Table 2 Minisequencing primers for the detection of the 13 Y-SNPs used in this study

Marker	Poly (dC)	Neutral sequence (5'→3')	Target specific sequence (5'→3')	(*)	Primer size (nt)	Concentration n (μM)
M3	-		- GGTACCAAGCTTCCTAATT ¹	R	20	0.08
M217	-		- TTATGTATTTCTCTGAAGAGTT ²	R	26	0.1
M242	-	GAAAGTCTGACAA	AAAAGGTGACCAAGGTGCT	F	32	0.2
M45	-	CGTCGTGAAAGTCTGACAA	CTCAGAAGGAGCTTTGC ²	R	38	0.58
M89	-	GCCACGTCGTGAAAGTCTGACAA	AACTCAGGCAAAGTGTGAGAGAT ³	R	44	0.2
RPS4Y ₇₁						
1	-	ACTAGGTGCCACGTCGTGAAAGTCTGACAA	GGCAATAAACCTTGGATTC	F	50	0.5
92R7	-	TAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATGAACACAAAAGACGTAGAAG ³	R	56	0.57
M17	-	TGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CCAAAATTCTACTAAAAAAACCC ²	R	60	0.18
M175	2	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CACATGCCCTCACTTCTC ²	F	62	0.2
TAT	2	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	GCTCTGAAATTTAAATTTAAACAC ²	R	68	0.13
M9	12	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CATGTCCTAAATTAAAGAAAAATAAGA	R	80	0.27
M216	21	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	TGCTAGTTATGTACCTGTTGAAT ²	R	86	0.2
M173	31	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA	CAATTCAAGGGCATTAGAAC ⁴	F	92	0.2

Primers picked from the literature are labeled with a number that indicates the publication of reference.

¹ Vallone and Butler 2004; ² Brion et al. 2005b; ³ Sanchez et al. 2005; ⁴ Alessandrini et al. 2005.

(*) Orientation of primer according to the YCC information (YCC, 2002)

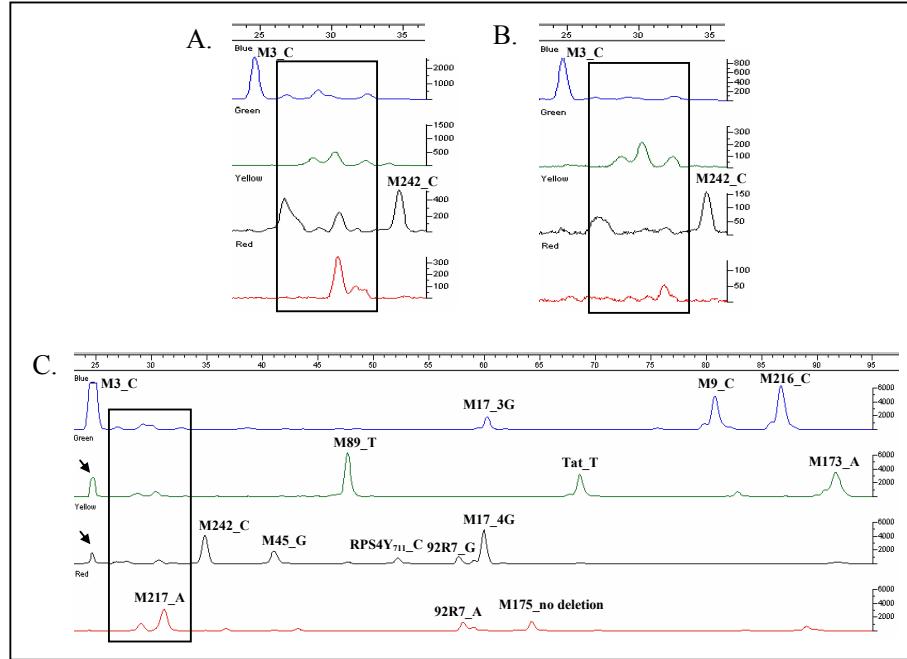


Fig. 2 Electropherograms obtained from the typing of 13 Y-SNPs with a modern European sample.

These plots, obtained using Genotyper v.3.7 software, show the RFUs versus measured size (nt) of SBE products relative to GS120 LIZ internal size standard. Typing of 13 Y-SNPs was achieved on 1 ng DNA template by two multiplex amplifications followed by one minisequencing reaction. Above each signal is indicated the name of the corresponding marker followed by the allele detected. For marker M175, the ancestral allele is detected by a positive amplification whereas no amplification is yielded with the derived allele, which is a 5 pb deletion. Various PCR products purification methods were tested in order to eliminate the background signals between 25 and 32 nt (black rectangle): (A) treatment with SAP and ExoI (Amersham Biosciences), (B) UltraClean PCR Clean-up Kit (MoBio), and (C) Genopure dsTM (Brucker Daltonics). Arrows indicate artifact pull-up peaks.

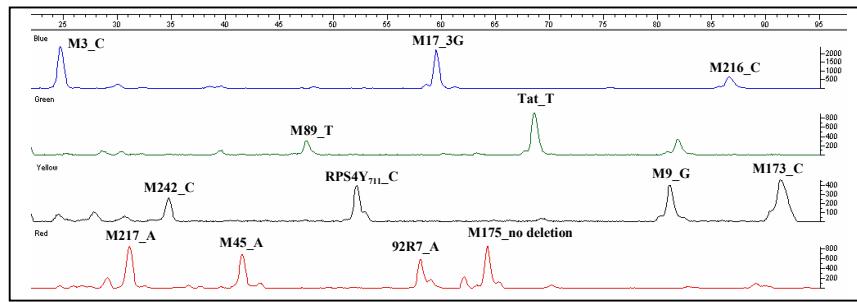


Fig. 3 Electropherogram obtained from an ancient male sample belonging to R1a1 Y-haplogroup.

This plot, obtained using Genotyper v.3.7 software, show the RFUs versus measured size (nt) of SBE products relative to GS120 LIZ internal size standard. Typing of 13 Y-SNPs was achieved by two multiplex amplifications followed by one minisequencing reaction. Above each signal is indicated the name of the corresponding marker followed by the allele detected. This electropherogram was obtained with approximately 450 pg aDNA.

VII.5 Annexe 5

A Rape Case Solved by Mitochondrial DNA Mixture Analysis

Didier Hatsch,^{1,2} Ph.D. ; Sylvain Amory,¹ ; Christine Keyser,¹ Ph.D. ; Rémi Hienne,² Ph.D.
; and Ludes Bertrand,¹ M.D., Ph.D.

¹ Institut de Médecine Légale, EA 3428, 11 rue Humann, 67085 Strasbourg Cedex, France.

² Laboratoire CODGENE, 11 rue Humann, 67085 Strasbourg Cedex, France.

ABSTRACT:

In cases of stains which contain mixed DNA from different contributors, analysing mtDNA requires the use of cloning techniques. We developed an efficient cloning technique which was applied in a rape case. After a differential lysis based DNA extraction from vaginal swabs, hypervariable region I and II (HVI, HVII) amplicons obtained from the male fraction were cloned. Although we mainly found the victim's haplotype, we were able to detect the suspect's haplotype in two clones for HVI and in one clone for HVII. As the mid piece of the flagellum, which contains mitochondria, can be lost during the differential lysis, we also investigated the female fraction by cloning, in order to evaluate the proportion of victim/suspect mtDNA. Unfortunately, only clones presenting the victim's haplotype were found. This case highlights the need for an optimal differential lysis protocol in order to enrich the male fraction with not only nuclear but also mitochondrial DNA

KEYWORDS: forensic science, cloning, mtDNA, rape

When faced with mixed DNA samples, forensic scientists commonly use STR analyses to distinguish the different contributors. However, such a study cannot be performed completely unless biological material compatible with STR investigations is available. For instance, when only shed hairs from suspected contributors are available, the resolution of a case relies mainly on mitochondrial DNA (mtDNA) analysis.

Mitochondrial DNA displays two important advantages over nuclear DNA. Firstly, mtDNA is present in thousands of copies in one cell, thus analyses based on it are highly sensitive even when working with degraded samples or minute amounts of biological material. Secondly, it is maternally inherited, which enlarges comparison possibilities to distant maternal relatives. It is therefore widely used for forensic investigations (1).

Direct sequencing of amplicons of the hypervariable regions I and II (HVI, HVII) is generally performed to retrieve the haplotype of a sample. Confronted with DNA mixture, such a procedure results in a haplotype combination of the different contributors. Thus amplicons have to be cloned in order to distinguish the different haplotypes.

In this report, we present a rape case where DNA mixture analysis was performed by means of an in-house cloning technique applied to mtDNA fragments.

Case Report

A 39-year old woman was raped at knife-point. As vaginal swabs were found to contain spermatozoa, DNA was extracted following a differential lysis protocol leading to two DNA fractions corresponding to the separated male and female cells (2). STR analysis, performed on the female fraction, yielded the victim's profile whereas those made on the male fraction resulted in a mix between the victim and the offender with an apparent mixture ratio of 1:1. Some weeks after these analyses, a suspect was found but the local police only supplied hair samples. Unfortunately, the transport of the hairs lasted several weeks and nuclear DNA extraction failed due to the long and inappropriate storage. In the meantime the suspect had fled in another country giving no possibility for obtaining easily new samples. As a consequence, for the comparison of the suspect's sample, the DNA mixture from the vaginal swab had to be investigated by means of mtDNA analysis.

Materials and Methods

Standard Molecular Techniques

Standard molecular biology procedures were performed (3). Unless otherwise stated, all enzymes used in this study came from MBI fermentas (Lithuania).

Vector Preparation

Vector preparation was performed by digesting 10 μ g pUC 19 with 100U HincII in 100 μ L reaction volume at 37°C for 2 hours. Digested plasmid was further purified using a Geneclean® Spin kit from Bio101 (Qbiogen, France). One base tailing was achieved using 3 μ g of digested vector, 2U of Terminal Desoxynucleotide Transferase and 500pmol of ddTTP in a final volume of 20 μ L at 37°C for 1 hour. The tailing product was purified by means of a Geneclean® Spin kit. Digestion and tailing efficiency was assayed by overnight self-ligation of 30ng vector at room temperature.

Cloning of mtDNA Fragments

mtDNA fragments were amplified as previously described (4) in 30 cycles and by adding to the PCR a final extension step of 30min, ensuring complete A-tailing of all PCR products. 100ng of PCR product were directly ligated in presence of 30ng of pUC19-tailed vector with 1U of T4-DNA ligase in 20 μ L final volume at 20°C overnight. The ligation products were treated 10min at 70°C and 1 μ L of each was transformed into *Escherichia coli* TOP10 (F-

mcrA $\Delta(mrr-hsdRMS-mcrBC)$ $\Phi 80lacZ\Delta M15$ $\Delta lacX74$ *recA1* *araD139* $\Delta(ara\ leu)$ 7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG*). After transformation, bacteria were plated on LB-agar plates containing Ampicillin and X-Gal. After overnight incubation at 37°C, white/blue screening allowed the selection of clones.

Plasmid Preparation

Individual clones were picked and cultivated in duplicated 2mL 96-well blocks for 24 hours at 37°C under 220rpm shaking. Bacteria were spun down by a 5min centrifugation at 1500g. The pellets obtained from duplicated plates were resuspended in 150 μ L of S1 (50mM glucose, 25mM Tris pH8, 10mM EDTA and 20 μ gmL⁻¹ DNase free RNase) and pulled together. Five hundred microliters of S2 (0,2M NaOH and 1% SDS) and 350 μ L S3 (3M potassium acetate, 10% acetic acid) were added with vortexing following each pipetting. This lysate was centrifuged for 10min at 3500g at 4°C to precipitate cell debris. The supernatant was transferred into a new 96-well block and 1mL of isopropanol was added to each well before centrifugation at 2500g for 15min. The precipitate was resuspended in 100 μ L of water and mixed with 100 μ L of 5M LiCl and further incubated at -20°C for 15min before centrifugation at 2500g for 15min at 4°C. The supernatant was recovered and plasmid DNA was precipitated by a 15min centrifugation at 2500g after addition of 500 μ L of ethanol. A final washing step in 70% ethanol was applied before resuspending the DNA in 40 μ L of water.

Plasmid Verification and Quantitation

Plasmids were diluted twenty times and 5 μ L of the dilution was used as PCR template under the following conditions: 500nM of each primer (PU+ 5'-GTG AAT TCG AGC TCG GTA CC-3' and RP+ 5'-CAG CTA TGA CCA TGA TTA CG-3'), 2mM MgCl₂, 200 μ M dNTP and 1 unit of *Taq* polymerase in 25 μ L final volume. After a first 5min denaturizing at 96°C, 15 cycles of 10s at 96°C, 30s at 60°C and 50s at 72°C and a final extension step of 2min at 72°C were carried out. A plasmid containing a 400bp HVI mtDNA fragment was used at 5, 10, 15, 20, 25, 35 and 50ng per reaction as a positive control and for quantitation of prepared plasmids. The subsequent PCR products were visualized on a 2% agarose gel.

Plasmid Sequencing

Sequencing was performed with 4 μ L of BigDye® terminator kit v1.1 mix (Applied Biosystems, CA), 1,6pmol PU (5'-GTT TTC CCA GTC ACG ACG TTG-3') or RP+ primer

and 5 μ L of plasmid in a final 10 μ L volume. Sequencing reactions were precipitated with 40 μ L 80% ethanol and further desalted by a 70% ethanol washing step. The purified sequencing products were separated by capillary electrophoresis and detected on an ABI Prism® 3100 genetic analyzer (Applied Biosystems, CA).

Restriction Analysis of Amplified Insert

The PCR reaction resulting from the insert check step was submitted again to 10 PCR cycles. Ten microlitters of the PCR product were digested by 20U *BccI* (New England Biolabs, MA) in a final volume of 20 μ L at 37°C for one hour.

Results and Discussion

Amplification of the HVI and HVII regions was performed for the suspect and victim reference samples and sequenced to determine their respective haplotypes (Table 1). Amplification of the same regions from the male fraction extracted from the vaginal swab and further cloning were performed in different PCR dedicated laboratory rooms to avoid contamination. As all negative controls were clean, the HVI and HVII fragments were cloned using our homemade T/A cloning vector. Vector preparation yields enough material for 50 cloning experiments. The prepared vector can be stored up to six months at -20°C as a ready-to-use stock. The cloning presented a very high efficiency, since almost 99% of the clones contained an insert. Plasmid preparation was performed in 96-well blocks following the adapted alkaline-lysis protocol described in material and methods. The quality, quantity and insert size of the plasmids were checked (Fig. 1). Only plasmids presenting amplification with intensities corresponding to quantities higher than 30 ng μ L⁻¹ were further sequenced. This procedure allows not only the quantitation of the plasmid DNA, but also quality assay by detection of DNA preparation containing PCR-inhibitors which could make the sequencing fail. Starting from the clone cultures, the whole preparation, quantitation and quality assay of the plasmid DNA of 192 to 384 clones can be achieved within a working day, which is very convenient knowing that cloning methods are generally labor-intensive and low-throughput. A sequencing strategy was applied in order to discover all the different mtDNA molecules present in the mixture rather than predigest the DNA mixture searching after the suspect's haplotype. After the sequencing of 48 HVI clones, only the victim's haplotype was recovered.

On the contrary, after sequencing of 32 HVII clones, the haplotypes of the suspect and the victim were found. Therefore, the mtDNA mixture ratio seemed to be lower than 1:10.

Thus, the mixture ratio in the HVI region amplification product was investigated. We searched for differential restriction sites in the victim's and suspect's HVI sequences. The restriction enzyme *BccI* gives a differential restriction product of the two different amplicons (Table 2) because the restriction site includes the position 16294 which is mutated (T) in the victim but not in the suspect. Thus the HVI amplification product from the male fraction was digested with *BccI*. The band of 330bp corresponding to a part of the suspect HVI sequence was barely visible on agarose gel, but was excised and cloned. Due to low insert concentration, only 8 clones were obtained. The plasmids were prepared and sequenced but only haplotypes presenting mutations T16126C, C16294A or G and C16296T were found. These molecules derive from the victim's mtDNA by Taq errors and were selected because of the lack of thymine at position 16294. In a first approach these results could point on a high Taq error rate because the same site seems to be hit several times by Taq point mutations. However this is not the case, but a result of the RFLP procedure which specifically selects molecules lacking the C16294T, putting this background noise to the foreground. Finding only haplotypes derived from the victim by Taq caused point mutations leads us to conclude that the mtDNA mixture ratio is extremely low.

Therefore to avoid sequencing hundreds of complete HVI clones, we selected them by means of restriction analysis. The bias of this approach is that the suspect was targeted for the match, whereas additional variants, from alternative suspects for example, will not be detected. The plasmids of 192 clones containing the complete HVI region were prepared and the PCR amplification used to check the quality and quantity of the preparation was submitted to 10 additional PCR cycles and further digested with *BccI*. Out of the 192 clones, 4 presented a restriction profile corresponding to the lack of the C16294T mutation as illustrated in figure 1. Sequencing these candidates revealed two mutated haplotypes as previously found belonging to the background noise of the procedure and two haplotypes corresponding to the suspect. This points to a mtDNA mixture ratio of roughly 1:100.

The mtDNA profiles found in the mixture correspond to the victim's and suspect's haplotypes. The latter is not found in the downloadable FBI forensic mtDNA population database (<http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm>) when considering both HVI and HVII regions in all populations (2008 records), as the French police never provide the ethnical origin of the suspect. A frequency estimator of $4,5 \cdot 10^{-4}$ was calculated, taking into account database size and distribution bias (5). These biological evidences allowed

the prosecution to require further biological samples from the police of the country in which the suspect fled.

It has been pointed out that the spermatozoa loose their flagellum and mid-piece, containing the mitochondria, during the first differential lysis step and therefore the male fraction was found to be exempt of sperm mtDNA (6). As these findings may explain our results, we decided to investigate the female fraction resulting from the differential lysis. The HVI region was amplified using the female fraction as a template and further cloned. The plasmid DNA from 96 clones was prepared and the PCR check amplification of the insert was digested with *BccI* in order to detect clones containing the suspect's sequence. Six clones, presenting a digestion profile indicating the absence of C16294T mutation, were sequenced but were identified as haplotype deriving from the victim's haplotype by a punctual mutation at position 16294 caused by the *Taq*. Thus we did not find higher male mtDNA content in the female fraction. Such results could be explained by the difference in the mtDNA copy number between the female epithelial cells and the spermatozoa. Indeed, 200 to 1700 mtDNA copies can be found per somatic cell (1). In opposite, the spermatozoa, which have only 70 to 80 mitochondria with only one mtDNA copy per mitochondrion, present only roughly 100 mtDNA copies per cell (7). Therefore a ratio of 1:10 can be expected with an equal amount of each cell type, however in forensic samples, dominance of the female cells lowers significantly this ratio.

Though the spermatozoa tails were supposed to be retained in the female fraction of the differential lysis extraction, in this case we observed male mtDNA in the male fraction only. The development of new protocols allowing a conservative separation of the two cell types would improve mtDNA mixture analysis from rape cases.

mtDNA was used as an investigational tool pointing out a preliminarily match with the suspect. Indeed, the obtained mtDNA haplotype match presented as an evidence for obtaining a sample adapted for STR typing. Fortunately, we were able to complete the match with STR profiling. Therefore, this investigation turned out to be a powerful application of the cloning approach.

Acknowledgements

Didier Hatsch was funded by a PostDoc fellowship from the Région Alsace. We thank Professor Jean-Marc Jeltsch and his team for great help in the cloning experiments. We warmly thank Elizabet Petkovski for constructive discussions. We thank Dr Lyndon Higgs for language revision of the manuscript.

References

1. Holland MM, Parson TJ. Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Sci Rev* 1999;11:21-50.
2. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints'. *Nature* 1985;318:577-9.
3. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, editors. *Current protocols in molecular biology*. New York: Wiley, 2002.
4. Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. Mitochondrial DNA sequences in single hairs from a southern African population. *PNAS* 1989;86:9350-4.
5. Buckleton J, Walsh S, Harbison S. Nonautosomal forensic markers. In Buckleton J, Triggs CM, Walsh SJ, editors. *Forensic DNA evidence interpretation*. CRC Press, 2005.
6. Anslinger KD, Bayer BR, Rolf B, Eisenmenger W. mtDNA investigations after differential lysis. *J Forensic Sci* 2005;50:579-81.
7. Wei YH, Kao SH. Mitochondrial DNA mutation and depletion are associated with decline of fertility and motility of human sperm. *Zool stud* 2000;39:1-12.

TABLE 1 - *Reference haplotypes for the suspect and the victim.*

HVII												
HVII												
		1	1	1	1					3	3	
		6	6	6	6			0	0	2	0	1
Start		1	2	2	3	Stop	Start	7	9	6	9	5
		2	9	9	0			3	3	3	.	.
		6	4	6	4					1	1	
Suspect	16025			C	16380	55		G	G	C	C	380
Victim	16025	C	T	T	16380	55	G		G	C	C	380

TABLE 2 - Digestion products of the HVI region amplification for the victim and the suspect.

	Size in bp
Suspect	330, 76
Victim	223, 107, 76

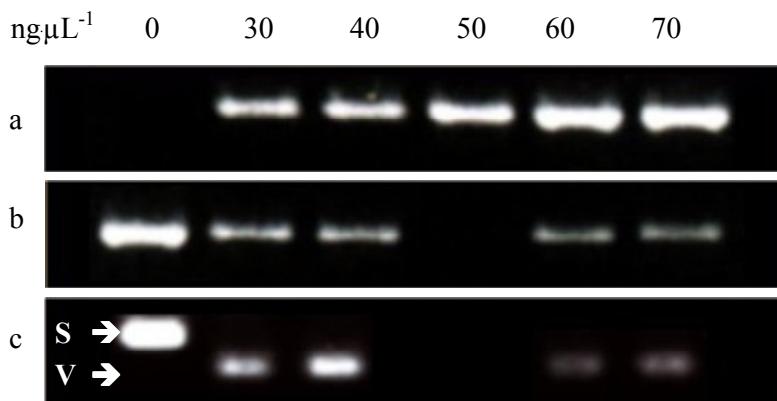


FIG. 1

FIG. 1 - Plasmid insert check by semi-quantitative PCR. **a** Standard of known concentration. **b** Amplification of the plasmid insert for six different clones. **c** Digestion of the amplification of the plasmid insert to detect clones containing a haplotype corresponding to the suspect (S, 330bp) and the one of the victim (V, 223bp). Gel electrophoresis was performed in a TBE-4% agarose gel.

VII.6 Annexe 6

Liste des populations incluses dans la base de données du logiciel YCDMA pour les STR du chromosome Y.

Population	Effectif	Référence
Berlin	314	YHRD
Bourriates	215	Wozniak <i>et al.</i> , 2006
Chine	55	YHRD
EG anciens	28	Keyser-Tracqui <i>et al.</i> , 2003
EG anciens moderne	69	Keyser-Tracqui <i>et al.</i> , 2006
Estonie	132	YHRD
Evenks	25	Xue <i>et al.</i> , 2005 (données non publiées)
Evenks	16	Karafet <i>et al.</i> (données non publiées)
Kazakhs	49	Perez-Lezaun <i>et al.</i> , 1999
Kirghiz-Highlands	43	Perez-Lezaun <i>et al.</i> , 1999
Kirghiz-Lowlands	41	Perez-Lezaun <i>et al.</i> , 1999
Latvia	145	YHRD
Leipzig	440	YHRD
Lithuanie	151	YHRD
Ljubljana	120	YHRD
Mongols	96	Zhu <i>et al.</i> , 2006
Mongols	37	Keyser-Tracqui <i>et al.</i> ,
Mongols	40	YHRD
Novgorod	50	YHRD
Ouigours	107	Zhu <i>et al.</i> , 2006
Ouigours	39	Perez-Lezaun <i>et al.</i> , 1999
Paris	104	YHRD
Russes	84	YHRD
Turcs	158	YHRD
Turcs Centre Anatolia	113	Cakir <i>et al.</i> , 2004
Turcs _Ankara	39	Nasidze <i>et al.</i> , 2004
Turcs-Bulgarie	60	YHRD
Ukraine	78	YHRD
Yakoutes actuels	25	Keyser-Tracqui <i>et al.</i> , 2006

VII.7 Annexe 7

Liste des populations incluses dans la base de données mitochondrielles de l’Institut de Médecine Légale de Strasbourg (<https://iml.u-strasbg.fr/>).

Zone Géographique	Population	Effectif	Séquence	Référence
Asie Centrale	Altai	16	16024-16383	Shields <i>et al.</i> , 1993
	Bourriates	40	16024-16383	Derenko <i>et al.</i> , 2000
	Bourriates	126	16024-16383	Pakendorf <i>et al.</i> , 2003
	Bourriates	130	16024-16383	Shimada 2001 non publié
	EgynGol anciens	39	16020-16380	Keyser <i>et al.</i> , 2003
	EgynGol actuels	124	16020-16380	Keyser <i>et al.</i> , 2006
	Kazakhs	55	16024-16383	Comas <i>et al.</i> , 1998
	Kazakhs	30	16024-16383	Yao <i>et al.</i> , 2002
	Kirghiz Highland	46	16024-16383	Comas <i>et al.</i> , 1998
	Kirghiz Lowland	48	16024-16383	Comas <i>et al.</i> , 1998
	Mongols	104	16020-16380	Keyser <i>et al.</i> , 2006
	Mongols	103	16020-16380	Kolman <i>et al.</i> , 1996
	Mongols	48	16020-16380	Kong <i>et al.</i> , 2003
	Mongols	15	16024-16383	Yao <i>et al.</i> , 2002
	Oroquen	44	16024-16383	Kong <i>et al.</i> , 2003
	Uzbeks	20	16024-16383	Comas <i>et al.</i> , 2004
Sibérie	Daurs	45	16020-16380	Kong <i>et al.</i> , 2003
	Evenks	47	16020-16380	Kong <i>et al.</i> , 2003
	Evenks	11	16024-16383	Torroni <i>et al.</i> , 1993b
	Evenks	29	16024-16383	Lebedeva 2000 non publié
	Sibérie Occidentale	38	16024-16383	Voevoda 1994 non publié
	Tuvas	36	16024-16383	Derenko <i>et al.</i> , 2000
	Udegey	3	16024-16383	Torroni <i>et al.</i> , 1993b
	Ouigours	45	16024-16383	Yao <i>et al.</i> , 2002
	Uighurs	55	16024-16383	Comas <i>et al.</i> , 1998
	Yakoutes	117	16024-16383	Pakendorf <i>et al.</i> , 2003
	Yakoutes	44	16020-16380	Keyser <i>et al.</i> , 2006
Béringie	Cercle Polaire	26	16024-16383	Shields <i>et al.</i> , 1993
	Tchoukches	65	16024-16383	Voevoda 1994 non publié
	Tchoukches	37	16024-16383	Starikovskaya <i>et al.</i> , 1998
	Esquimaux	23	16024-16383	Voevoda 1994 non publié
	Esquimaux (Sibérie)	77	16024-16383	Starikovskaya <i>et al.</i> , 1998
	Itelmens	45	16024-16383	Schurr <i>et al.</i> , 1999
	Koryaks	139	16024-16383	Schurr <i>et al.</i> , 1999
	Yupiks	52	16024-16361	Simonson 1997 non publié
Europe/Moyen-Orient	Karelle	83	16024-16383	Sajantila <i>et al.</i> , 1995
	Mansi	98	16024-16383	Derbeneva <i>et al.</i> , 2002
	Saamis	115	16024-16383	Sajantila <i>et al.</i> , 1995
	Turcs	29	16020-16380	Calafell <i>et al.</i> , 1996
	Turcs	44	16020-16380	Comas <i>et al.</i> , 2000
	Turcs	72	16020-16380	Di Benedetto <i>et al.</i> , 2001
Asie de l'est	Coréens	48	16020-16380	Kong <i>et al.</i> , 2003
Amérique du nord	Bella Coola	40	16024-16383	Ward <i>et al.</i> , 1993
	Haïda	41	16024-16383	Ward <i>et al.</i> , 1993
	Inuits	46	16024-16361	Simonson 1999 non publié
	Inupiak	10	16024-16361	Simonson 1997 non publié
	Nuu Chah nulth	63	16024-16383	Ward <i>et al.</i> , 1991
	Tlingit	6	16024-16361	Simonson 1997 non publié
	Yupiks	25	16024-16361	Simonson 1997 non publié

VII.8 Annexe 8

Bilan des Résultats de quantification par PCR en temps réel pour les échantillons osseux.

n° Laboratoire	Ct	Quantité
YAKa15	31,58	1,23E-01
IPC	28,9	
YAKa16-E1	33,56	5,44E-02
IPC	29,53	
YAKa16-E2	31,11	1,70E-01
IPC	28,75	
YAKa17-E1	32,83	2,05E-02
IPC	28,38	
YAKa17-E2	33,22	2,92x10-2
IPC	28,48	
YAKa18-E1	30,03	2,40E-01
IPC	28,62	
YAKa18-E2	28,22	1,23
IPC	29,11	
YAKa19-E1	26,38	2,44
IPC	28,85	
YAKa19-E2	26,49	3,05
IPC	28,98	
YAKa20-E1	28,65	8,00E-01
IPC	28	
YAKa20-E2	27,93	1,43E+00
	29,32	
YAKa21-E1	29,7	2,60E-01
IPC	28,65	
YAKa21-E2	28,62	9,60E-01
	29,26	
YAKa22-E1	36,34	7,02E-03
IPC	29,18	
YAKa22-E2	31,44	1,02E-01
IPC	28,8	
YAKa23-E1	ND	ND
IPC	29,49	
YAKa23-E2	34,57	1,09E-02
IPC	29,1	
YAKa24-E1	34,36	3,02E-02
IPC	29,06	
YAKa24-E2	35,49	5,65E-03
IPC	29,22	
YAKa25	32,89	8,92E-02
IPC	29,08	
YAKa26-E1	31,7	1,13E-01
IPC	28,77	

n° Laboratoire	Ct	Quantité
YAKa33	28,28	0,765
IPC	33,28	
YAKa34-E1	30,93	2,50E-01
IPC	29,79	
YAKa34-E2	33,86	3,62E-02
IPC	29,69	
YAKa35-E1	ND	ND
IPC	29,25	
YAKa35-E1	37,03	3,76E-03
IPC	29,44	
YAKa36-E1	28,8	1,18
IPC	37,78	
YAKa36-E2	30,53	3,92E-01
IPC	ND	
YAKa36-PTB	31,12	6,35E-02
IPC	ND	
YAKa37-E1	28,32	1,66
IPC	ND	
YAKa37-E2	29,85	0,635
IPC	ND	
YAKa37-PTB	29,14	2,70E-01
IPC	ND	
YAKa38-E1	34,5	1,87E-02
IPC	30,84	
YAKa38-E2	34,08	3,11E-02
IPC	31,31	
YAKa38-PTB	32,66	2,08E-02
IPC	29,24	
YAKa39-E1	31,84	0,054
IPC	24,63	
YAKa39-E2	ND	ND
IPC	36,04	
YAKa40	30,13	0,225
IPC	24,88	
YAKa41-E1	34,64	4,60E-02
IPC	27,26	
YAKa41-E2	ND	ND
IPC	26,86	
YAKa41	ND	ND
IPC	25,08	
YAKa43-E1	ND	ND
IPC	27,05	

n° Laboratoire	Ct	Quantité
YAKa45-E2	34,68	4,49E-02
IPC	27,2	
YAKa45-E3	ND	ND
IPC	25,07	
YAKa46-E1	30,76	6,47E-01
IPC	27,18	
YAKa46-E1	37,2	8,06E-03
IPC	27,05	
YAKa46-E2	32,94	1,46E-01
IPC	27,02	
YAKa46-E3	ND	ND
IPC	25,09	
YAKa47-E1	34,08	6,75E-02
IPC	27,25	
YAKa47-E2	34,57	4,82E-02
IPC	27,25	
YAKa47-E3	32,25	3,90E-02
IPC	27,82	
YAKa47-PTB	33,9	1,35E-02
IPC	25,94	
YAKa48-E1	ND	ND
IPC	26,99	
YAKa48-E2	ND	ND
IPC	25,04	
YAKa49-E1	33,02	1,39E-01
IPC	32,78	
YAKa50-E1	33,21	1,22E-01
IPC	26,98	
YAKa50-E2	31,92	4,97E-02
IPC	27,06	
YAKa50-E3	34,21	3,71E-03
IPC	25,12	
YAKa51-E1	32,09	5,54E-02
IPC	28,5	
YAKa51-E2	37,76	1,325E-03
IPC	29,47	
YAKa52-E1	36	3,20E-03
IPC	27,74	
YAKa52-E2	37,16	2,070E-03
IPC	29,87	

n° Laboratoire	Ct	Quantité
YAKa55-E3	ND	ND
IPC	25,23	
YAKa55-PTB	36,81	1,68E-03
IPC	25,98	
YAKa56-E1	26,98	2,32
IPC	27,48	
YAKa56-E2	26,18	2,085
IPC	27,59	
YAKa56-E3	26,75	2,19E+00
IPC	27,75	
YAKa57-E1	34,33	5,68E-02
IPC	27,04	
YAKa57-E2	31,14	1,11E-01
IPC	26,83	
YAKa57-E3	ND	ND
IPC	25,28	
YAKa58	35,98	2,56E-03
IPC	26,96	
YAKa58	34,96	2,07E-03
IPC	25,11	
YAKa58-E3	ND	ND
IPC	27,07	
YAKa58-PTB	32,2	4,10E-02
IPC	26,15	
YAKa59-E1	35,89	2,72E-03
IPC	26,98	
YAKa59-E2	34,15	1,23E-02
IPC	27,01	
YAKa59-E3	36,4	7,10E-04
IPC	25,1	
YAKa60-E1	ND	ND
IPC	27,11	
YAKa60-E2	ND	ND
IPC	25,06	
YAKa64-E2	34,63	4,64E-02
IPC	27	
YAKa64-E3	33,41	2,12E-02
IPC	26,86	
YAKa64-E4	35,18	9,000E-03
IPC	29,71	
YAKa65-E1	34,12	1,26E-02
IPC	27,12	

n° Laboratoire	Ct	Quantité
YAKa69	28,25	7,30E-01
IPC	ND	
YAKa70	28,8	4,87E-01
IPC	ND	
YAKa71-E1	29,71	4,43E-01
IPC	ND	
YAKa71-E2	31,75	5,90E-02
IPC	26,1	
YAKa72-E1	35,93	5,90E-03
IPC	27,89	0
YAKa72-E2	35,68	2,39E-03
IPC	27,73	
YAKa72-PTB	ND	ND
IPC	26,12	
YAKa73-E1	32,59	6,00E-02
IPC	28,36	0
YAKa73-E2	32,41	3,47E-02
IPC	29,21	
YAKa74-E1	32,55	6,20E-02
IPC	26,18	0
YAKa74-E2	33,41	1,53E-02
IPC	26,53	
YAKa74-PTB	33,08	2,42E-02
IPC	26,31	
YAKa75-E1	33	2,14E-02
IPC	26	
YAKa75-PTB	33,9	1,34E-02
IPC	26,04	
YAKa76-E1	35,53	2,71E-03
IPC	26,38	
YAKa76-PTB	37,05	1,42E-03
IPC	25,98	
YAKa78	32,27	3,89E-02
IPC	26,25	
YAKa79DIL	33,17	4,020E-02
IPC	30,9	
YAKa80-E1	31,78	5,75E-02
IPC	26,35	
YAKa80-PTB	32,37	4,01E-02
IPC	25,86	
YAKa81-E1	ND	ND
IPC	26,39	

Bilan des Résultats de quantification par PCR en temps réel pour les échantillons osseux (suite).

n° Laboratoire	Ct	Quantité
YAKa27	30,15	0,141
IPC	30,46	
YAKa 28	ND	ND
IPC	ND	
YAKa 29	31,16	1,02E-01
IPC	29,46	
YAKa 30	34,54	9,60E-03
IPC	33,27	
YAKa 31	27,64	1,19
IPC	30,1	
YAKa 32	27,55	1,27
IPC	31,11	

n° Laboratoire	Ct	Quantité
YAKa43-E2	ND	ND
IPC	27,11	
YAKa43-E3	ND	ND
IPC	25,16	
YAKa44-E1	ND	ND
IPC	26,88	
YAKa44-E2	ND	ND
IPC	27,02	
YAKa44-E3	ND	ND
IPC	25,12	
YAKa45-E1	32,64	1,80E-01
IPC	27,01	

n° Laboratoire	Ct	Quantité
YAKa53-E1	ND	ND
IPC	ND	
YAKa53-E2	ND	ND
IPC	37,83	
YAKa54-E1	27,8	1,27
IPC	29,17	
YAKa54-E2	30,85	2,245E-01
IPC	30,62	
YAKa55-E1	36	2,51E-03
IPC	27,09	
YAKa55-E2	ND	ND
IPC	27,01	

n° Laboratoire	Ct	Quantité
YAKa65-E2	32,52	4,06E-02
IPC	27,82	
YAKa66	28,47	6,25E-01
IPC	27,52	
YAKa67	34,33	1,695E-02
IPC	29,43	
YAKa68-E1	28,4	6,50E-01
IPC	31,61	
YAKa68-E2	30,47	8,65E-02
IPC	26,47	
YAKa68-PTB	31,08	1,01E-01
IPC	26,98	

n° Laboratoire	Ct	Quantité
YAKa81-PTB	ND	ND
IPC	25,98	
YAKa82	31,34	8,30E-02
IPC	26,35	
YAKa84	37,26	2,36E-03
IPC	30,54	
YAKa86	31,85	5,50E-02
IPC	25,94	

Bilan des Résultats de quantification par PCR en temps réel pour les échantillons de cheveux.

n° Laboratoire	Extraction	Ct	Quantité (ng/μl)
YAKa34L	1	30,84	1,25E-01
		24,74	
YAKa34NL	1	30,72	1,37E-01
		24,83	
YAKa34L	2	31,16	8,70E-02
		27,06	
YAKa34NL	2	30,8	1,14E-01
		26,96	
YAKa37L	1	36	9,55E-04
		25,33	
YAKa37NL	1	ND	0,00E+00
		25,57	
YAKa37L	2	36,31	7,40E-03
		26,97	
YAKa37NL	2	ND	0,00E+00
		26,96	
YAKa39L	1	30,3	1,90E-01
		24,81	
YAKa39NL	1	30,86	1,23E-01
		24,9	
YAKa39L	2	29,13	3,83E-01
		27	
YAKa39NL	2	29,54	2,84E-01
		26,97	
YAKa45L	1	35,01	2,00E-03
		25,08	
YAKa45NL	1	ND	0,00E+00
		25,19	
YAKa45L	2	ND	0,00E+00
		26,96	
YAKa46L	1	35,42	1,48E-03
		25,18	
YAKa46NL	1	34,92	2,14E-03
		26,16	

n° Laboratoire	Extraction	Ct	Quantité (ng/μl)
YAKa46L	2	39,49	8,50E-04
		27,2	
YAKa55L	1	ND	0
		25,1	
YAKa55NL	1	39,76	5,75E-05
		25,28	
YAKa55L	2	39,07	1,13E-03
		26,93	
YAKa66L	1	28,1	5,10E-01
		26,96	
YAKa66NL	1	28,11	5,10E-01
		26,94	
YAKa66L	2	28,78	4,97E-01
		26,99	
YAKa66NL	2	27,9	9,45E-01
		26,96	
YAKa67L	1	35,4	2,49E-03
		27,05	
YAKa67NL	1	35,82	1,82E-03
		26,96	
YAKa68L	1	ND	0,00E+00
		27,11	
YAKa68NL	1	ND	0,00E+00
		27,03	
YAKa68L	2	ND	0,00E+00
		27,03	
YAKa68NL	2	35,9	2,71E-03
		27,07	
YAKa69L	2	34,85	5,85E-03
		26,92	
YAKa69NL	2	34,4	8,10E-03
		26,91	

n° Laboratoire	Extraction	Ct	Quantité (ng/μl)
YAKa70L	1	33,33	1,77E-02
		26,89	
YAKa70NL	1	33,44	1,64E-02
		26,84	
YAKa70L	2	33,3	1,82E-02
		27,01	
YAKa70NL	2	33,42	1,67E-02
		27,09	
YAKa79L	2	33,8	1,26E-02
		27,11	
YAKa79NL	2	31,19	8,45E-02
		26,91	
YAKa79L	3	31,21	8,35E-02
		26,98	
YAKa79NL	3	31,53	6,60E-02
		26,99	

VII.9 Annexe 9

Résultats des analyses STR obtenus avec le kit AmpflSTR® Profiler Plus™ sur les matrices amplifiées à l'aide du kit Genomiphi™ (Phi), avec (O) ou sans (N) étape de purification.

Echantillon	Purif	Amel	D13	D18	D21	D3	D5	D7	D8	FGA	vWA
YAKa33 Phi	O	XX		(32.2)		(12/12)			13/13		
YAKa33 Phi	N	XX				12/12			13/13		
YAKa33		XX	10/13	17/19	30/32.2	16/17	12/12	11/13	13/13	23/24	14/18
YAKa21 Phi	O	X	9		30/30			(8)			17/17
YAKa21 Phi	N	(XY)				(12/12)					
YAKa21		XY	8/12	14/14	30/30	15/15	12/12	8/11	13/13	21/24	17/17
YAKa31 Phi	O	XY	8		29/29	17	10	8/11	10/16	22	16
YAKa31 Phi	N	XY	(8)		(29)	16/17	7/10		10		
YAKa31		XY	8/14	15/18	29/29	16/17	7/10	8/8	10/16	22/23	16/19
YAKa32 Phi	O	XX	11/11		28	17		9	13		14
YAKa32 Phi	N	XX			28	15/(17)			12		20
YAKa32		XX	11/11	15/18	28/30	15/17	9/11	9/10	12/13	22/24	14/20
YAKa18 Phi	O	XX	11		30/32	14/17	11/11	(8)	13	24/24	16
YAKa18 Phi	N	XX	9/11		(30)	14/17	11/11		10/13		16/17
YAKa18		XX	9/11	15/18	29/30	14/17	11/11	8/10	10/13	24/24	16/17
YAKa20 Phi	O	XX			31			8	13/13		19
YAKa20 Phi	N	XX	8/(13)				11/13		13/13		18
YAKa20		XX	8/13	14/15	30/31	15/16	11/13	8/11	13/13	23/24	18/19
YAKa19 Phi	O	XY	11		29	15/15	12/12		13/13	23	14/14
YAKa19 Phi	N	XY	11			15/15	10/12	9	13/13		14/14
YAKa19		XY	11/12	14/15	29/34.2	15/15	12/12	8/9	13/13	23/24	14/14
SyA Phi	O	XY	12/13	14/14	31/32.2	16/17	11/12	9/12	13/13	22/25	17/19
SyA Phi	N	XY	12/13	14/14	31/32.2	16/17	11/12	9/12	13/13	22/25	17/19
SyA		XY	12/13	14/14	31/32.2	16/17	11/12	9/12	13/13	22/25	17/19

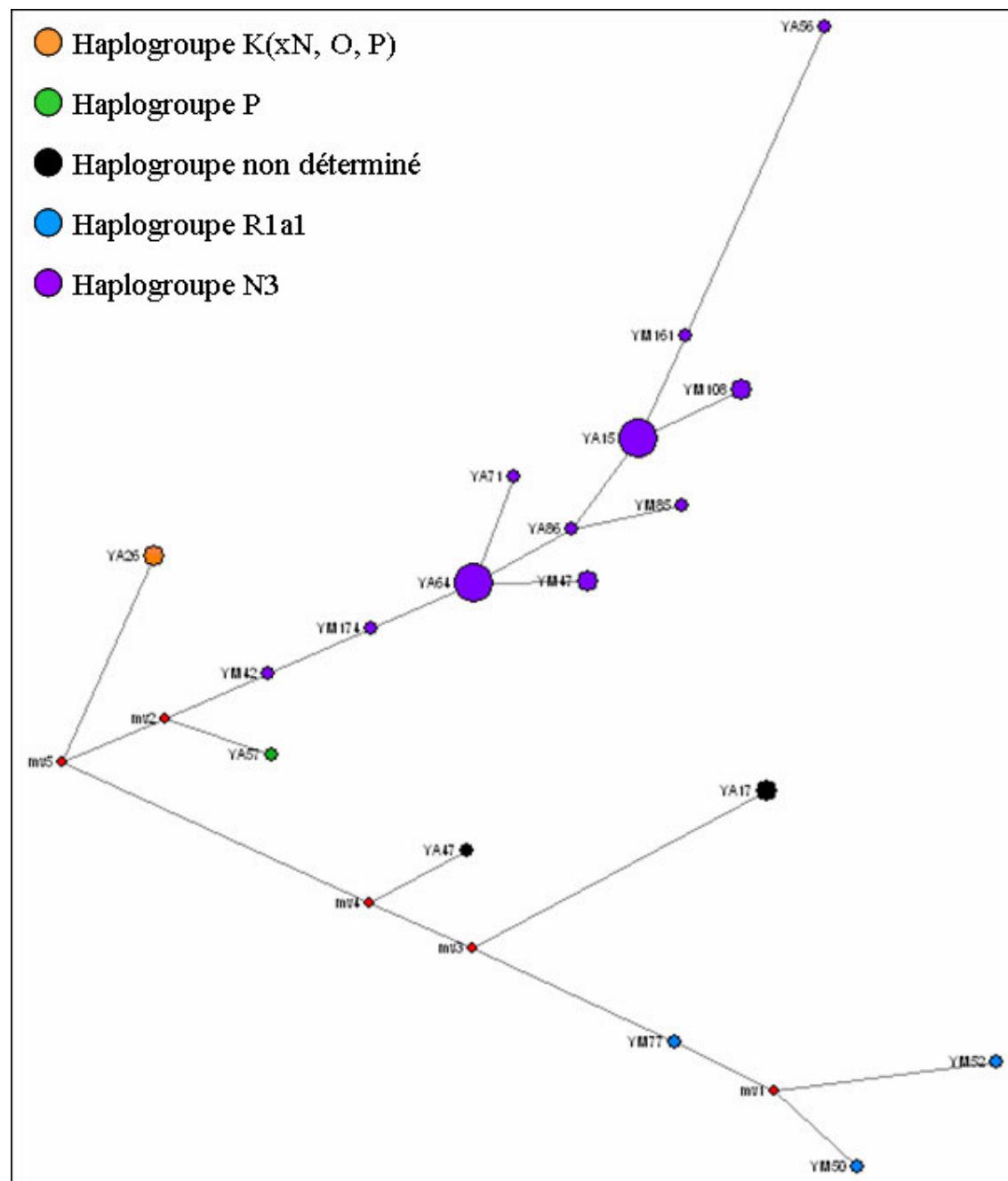
VII.10 *Annexe 10*

Diversités génétiques des loci STR du chromosome Y pour différentes populations. Les populations dont la référence n'est pas précisée correspondent aux populations de la base de données (cf Annexe 6).

Population	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439
Yakoutes anciens	0,3118	0,2897	0,6376	0,2355	0,3913	0,706	0,2355	0,3549	0,2446
Yakoutes actuels (échantillons du laboratoire)	0,1385	0,2051	0,583	0,1192	0,2672	0,5999	0,1351	0,1366	0,0611
Yakoutes actuels (Keyser-Tracqui, 2006)	0,0769	0,1476	0,6093	0,0768	0,1509	0,2797	0,0769	0,0768	0,0769
Yakoutes actuels (Pakendorf, 2002)	0,2622	0,2589	0,5949	0,2151	0,1613	0,4317	0,2082	ND	ND
EyginGol anciens	0,7213	0,4678	0,821	0,6265	0,4678	0,7213	0,6526	ND	ND
EyginGol modernes	0,7771	0,4234	0,6404	0,6786	0,4688	0,5886	0,5601	ND	ND
Mongols (Keyser-Tracqui, 2006)	0,8175	0,5708	0,7838	0,6941	0,5058	0,6702	0,5162	0,3555	0,6036
Mongols (Zhu, 2005)	0,6978	0,6749	0,7566	0,7207	0,4242	0,8113	0,6813	0,4354	0,7195
Mongols (Kwak, 2005)	0,6921	0,5896	0,791	0,6711	0,4169	0,603	0,6147	0,1367	0,0611
Bouriates (Wozniak, 2006)	0,58	0,4103	0,6063	0,4103	0,2519	0,3948	0,3738	0,2118	0,6805
Ouïgurs (Zhu, 2005)	0,7592	0,6102	ND	0,7239	0,5388	0,7593	0,6369	ND	ND
Siberie (Derenko, 2006)	0,5008	0,4665	0,6069	0,5663	0,4395	0,0271	0,0704	0,0622	0,4641
Kazakhs	0,2924	0,2789	ND	0,5697	0,1557	0,1191	0,0809	ND	ND
L_Kirghiz	0,5254	0,5157	ND	0,5427	0,6486	0,4573	0,4743	ND	ND
H_Kirghiz	0,3586	0,2855	ND	0,2592	0,31	0,2581	0,092	ND	ND
Turcs_Anatolie	0,6963	0,5701	0,755	0,744	0,4865	0,5441	0,6551	0,7161	0,6897
Turcs_centre_Anatolie	0,706	0,5012	0,7452	0,716	0,4673	0,5994	0,6656	0,7212	0,748

VII.11 Annexe 11

Median joining network réalisé à partir des données des STR du chromosome Y pour l'ensemble des échantillons anciens.



Résultats obtenus pour l'analyse des SNP du chromosome Y par la méthode SNaPshot. Les lignes grisées correspondent aux échantillons actuels analysés.

n° laboratoire	M89 C/T	M9 C/G	RPS4Y C/T	M216 C/T	M217 A/C	TAT T/C	M175 del	M45 G/A	92R7 GA/A	M242 C/T	M3 C/T	M173 A/C	M17 4G/3G	Haplogroupe Y
YAKa19	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa26	T	G	C	C	A	T	T	G	GA	C	C	A	4G	K(xN,O,P)
YAKa34	T	G	C	C	A	C	T	G	GA	C	C	-	-	N3
YAKa51	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa56	T	G	C	C	T	C	T	G	GA	C	C	A	4G	N3
YAKa65	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAKa69	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa69d	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAKa71	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa80	-	-	-	-	-	-	-	-	-	-	-	-	-	-
YAKa81	-	-	C	-	A	-	T	G	-	-	-	A	4G	-
YAKa86	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK22	T	G	C	C	A	C	T	G	GA	C	C	-	-	N3
YAK42	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK47	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK50	T	G	C	C	A	T	T	A	A	C	C	C	3G	R1a1
YAK52	T	G	C	C	A	T	T	A	A	C	C	C	3G	R1a1
YAK61	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK77	T	G	C	C	A	T	T	A	A	C	C	C	3G	R1a1
YAK85	T	G	C	C	A	C	T	G	GA	C	C	-	-	N3
YAK94	T	G	C	C	A	T	T	G	GA	C	C	A	4G	K(xN,O,P)
YAK108	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK161	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3
YAK174	T	G	C	C	A	C	T	G	GA	C	C	A	4G	N3

VII.12 *Annexe 12*

Matrice des distances F_{st} établies à partir des données des STR du chromosome Y pour l'ensemble des échantillons anciens et différentes populations de notre base de données.

Populations	Kazakh	Kirghiz_L	Kirghiz_H	Ouigours	Russes	Eygin Gol anciens	Eygin Gol actuels	Ouigours_Zhu	Mongols_Zhu	Bourriates	Mongols_Kolman	Yakoutes actuels	Yakoutes anciens
Kazakh	0												
Kirghiz_L	0,21253	0											
Kirghiz_H	0,41173	0,06821	0										
Ouigours	0,15012	0,07036	0,25035	0									
Russes	0,28761	0,17329	0,29079	0,14688	0								
Eygin Gol anciens	0,33583	0,13098	0,36469	0,05056	0,01287	0							
Eygin Gol actuels	0,29376	0,19934	0,36582	0,12326	0,03707	0,02761	0						
Ouigours_Zhu	0,29635	0,18821	0,32782	0,10575	0,07322	0,01017	0,03752	0					
Mongols_Zhu	0,30728	0,20257	0,35045	0,10555	0,1147	0,0278	0,06527	0,00917	0				
Bourriates	0,47386	0,39241	0,53363	0,29263	0,23136	0,19104	0,15338	0,18177	0,18013	0			
Mongols_Kolman	0,3002	0,21256	0,37348	0,12101	0,0725	0,00053	0,01246	0,03261	0,0423	0,08972	0		
Yakoutes actuels	0,73317	0,58513	0,69675	0,51506	0,52381	0,52524	0,50594	0,39179	0,408	0,47307	0,439	0	
Yakoutes anciens	0,68779	0,45536	0,64725	0,35442	0,40537	0,35766	0,36522	0,26928	0,27915	0,37147	0,30555	-0,01029	0

VII.13 *Annexe 13*

Matrice des distances F_{st} établies à partir des données des STR du chromosome Y pour l'ensemble des échantillons anciens répartis par périodes chronologiques et différentes populations de notre base de données.

Populations	Kazakh	Kirghiz_L	Kirghiz_H	Ouigours	Russes	Eygin Gol anciens	Eygin Gol actuels	Ouigours_Zhu	Mongols_Zhu	Bourriates	Mongols_Kolman	Yakoutes actuels	YAKa_P1	YAKa_P2	YAKa_P3
Kazakh	0														
Kirghiz_L	0,21253	0													
Kirghiz_H	0,41173	0,06821	0												
Ouigours	0,15012	0,07036	0,25035	0											
Russes	0,28761	0,17329	0,29079	0,14688	0										
EyginGol anciens	0,33583	0,13098	0,36469	0,05056	0,01287	0									
EyginGol actuels	0,29376	0,19934	0,36582	0,12326	0,03707	0,02761	0								
Ouigours_Zhu	0,29635	0,18821	0,32782	0,10575	0,07322	0,01017	0,03752	0							
Mongols_Zhu	0,30728	0,20257	0,35045	0,10555	0,1147	0,0278	0,06527	0,00917	0						
Bourriates	0,47386	0,39241	0,53363	0,29263	0,23136	0,19104	0,15338	0,18177	0,18013	0					
Mongols_Kolman	0,3002	0,21256	0,37348	0,12101	0,0725	0,00053	0,01246	0,03261	0,0423	0,08972	0				
Yakoutes actuels	0,73317	0,58513	0,69675	0,51506	0,52381	0,52524	0,50594	0,39179	0,408	0,47307	0,439	0			
YAKa_P1	0,67197	0,37065	0,63465	0,22897	0,30157	0,21425	0,23221	0,13559	0,15704	0,30335	0,18942	-0,02799	0		
YAKa_P2	0,77148	0,51874	0,72864	0,42114	0,4945	0,47455	0,46071	0,3524	0,36617	0,49932	0,40359	-0,01301	0,11765	0	
YAKa_P3	0,67083	0,401	0,63014	0,28518	0,36105	0,26827	0,31268	0,21548	0,21842	0,32949	0,25022	-0,0041	-0,17857	0,04407	0

VII.14 Annexe 14

Matrice des distances F_{st} établies à partir des données des STR du chromosome Y pour l'ensemble des échantillons anciens répartis selon la localisation géographique et différentes populations de notre base de données.

Populations	Kazakh	Kirghiz_L	Kirghiz_H	Ouigours	Russes	Eygin Gol anciens	Eygin Gol actuels	Ouigours_Zhu	Mongols_Zhu	Bourriates	Mongols_Kolman	Yakoutes actuels	YAKa_Khang	YAKa_Chur	YAKa_Tatt
Kazakh	0														
Kirghiz_L	0,21253	0													
Kirghiz_H	0,41173	0,06821	0												
Ouigours	0,15012	0,07036	0,25035	0											
Russes	0,28761	0,17329	0,29079	0,14688	0										
Eygin Gol anciens	0,33583	0,13098	0,36469	0,05056	0,01287	0									
Eygin Gol actuels	0,29376	0,19934	0,36582	0,12326	0,03707	0,02761	0								
Ouigours_Zhu	0,29635	0,18821	0,32782	0,10575	0,07322	0,01017	0,03752	0							
Mongols_Zhu	0,30728	0,20257	0,35045	0,10555	0,1147	0,0278	0,06527	0,00917	0						
Bourriates	0,47386	0,39241	0,53363	0,29263	0,23136	0,19104	0,15338	0,18177	0,18013	0					
Mongols_Kolman	0,3002	0,21256	0,37348	0,12101	0,0725	0,00053	0,01246	0,03261	0,0423	0,08972	0				
Yakoutes actuels	0,73317	0,58513	0,69675	0,51506	0,52381	0,52524	0,50594	0,39179	0,408	0,47307	0,439	0			
YAKa_Khang	0,67197	0,37065	0,63465	0,22897	0,30157	0,21425	0,23221	0,13559	0,15704	0,30335	0,18942	-0,02799	0		
YAKa_Chur	0,77148	0,51874	0,72864	0,42114	0,4945	0,47455	0,46071	0,3524	0,36617	0,49932	0,40359	-0,01301	0,11765	0	
YAKa_Tatt	0,67083	0,401	0,63014	0,28518	0,36105	0,26827	0,31268	0,21548	0,21842	0,32949	0,25022	-0,0041	-0,17857	0,04407	0

VII.15 Annexe 15

Matrice des distances F_{st} établies à partir des données des mitochondriales pour l'ensemble des échantillons anciens répartis par périodes chronologiques et différentes populations de notre base de données.

Populations	Altaï	Bourriates	Eygin Gol anciens	Eygin Gol actuels	Evenks	Kazakhs	Sibérie Occidentale	Kirghiz	Koryaks	Mansi	Mongols	Yakoutes actuels	Tuvas	Ouigours	Uzbeks	YAKa old	YAKa P1	YAKa P2	YAKa P3
Altaï	0,0000																		
Bourriates	0,0152	0,0000																	
Eygin Gol anciens	0,0327	0,0011	0,0000																
Eygin Gol actuels	0,0189	0,0023	0,0082	0,0000															
Evenks	0,0180	0,0062	0,0088	0,0115	0,0000														
Kazakhs	0,0039	0,0090	0,0251	0,0069	0,0194	0,0000													
Sibérie Occidentale	0,0725	0,0725	0,1168	0,0692	0,0937	0,0352	0,0000												
Kirghiz	0,0099	0,0064	0,0211	0,0055	0,0131	0,0023	0,0485	0,0000											
Koryaks	0,0996	0,0714	0,1022	0,0710	0,0842	0,0766	0,1733	0,0861	0,0000										
Mansi	0,0347	0,0527	0,0813	0,0508	0,0592	0,0267	0,0207	0,0318	0,1298	0,0000									
Mongols	0,0081	0,0035	0,0047	0,0028	0,0094	0,0050	0,0544	0,0032	0,0717	0,0456	0,0000								
Yakoutes actuels	0,0390	0,0241	0,0336	0,0218	0,0188	0,0338	0,1155	0,0326	0,0678	0,0787	0,0279	0,0000							
Tuvas	0,0419	0,0162	0,0305	0,0116	0,0141	0,0320	0,1262	0,0272	0,0946	0,0732	0,0238	0,0098	0,0000						
Ouigours	0,0129	0,0147	0,0280	0,0113	0,0292	0,0013	0,0367	0,0034	0,0963	0,0327	0,0059	0,0505	0,0471	0,0000					
Uzbeks	0,0305	0,0307	0,0669	0,0197	0,0477	0,0055	0,0175	0,0106	0,1288	0,0084	0,0224	0,0648	0,0434	0,0056	0,0000				
YAKa P0	0,0199	-0,0143	-0,0116	0,0028	-0,0207	0,0263	0,2118	0,0147	0,0761	0,0945	0,0025	-0,0057	-0,0062	0,0530	0,1063	0,0000			
YAKa P1	0,0624	0,0239	0,0383	0,0271	0,0114	0,0518	0,2193	0,0545	0,0804	0,1085	0,0371	-0,0214	-0,0254	0,0854	0,1050	-0,0442	0,0000		
YAKa P2	0,0277	0,0146	0,0333	0,0077	0,0149	0,0149	0,1073	0,0138	0,0520	0,0649	0,0076	-0,0057	0,0135	0,0191	0,0356	0,0367	0,0159	0,0000	
YAKa P3	0,0396	0,0201	0,0327	0,0151	0,0192	0,0280	0,1276	0,0283	0,0503	0,0758	0,0208	-0,0090	0,0032	0,0428	0,0541	-0,0074	-0,0182	-0,0293	0,0000

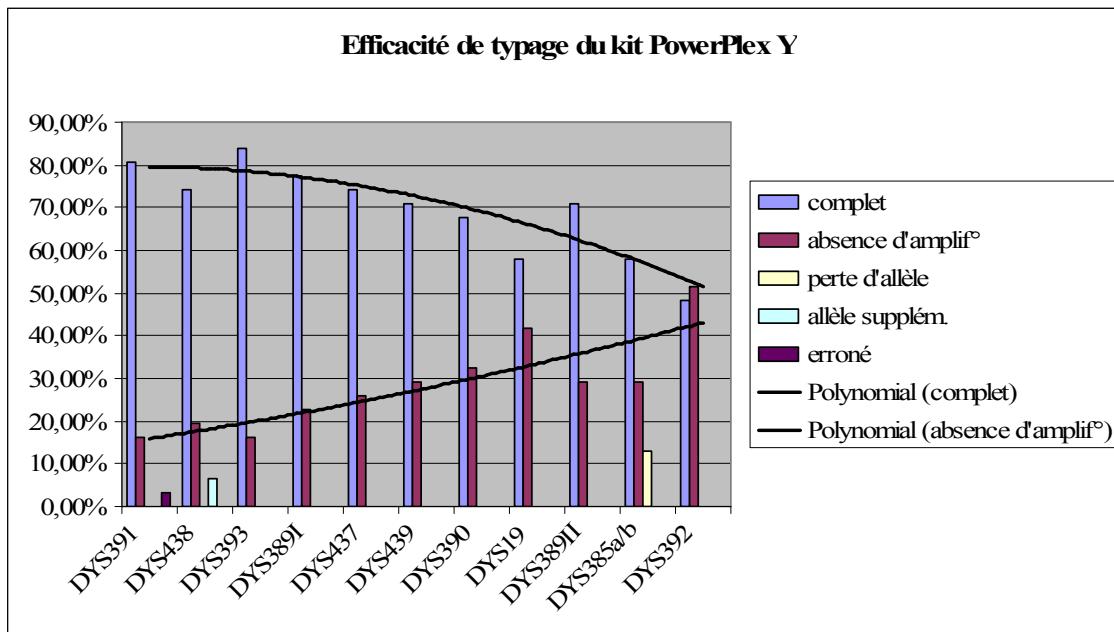
VII.16 Annexe 16

Matrice des distances F_{st} établies à partir des données mitochondrielles pour l'ensemble des échantillons anciens répartis selon la localisation géographique et différentes populations de notre base de données.

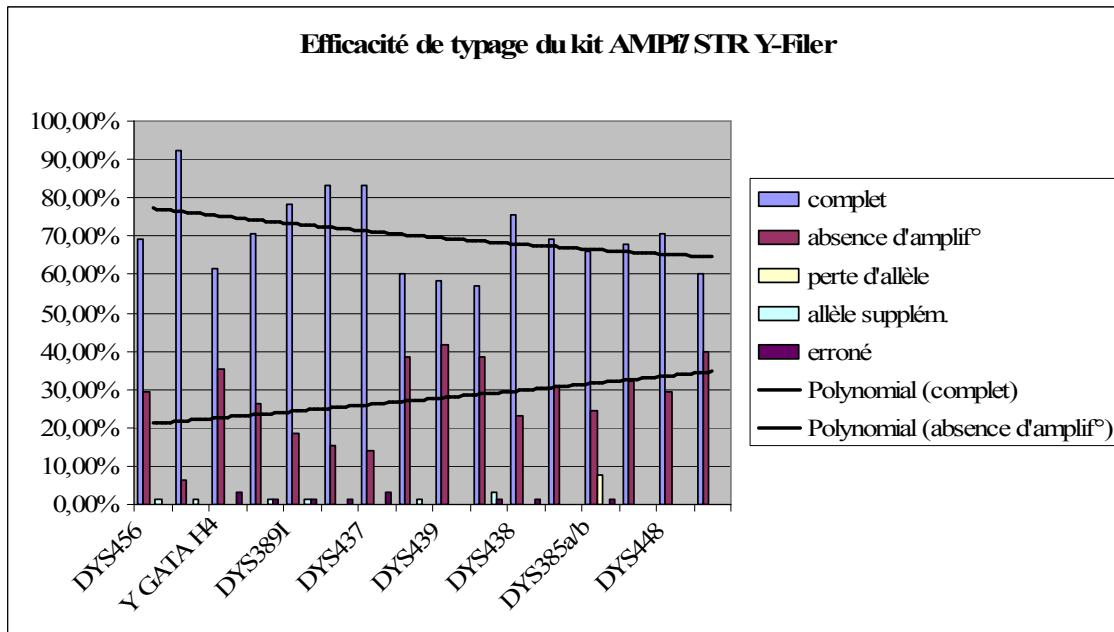
Populations	Altaï	Bourriates	Eygin Gol anciens	Eygin Gol actuels	Evenks	Kazakhs	Kirghiz	Mansis	Sibérie Orientale	Koryaks	Mongols	Yakoutes actuels	Tuvas	Ouigours	Uzbeks	YAKA_Kangh	YAKA_Chur	YAKA_Tatta
Altaï	0																	
Bourriates	0,01519	0																
Eygin Gol anciens	0,0327	0,00107	0															
Eygin Gol actuels	0,01887	0,00229	0,00822	0														
Evenks	0,01797	0,00615	0,00882	0,01152	0													
Kazakhs	0,00391	0,00902	0,02508	0,00686	0,01936	0												
Kirghiz	0,00994	0,00643	0,02113	0,0055	0,01311	0,00232	0											
Mansis	0,03473	0,05269	0,08132	0,05075	0,05915	0,02667	0,03177	0										
Sibérie Orientale	0,07254	0,07247	0,11681	0,0692	0,09369	0,03521	0,0485	0,02069	0									
Koryaks	0,09958	0,07135	0,10223	0,07098	0,08422	0,07662	0,08614	0,12982	0,1733	0								
Mongols	0,00811	0,00352	0,00468	0,00278	0,00943	0,00502	0,00317	0,04561	0,05437	0,07172	0							
Yakoutes actuels	0,03897	0,02411	0,03356	0,02175	0,01881	0,0338	0,03264	0,07868	0,11547	0,06775	0,02787	0						
Tuvas	0,04192	0,01617	0,03048	0,01158	0,01413	0,03198	0,02724	0,07319	0,12619	0,09459	0,02375	0,00975	0					
Ouigours	0,01407	0,01413	0,02729	0,00958	0,02896	0,00077	0,00353	0,0332	0,03644	0,09535	0,00556	0,04878	0,04406	0				
Uzbeks	0,03054	0,03067	0,06686	0,01965	0,04767	0,00549	0,01056	0,00844	0,01752	0,12879	0,02241	0,06482	0,04337	0,00372	0			
YAKA_Kangh	0,03266	0,02038	0,03965	0,02333	0,02268	0,0277	0,03217	0,06598	0,13325	0,04354	0,02198	0,0094	0,01348	0,04399	0,06268	0		
YAKA_Chur	0,11413	0,08117	0,09086	0,06359	0,07416	0,09386	0,08959	0,15558	0,21221	0,15205	0,07403	0,0374	0,04766	0,09697	0,11261	0,1027	0	
YAKA_Tatta	0,04653	0,02078	0,04044	0,02551	0,02055	0,03024	0,03377	0,07738	0,15172	0,03452	0,02648	-0,00079	0,01687	0,05307	0,08074	-0,03467	0,0969	0

VII.17 Annexe 17

Courbes polynomiales pour les profils complets (courbes supérieures) et pour les absences d'amplification (courbes inférieures) calculées à partir des efficacités de typages des kits PowerPlex® Y (Annexe 17a) et AMPflSTR® Y-Filer™ (Annexe 17b).



Annexe 17a.



Annexe 17b.

VIII PUBLICATIONS

Early Influence of the Steppe Tribes in the Peopling of Siberia

Amory S^{1,2}, Crubézy E², Keyser C¹, Alekseev A N³, Ludes B^{1,2}.

¹Laboratory of Molecular Anthropology, Institute of Legal Medicine, Strasbourg, France 67085.

²Laboratory of Anthropobiology, University Paul Sabatier, CNRS, UMR 8555, Toulouse, France 31400.

³Institute of Archaeology and Ethnography, Yakutsk State University, Belinsky Street 58, 677891 Yakutsk, Russia

KEY WORDS: ANCIENT DNA, ORIENTAL SIBERIA, MITOCHONDRIAL DNA, AUTOSOMAL STRs, Y CHROMOSOME STRs.

¹ Corresponding author: Sylvain Amory, Laboratory of Molecular Anthropology, Institute of Legal Medicine, 11 rue Humann Strasbourg Cedex, France 67085.

Tel: +33 (0) 390 243348. Fax: +33 (0) 390 243362.

E-mail address: sylvain.amory@iml-ulp.u-strabg.fr

Abstract The Yakuts, Middle Age Turkic speakers (XV/XVIth century), are widely accepted as the first settlers from the Altaï/Baïkal area in east Siberia. They are supposed to have introduced horses and developed metallurgy in this geographical area during the XVth or XVIth century AD. The analysis of the Siberian grave of Pokrovsk, recently discovered near the Lena River (61° 29' N) and dated by accelerator mass spectrometry from 2400 to 2200 years BP may provide new elements to test this hypothesis. The exceptional combination of various artefacts as well as the mitochondrial DNA data extracted from the bone remains of the Pokrovsk man might prove the existence of previous contacts between autochthonous hunters of Oriental Siberia and the nomadic horse breeders from the Altaï/Baïkal area (Mongolia and Buryatia). Indeed, the stone arrow head and the harpoons relate this man to the traditional hunters of the Taïga. Some artefacts made of horse bone and the pieces of armour, however, are related to the tribes of Mongolia and Buryatia of the Xiongnu period (IIIrd century BC). This affinity has been confirmed by the match of the mitochondrial haplotype of this subject with a woman of the Egyin Gol necropolis (Mongolia, IInd/IIrd century AD) as well as with two modern Buryats. This result allows us to postulate that contacts between southern steppe populations and Siberian tribes occurred before the XVth century.

Several hypotheses have been proposed concerning the ethnogenesis of the Yakuts (or Sakhas as they call themselves), but the question of the origin of this ethnic group remains unclear. According to the Russian archaeologist AP Okladnikov, the origin of the Yakuts could be traced back to the Kurykan tribe (Okladnikov, 1955). This group of Türkic cattle breeders from the Baikal region, pressed by the expansion of the Mongolic tribes in the XIth-XIIIth centuries, moved northward to the Lena Basin. During this migration throughout Oriental Siberia the Kurykans displaced and mixed with the autochthonous populations. Thus, the Sakhas are supposed to be the descendants of these two ethnic groups. These two influences are still present in the Yakut population as evidenced by their language composed by Turkic and Mongolic words (Balzer, 1994) and their genetic affinities. Indeed, the genetic proximity between modern Yakuts and Evenks was confirmed by various genetic studies (Derenko et al., 2002; Pakendorf et al., 1999, Pakendorf et al., 2003).

Taking into account more recent archaeological findings, A Gogolev (2000) proposed an alternative theory. According to this author the Yakut culture originated in the culture of the nomadic tribes from southern Siberia and Central Asia of the Scythian period. For this author, although some türkic and Mongolic components are apparent in the culture of the Sakhas, the real origin of the Yakuts is to be found in the Kulun Atakh culture of the XVth century. People of the Kulun-Atakh culture reached Siberia by the rivers Lena, Vilyuy and Aldan assimilating small ethnic groups inhabiting these regions and introduced horse and cattle-breeding as well as metallurgy. The rise of annual temperatures during the Medieval Warm Period, between the XIth and the XIVth Century AD, (Bradley et al. 2003; Mackay et al., 2005) could have been favourable to the expansion of nomadic tribes through Oriental Siberia. The contacts between this group and aboriginal cultures combined with efforts to adapt their lifestyle to the extreme climatic conditions resulted in the formation of the Yakuts at the end of the XVIth century (Gogolev A, 2000). Thus, the Yakuts and their predecessors from the Kulun-Atakh culture are considered as the earliest evidence of human beings from the Altai/Baikal area (a vast region including the Altai-Sayan Mountains, the Baikal Lake and northern Mongolia) in east Siberia. Thereafter, the beginning of the Russian colonization, characterized by the establishment of a fort in Yakutsk in 1632, broke the isolation of the Sakhas and allowed their expansion throughout Oriental Siberia (Sierszewski, 1993).

Prior to the Yakut period Oriental Siberia was occupied by hunters/fishers groups until the end of the first millennium AD (Alekseev, 1996). These first inhabitants are considered as

the ancestors of modern Youkaguirs. They were later assimilated or pushed back by the expansion of the Tungus-Manchurian nomadic tribes. Despite the extremely harsh continental climate, contacts between nomadic tribes from Eurasian steppe and autochthonous people from central and northern Siberia seem to have taken place from the Neolithic to the Iron Age (Alekseev et al., 1996). These contacts could have been favoured by the deeply frozen rivers which became preferential roads of migration during winter. Furthermore, since the expeditions of Okladnikov (1945-1955) (Okladnikov, 1955) the question of the origin of the metallurgy has been frequently asked. Indeed, bronze and iron techniques may have been developed under a southern influence (from the Amur River) (Konstantinov, 1975). The genetic study of individuals dated from this period could provide new elements to test this theory.

The aim of this article was to determine the origin of the Pokrovsk individual using a combination of archaeological and molecular data. Even though a single individual is not representative of a whole population, the analysis of a 2400/2200 years old subject can give information on the relationship between nomadic tribes from the Altaï/Baïkal area and autochthonous people from Siberia prior to the invasions of the Mongols in the XIth-XIIIth centuries.

Indeed, molecular analysis is an efficient tool to appreciate the ethnic origins of an individual and his or her genetic relationships with neighbouring populations. In this study, three types of molecular markers were used to analyse the DNA extracted from a femora of the Pokrovsk subject. Autosomal Short Tandem Repeats (STR) allowed to determine whether the sample was contaminated by exogenous DNA. Y chromosome and mitochondrial DNA allowed the investigation, respectively, of the paternal and maternal lineage. Furthermore, the mitochondrial DNA, due to its haploid and strictly maternal mode of transmission, is a powerful tool to investigate relationships between populations.

Material and Methods

The Pokrovsk grave. The survey of archaeological sites is difficult in Siberia due to the forest cover and to the low density of building sites. The grave was located at the top of a glacial terrace ($61^{\circ} 29' 44''$ N, $129^{\circ} 09' 51''$ E) on the western side of Pokrovka River near the junction with the Lena River, 75 km south of Yakutsk. Radiocarbon dating indicated an age comprised between 390 and 190 BC (Beta Analytic Radiocarbon Dating Laboratory,

Miami, Florida). The depth of the permafrost was 60cm and, although no leather, muscle or any soft tissue were retrieved, the bones were well preserved. The skeleton of this adult male is gracile, the skull is brachycranic with some Mongolic traits but less accentuated than in the Middle Age Yakut population and without torus mandibularis which is common in this population (Crubézy et al., submitted). The osseous remains and the cultural material were fully recovered and deposited in 1992 in the Pokrovsk museum where they were studied by a multidisciplinary team.

Artefacts found in the grave include a wide variety of bone tools: harpoon heads, elements of an armour made of reindeer bone and a flint arrowhead. These artefacts are connected to Siberian archaic traditions. Two other artefacts were also discovered: (i) an iron arrowhead showing similarities with the ones found in the south Siberia; (ii) a punch made in vestigial horse metatarsal that indicates the earliest presence of this animal in this region (Full description in Crubézy et al., submitted).

DNA extraction and purification. To eliminate surface contamination we abraded the bone to a depth of 2-3mm with a sanding machine (Dremel, Breda, Netherlands) and used a column drill with a surgical trepan to generate bone powder.

DNA was extracted according to Keyser-Tracqui and Ludes (2005) with a 16 hour incubation of 2g of bone powder in extraction buffer at 50°C.

DNA quantitation. DNA quantitation enabled us to adjust the DNA amount necessary to the different analyses and to determine the best PCR cycling conditions (data not shown). It also allowed us to verify the presence of PCR inhibitors as an internal PCR control (IPC, synthetic sequence not existing in nature) is included in each reaction. The IPC threshold cycle (C_t) monitors PCR inhibitors. Quantitation was performed on an ABI Prism 7000 SDS (Applied Biosystems, Courtaboeuf, France) using the Quantifiler® Human DNA Quantification Kit (Applied Biosystems) according to the manufacturer's protocol.

Autosomal STR analysis. Amplifications were performed using the AmplifSTR® Profiler Plus® kit (Applied Biosystems) which allows the simultaneous amplification of 9 Short Tandem Repeat (STR) loci and the amelogenin marker for sex determination.

PCR conditions were modified from the manufacturer's protocol in order to reduce the

amount of ancient DNA used. The number of cycles was increased to 34 cycles according to the quantitation results (Gill et al, 2000, Whitaker et al, 2001). PCR amplification was carried out on a T3 Biometra thermalcycler (Whatman Biometra, Goettingen, Germany). Analysis was performed on an ABI Prism 3100® (Applied Biosystems) with the Genemapper software (Applied Biosystems).

Y-chromosome STR analysis. The DNA of the Pokrovsk specimen was analyzed with the PowerPlex Y® System (Promega, Madison Wisc.) which includes 11 STR loci. For PCR amplification (performed on a T3 Biometra thermalcycler) we used the conditions recommended by the manufacturer.

The TAT Single Nucleotide Polymorphism (SNP), located on the Y-chromosome, was studied by Matrix Assisted Laser Desorption Ionisation-Time Of Flight Mass Spectrometry (MALDI-TOF MS) (Keyser-Tracqui et al, 2004).

Mitochondrial DNA analysis. The first hyper variable segment (HV1) of the mitochondrial control region was amplified and sequenced. A 421-base pair (bp) fragment was sequenced from position 15989 to 16410 of the Cambridge Reference Sequence (Anderson et al, 1981). To avoid DNA degradation-induced amplification problems we amplified separately the two HV1 sub regions (A and B) with the F15989/R16239 and F16190/R16410 primer couples, respectively (Gabriel et al, 2001, Ivanov et al, 1996). The amplification was performed using the Hot Goldstar DNA polymerase (Eurogentec, Seraing, Belgium) according to the manufacturer's instructions. The amplification products were separated on a 1% agarose gel and purified on Microcon PCR filters (Millipore, Billerica, MA). Sequencing reactions were performed with the ABI Prism BigDye Terminator® Cycle Sequencing Ready Reaction kit (Applied Biosystems) according to the manufacturer's instructions. The products were detected on an ABI Prism 3100® automatic sequencer (Applied Biosystems) and analyzed with the Sequence Navigator Software package.

Cloning. The 421bp fragment was amplified using the F15989 and R16410 primers. Bacterial clones containing HVI fragment carrying plasmids were obtained by the use of a laboratory made T/A vector based on pUC19 and electroporation. The plasmids were prepared according to a 96-well adapted protocol (Hatsch et al, unpublished).

Comparative analysis. In order to investigate the population affinities of this individual we compared the HV1 sequences obtained with modern and ancient population data.

Mitochondrial DNA sequences were aligned using the Bioedit Software (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) with 2843 HV1 sequences of various ethnic groups from literature (Table 1). The distribution of this haplotype was analysed using the Blast2 software (http://www.infobiogen.fr/services/analyseq/cgi-bin/blast2_in.pl).

Table 1

Measures against contamination. A number of measures were applied to minimize the risk of contamination. Excavation was performed wearing facial masks and latex gloves. All people handling the material or working in the laboratory were genotyped and their profiles were compared to that of the ancient sample. All the steps of the study, from abrasion to amplification, were performed in a laboratory dedicated to ancient DNA with all the precautions against modern DNA contamination.

Open tube experimentations were made in a room separated from the rest of the laboratory and cloning was performed in a different laboratory. Laboratories, laboratory ware, coats, plastic ware, reagents, pipettes, benches and equipment were irradiated under ultraviolet light (250nm for 45 min). Samples were handled wearing gloves, face masks and laboratory coats dedicated to ancient DNA. Reagents were prepared in small volumes to avoid multiple uses. Aerosol-resistant filter tips were used for all reactions with ancient DNA. Extraction and negative blanks were included in all PCR assays.

Multiple, time-spaced extractions were performed and PCR amplifications were made at least three times to ensure the accuracy of the results.

Results

Autosomal STRs results. The genetic typing by megaplex amplifications of the Pokrovsk subject was successful since it was possible to establish a complete allelic profile. To avoid errors due to allelic dropout, spurious alleles or other stochastic variations which occur frequently with low copy number templates, we increased the number of amplification cycles. Even though the DNA concentration of this extract was low, it yielded a complete profile for two different amplifications, showing that the template was not highly degraded.

The results obtained were compared with the STR profiles of all technicians and archaeologists and no matches were found. Moreover, the morphological and genetic sex determinations are in accordance. These findings suggest that the results obtained from amplified products are genuine.

Table 2

Table 3

Y-chromosome STRs results. None of the amplifications performed on the three different DNA extractions gave results usable for comparison with Y-chromosome STRs databases (Table 4). Thus, it was impossible to determine the putative affinities of this subject with other populations. As far as the SNP marker is concerned, this subject presents the TAT T allele.

Table 4

Table 5

Mitochondrial DNA results. A 421 bp sequence was obtained for two different extractions and compared with the CRS. Three variable nucleotide positions were observed: 16223 T, 16362 C and 16368 C, all corresponding to transitions (Table 6). This haplotype was attributed to the D haplogroup.

We employed the mtDNA nomenclature used in Macaulay et al. (Macaulay et al, 1999) for the haplogroup determination. This sequence contains two mutations characteristic of haplogroup D.

Table 6

Table 7

Cloning. Sixteen positive clones were directly sequenced with the following primers: M13 (GTAAAACGACGGCCAG) and RP (AACAGCTATGACCATGATTAC). The validity of the haplotype was confirmed since the polymorphic positions 16223, 16362 and 16368 were retrieved in all clones. Additional positions were found in some particular and unique clones, which were not considered for the consensus (Table 8). The Consensus

Confidence Program (Consensus Confidence Program Test Version 1.12 Copyright Mim Bower 2004) was used to test the validity of the cloning of our sample. The results of this test confirm that this haplotype is genuine, with a confidence rate over 95% for the haplotype of the Pokrovsk subject.

Discussion

Exogenous contamination seems to have been eliminated since autosomal STRs profile and HV1 sequence were reproducible and did not correspond to genotypes observed in members involved either in the archaeological or in the genetic study (Tables 3, 5, 7). Moreover, the same results were obtained on the three different extractions and for at least two different amplifications. The sample clearly showed ancient DNA specificities. Thus, we concluded that the high molecular weight range of the D7S820 and D18S51 markers was responsible for the absence of genotypes for two PCR amplifications with the Profiler Plus kit. Indeed, it has been demonstrated that a negative correlation exists between the amplified sequence length and the PCR efficiency on ancient DNA templates (Pääbo et al, 1989).

The paternal lineage of the Pokrovsk subject seems to differ from the ones found in the modern population. Indeed, this individual shows a T allele at the TAT SNP locus, whereas 86% of modern Yakuts present the C allele (Pakendorf et al, 2002) which is the highest frequency in Siberia. Unfortunately, the absence of complete Y STR results prevented the determination of the haplotype of this individual and its comparison with modern Siberian population.

The HV1 sequence obtained for the Pokrovsk individual can be attributed to the Asian haplogroup D, found in the modern Yakut population at a frequency of 30% (Fedorova et al, 2000). Furthermore, haplogroup D is the second most common haplogroup found in modern Siberian populations with the highest frequencies found in Buryats and Sojots of the Baïkal region (Derenko et al, 2003). A common ancestral population or genetic drift could explain the wide distribution of this haplogroup among Siberian populations.

This sequence was compared with the data available in the Genbank database and in the database developed our laboratory. This haplotype matched with two Buryats (from the Baïkal area) (GenBank accession numbers: AB059982, AB059902); two West Siberians (AF214088, AF214087), two Mansis (from the Altaï) (Derbeneva et al, 2002); one Evenk (Starikovskaya et al, 2005); one ancient Yakut, two modern Yakuts and one female from the

Egyin gol necropolis (North Mongolia, near the Baïkal area) from the later part of the necropolis (II/III century AD) (Keyser-Tracqui et al., 2003). This mitotype is found neither in Koryaks, Chukchi, Itel'men or Youkaguirs, sometimes considered as “paleoasiatic” ethnic groups, nor in Central Asian populations.

The comparison with ancient and modern Yakuts demonstrates 1) the early participation to the Yakut gene pool of women belonging to tribes who lived on the Lena River during antiquity and 2) the continuity of this maternal lineage through the ages in the Yakut population but also in other Siberian ethnic groups. This hypothesis confirms the results obtained by Ricaut et al. (2005) indicating the transmission of ancestral mitochondrial sequences in Siberian populations.

The similarity of the mitotype of the Pokrovsk subject with Buryats and a skeleton from the Egyin Gol necropolis, located 2000 kilometers to the south, confirms the occurrence of ancient contacts between the Altaï/Baïkal region and oriental Siberia, prior to the end of the Xiong Nu period (III BC/IIAD). Some female ancestors of this hunter may originate from the First Empire of the Steppes, well known for its military expansion to the south (China) and to the west. However, the man of the Pokrovsk grave shows that these nomadic people may have also tried to explore the north by diffusion along the rivers. Indeed, the main rivers of Yakutia (Lena, Vilyuy and Aldan) were already used during the Late Sartan Time (Final Pleistocene) as dispersion roads during the human migrations through north-eastern Siberia (Vasil'ev et al, 2002). These settlers or their descendants adopted an arctic way of life after merging in the whole population. Furthermore, links with the Buryats are evidenced by the genetic proximity of these two populations (Pakendorf et al., 2003) as well as by anthropological record as the Western Buryats seem to be morphologically close to the Yakuts (Khintrinskaya et al. 2003).

As mentioned previously, the proximity between Yakuts and Evenks has been demonstrated in several population genetic studies (Pakendorf et al., 1999; Pakendorf et al., 2003; Derenko et al, 2002; Puzyrev et al, 2003). This genetic similarity between Evenks and Yakuts revealed by means of mtDNA analysis suggests that their gene pools were formed on the basis of very similar maternal lineages (Derenko et al., 2002). The ancient DNA analysis of the Pokrovsk individual confirms the dual origin of the Yakuts and brings new elements to this hypothesis in addition to archaeological and ethnological data. Indeed, this shared haplotype may indicate that an admixture between ethnic groups from the southern steppe and Tungusic reindeer herders occurred during the IIIrd century BC, before the arrival of the Kurykans or the Kulun-Atakh tribes.

The match of our sequence with two Mansis from the Ural Mountains and two western Siberians could be related to an extensive east/west gene flow (Starikovskaya et al, 2005) along the Ienisseï River. Considering the important frequency of Asian haplogroups present in the Mansi (Derbeneva et al, 2002), this similarity may however stem from the wide expansion of the nomadic tribes from the southern steppe westward in the Ural Mountains. Thus, the gene flow from Mongolia seems to have affected autochthonous populations from Oriental and Occidental Siberia during the Xiong Nu period since the 3rd century BC.

Our study confirms that archaeology and direct access to the gene pool of ancient autochthonous Siberian populations are privileged tools to clarify northern Asian prehistory and the ancestor-descendant relationships between ancient and modern populations. Indeed, the analysis of the Pokrovsk grave allowed us to corroborate the great influence of the Xiong Nu Empire over the Siberian populations and an early admixture between populations from the southern steppe and Central Siberia aborigines. The fact that the mtDNA sequence of the Pokrovsk individual is found in Mongols and Buryats suggests an early Mongolic expansion eastward from Mongolia to the Baïkal Lake before the migration in the XIVth and XVth centuries. Furthermore, this shared sequence indicates the occurrence of contacts between populations from Oriental Siberia and the nomadic peoples of the steppe, especially those of the Altaï/Baïkal area which were unified for the first time during the Xiong Nu period. Therefore, it seems that the migrations of the Kurykhans or the expansion of the Kulun Atakh culture did not constitute first contacts between these two geographical areas. Moreover, an admixture between the new settlers and Tungus autochthonous people during the IVth and IIIrd centuries BC could explain the genetic proximity observed nowadays between modern Evenks and Yakuts (Pakendorf et al. 2003). Further investigations on ancient samples will be necessary to conclude whether these relations were due to short expansions or to a real settlement of horse breeders before the Middle Age.

Literature cited

- Alekseev AN. 1996. Ancient Yakutia: The Iron Age and the Medieval Epoch (in Russian). Izdatel'stvo Instituta Arkheologii SO RAN, Novosibirsk.
- Anderson S, Bankier AT, Barrell BG, de Brujin MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Balzer MM. 1994. Yakut. In: Friedrich P, Diamond N, editors. *Encyclopedia of world cultures*, volume VI—Russia and Eurasia/China. New York: G.K. Hall & Co. p 404–407.
- Bradley R, Huges M, Diaz H. 2003. Climate in Medieval Time. *Science*. 302(5644):404-5.
- Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet*. Jan;60 (Pt 1):35-49.
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetti J. 1996. Geographic variation in human mitochondrial DNA control region : the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13 :1067-1077.
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetti J. 1998. Trading Genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 63:1824-1838.
- Comas D, Calafell F, Bendukidze L, Fanamas L, Bertranpetti J. 2000. Georgian and Kurd mtDNA Sequence Analysis Shows a Lack of Correlation between languages and Female Genetic Lineages. *Am J Phys Anth* .112:5–16.
- Crubézy E, Alekseev A N, Amory S, Keyser C, Ludes B. The Pokrovsk grave: early influence of the populations of the Altaï/Baïkal spatial complex in the peopling of Oirental Siberia. *Antiquity* (submitted).
- Derbeneva OA, Starikovskaya EB, Wallace DC, Subernik RI. 2002. Traces of early eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70:1009-1014.
- Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM, Nimaev DD, Zakharov AIA. 2000. Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of north Asia. *Hum Biol* 72(6):945-973.

- Derenko M. V., Malyarchuk B. A., Denisova G. A., Dambueva I. K., Kakpakov V. T., Dorzhu Ch. M., Luzina F. A., Lotosh E. A., Ondar U. N., Kaplina M. I. and Zakharov I. A. 2002. Molecular Genetic Differentiation of the Ethnic Populations of South and East Siberia Based on Mitochondrial DNA Polymorphism. *Russ J Genet*, Vol. 38, No. 10, pp. 1196–1202.
- Derenko MV, Malyarchuk BA, Dambueva IK, Zakharov IA. 2003. Structure and diversity of the mitochondrial gene pools of south Siberians. *Dokl Biol Sci*. Nov-Dec;393:557-61.
- Fedorova SA, Bermisheva MA, Villemans R, Maksimova NR, Khusnutdinova EK. 2000. Analysis of mitochondrial DNA lineages in Yakuts. *Mol Biol* 37:4 544-553.
- Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. 2001. Improved mtDNA sequence analysis of forensic remains using "mini-primer set" amplification strategy. *J Forensic Sci* 46:247-253.
- Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. 2000. An investigation of the rigor of interpretation rules for STRs derived from less than 100pg of DNA. *Forensic Sci Int*. 112: 17-40.
- Gogolev AI. 2000. *Istorija Jakutii* (History of Yakutia). 201p. (In Russian). Izdatel'stvo JaGU, Yakutsk.
- Ivanov P, Wadhams M, Roby R, Holland M, Weedn V, Parson T. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12 :417-420.
- Keyser-Tracqui C, Crubézy E, 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*. Aug;73(2):247-60
- Keyser-Tracqui C., Blandin P., Ricaut F.X., Petkovski E., Crubézy E. and Ludes B. 2004. Does the Tat polymorphism originate in northern Mongolia? International Congress Series, 1261 *Progress in Forensic Genetics* 10, pp. 325-327.
- Keyser-Tracqui C, Ludes B. 2005. Methods for the study of ancient DNA. *Methods Mol Biol*;297:253-64.
- Keyser-Tracqui C, Crubézy E, Pamzsav H, Varga T, Ludes B. 2006. Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. *Am J Phys Anthropol*. 131(2): 272-81.
- Khitrinskaya IY, Stepanov VA, Puzyrev VP, Spiridonova MG, Puzyrev KV, Maksimova NR, Nogovitsyna AN. 2003. Genetic Peculiarity of the Yakut Population as Inferred from Autosomal Loci. *Mol. biol.* 37 (2): 205-209.

- Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of new world founders. *Genetics* 142:1321-1334.
- Konstantinov IV. 1975. The origins of the Yakut people and their culture. Yakutia and her neighbors in Antiquity. *Publications of the Prilenskaya Archaeological expedition*. Yakutsk :Yakutskiy.
- Macaulay V, Richards M, Hickey E 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control region sequences and RFLPs. *Am J Hum Genet* 64:232-249.
- Mackay A, Ryvesa D, Battarbeea R, Flowera R, Jewsonb D. 2005. 1000 years of climate variability in central Asia: assessing the evidence using Lake Baikal (Russia) diatom assemblages and the application of a diatom-inferred model of snow cover on the lake. *Global and Planetary Change* 46: 281–297.
- P. Rioualc, M. Sturmd
- Okladnikov AP. 1955. The history of the Yakut ASSR. Volume 1 : Yakutia before its incorporation into the Russian state. Moscow : Izdatel'stvo Akademii Nauk SSSR.
- Orehov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankosky N. 1999. Mitochondrial DNA sequence diversity in Russians. *FEBS Lett* 445:197-201.
- Pääbo S, Higuchi RG, Wilson AC. 1989. Ancient DNA and the Polymerase Chain Reaction. *J Biol Chemistry*. 264; vol 17: 9709-9712.
- Pakendorf B, Spitsyn V, Rodewald A. 1999. Genetic structure of a Sakha population from Siberia and ethnic affinities. *Hum Biol*. 71(2):231-44.
- Pakendorf B, Morar B, Tarskaia LA, Kayser M, Soodyall H, Rodewald A, Stoneking M. 2002. Y-chromosomal evidence for a strong reduction in male population size of Yakuts. *Hum Genet* 110 :198-200.
- Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M.. 2003. Mitochondrial DNA evidence for admixed origins of Central Siberian Populations. *Am J Phys Anthropol* 120:211-224.
- Puzyrev VP, Stepanov VA, Golubenko MV, Puzyrev KV, Maximova NR, Kharkov VN, Spiridonova MG, and Nogovitsina AN. 2003. mtDNA and Y-Chromosome Lineages in the Yakut Population. 2003. *Russ J Genet* Vol. 39, No. 7 pp. 816–822.
- Ricaut F-X, Fedoseeva A, Keyser-Tracqui C, Crubézy E, Ludes B. 2005. Ancient DNA analysis of human Neolithic remains found in North-eastern Siberia. *Am J Phys Anthropol* 126: 458-462.

- Sajantila A, Lahermo P, Anttilanen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjærg L, Gedde-Dahl T, Issel-Tarver L, DiRienzo A, Pääbo S. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Gen Res* 5:42-52.
- Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC. 1999. Mitochondrial DNA variation in koryaks and itel'men: population replacement in the okhotsk sea-bering sea region during the neolithic. *Am J Phys Anthropol* 108:1-39.
- Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH.. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet*. 1993 Sep; 53(3):549-62.
- Sieroszewski W. 1993. Jakuty. Opyt etnografi eskogo issledovaniya (The Yakuts: Experiment of an Ethnological Research). Moscow, Assotsatsiiia 'Rossiiskaia Polit. Entsiklopedia' [eHRAF], 736 p.
- Simonson, J. H. and G. F. Shields. 1997. Genetic evidence for a Beringian bottleneck in northern Native Americans during the late Pleistocene. NCBI published data.
- Starikovskaya EB, Subernik RI, Schurr TG, Kogelnik AM, Wallace DC. 1998. mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the new world. *Am J Hum Genet* 63:1473-1491.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, 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*. Jan; 69:67-89.
- Torroni A, Subernik RI, Schurr TG, Starikovskaya YB, Cabell MA, Crawford MH, Comuzzie AG, Wallace DC. 1993. mtDNA variation of aboriginal siberians reveals distinct genetic affinities with native Americans. *Am J Hum Genet* 53:591-608.
- Vasil'ev SA, Kuzmin Lyubov Y V, Orlova A, Dementiev VN. 2002. Radiocarbon-based chronology of the Palaeolithic in Siberia and its relevance to the peopling of the New World. *Radiocarbon*, Vol 44, Nr 2 p 503–530.
- Voevoda MI., Avksentyuk AV., Ivanova AV., Astakhova TI., Babenko VN., Kurilovich SA., Duffy LK., Segal B. and Shields GF. 1994. Molecular genetic studies in the population of native inhabitants of Chukchee Peninsula. Analysis of polymorphism of mitochondrial

- DNA and of genes controlling alcohol metabolizing enzymes. *Sibirskii Ekolog. Z.* I (2), 139-151 (1994)
- Ward RH, Frazier BL, Dew-Jager K, Pääbo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88:8720-8724.
- Ward RH, Redd A, Valencia D, Frazier B, Pääbo S. 1993. Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90:10663-10667.
- Whitaker JP, Cotton EA, Gill P. 2001. A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci Int.* 1;123(2-3):215-23.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635-651.
- Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP. 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118:63-76.

Tables

Population/Region	N	Reference
Central Asia		
Altai	17	Shields et al. (1993)
Buryats	296	Derenko et al. (2000); Derenko et al. (2003); Pakendorf et al. (2003)
Kazakhs	85	Comas et al. (1998); Yao et al. (2002b)
Kirghiz	92	Comas et al. (1998)
Mongols	118	Kolman et al (1996) ; Yao et al. (2002b)
Uzbeks	20	Comas et al. (2000)
Yupiks	24	Simonson et al. (1997)
Siberia		
Evenks	71	Torroni et al. (1993), Starikovskaya et al. (2005)
Udegey	3	Torroni et al. (1993)
Uighurs	100	Yao et al. (2002b); Comas et al. (1998)
Tuvas	36	Derenko et al. (2000)
West Siberia	38	Voevoda et al. data published on GenBank
Yakuts	161	Keyser-Tracqui et al. submitted; Pakendorf et al. (2003)
Beringia		
Chukchi	135	Voevoda et al. (1994); Starikovskaya et al. (1998)
Eskimos	126	Starikovskaya et al. (1998); Voevoda et al. NCBI published data; Shields et al. (1993)
Itel'men	47	Schurr et al. (1999)
Koryaks	147	Schurr et al. (1999)
Europe/Middle-East		
Bulgarians	30	Calafell et al. (1996)
Estonia	28	Sajantila et al. (1995)
Finn-Volga	34	Sajantila et al. (1995)
Georgians	45	Comas et al. (1996)
Karelia	83	Sajantila et al. (1995)
Mansi	98	Derbeneva et al. (2002)
Saami	135	Sajantila et al. (1995)
Russians	103	Orekov et al. (1999)
Iranians	20	Comas et al. (2000)
Turks	72	Calafell et al. (1996) ; Comas et al. (2000)
Kurds	29	Comas et al. (1996)
East-Asia		
Han Chinese	183	Yao et al. (2002a)
Non Han Chinese	253	Yao et al. (2002b)
Sali	31	Yao et al. (2002b)
Tibetan	39	Yao et al. (2002b)
North America		
Bella Coola	40	Ward et al. (1993)
Haïda	41	Ward et al. (1993)
Nuu Chah Nutlh	63	Ward et al. (1991)

Table 1. Populations and references of the HV1 sequences database used for the comparative analysis.

Name	Extractions	Allele(s) at marker									
		Amel*	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA
Pokrovsk	1	XY	.	.	29/29	14/14	.	.	.	22/23	19/19
		XY	9/10/12	13/16/17/18	29/29	14/	11/12	10/13	13/14	22/23	16/17/18
	2	XY	9/12	13/17	29/29	14/15	11/12	19/13	13/13	22/23	17/18
		XY	9/12	13/17	29/29	14/15	11/12	.	13/13	22/23	17/18
	3	XY	9/12	.	29/29	14/15	11/12	.	13/13	22/23	17/18
		XY	9/12	13/17	29/29	14/15	11/12	10/13	13/13	22/23	17/18
Consensus sequence		XY	9/12	13/17	29/29	14/15	11/12	10/13	13/13	22/23	17/18

Table 2. Summary of the allelic profiles obtained with Profiler Plus kit from the Pokrovsk subject. * Amelogenin locus.

Name	Allele(s) at marker									
	Amel	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA
Team 1	XX	11/14	13/15	27/31	16/17	12/12	7/11	13/13	20/23	17/19
Team 2	XY	12	14/15	27/31,2	14/18	9/13	8/10	8/15	20	17/19
Team 3	XY	12/13	14/14	31/32,2	16/17	11/12	9/12	13/13	22/25	17/19
Team 4	XY	19/19	16/21	29/30	8/11	11/11	8/11	13/16	21/24	16/17
Team 5	XX	10/11	13/14	29/30	17/17	12/13	9/11	10/13	21/23	17/18

Table 3. STR obtained for the Profiler Plus kit of all persons involved in processing samples.

Name	Extraction	Allele(s) at marker										
		DYS 389I	DYS 389II	DYS 391	DYS 439	DYS 19	DYS 392	DYS 437	DYS 438	DYS 385	DYS 390	DYS 393
Pokrovsk	1	14	.	9	.	.	.	14	.	.	.	13
	2	14	.	10	10	.	.	.
	3

Table 4. Summary of the results obtained with the PowerPlex Y kit for the Pokrovsk specimen.

Name	Allele(s) at marker										
	DYS 389I	DYS 389II	DYS 391	DYS 439	DYS 19	DYS 392	DYS 437	DYS 438	DYS 385	DYS 390	DYS 393
Team 2	12	28	10	11	14	11	.	10	13/16	22	14
Team 3	12	28	11	.	14	13	.	.	11/14	24	13
Team 5	14	32	11	10	14	16	.	11	11/13	23	14

Table 5. Y chromosome STR haplotypes of all the persons involved in processing sample.

Sample	Extraction	Polymorphic positions*												
		16039	16093	16126	16223	16224	16249	16263	16292	16294	16296	16311	16362	16368
CRS		G	T	T	C	T	T	T	C	C	C	T	T	T
Pokrovsk	1	.	.	.	T	C	C	
	2	.	.	.	T	C	C	

Table 6. Polymorphic positions of the mitochondrial DNA sequence of the Pokrovsk subject.

*numbered according to Cambridge Reference Sequence (Anderson et al., 1981).

Sample	Polymorphic positions*															
	16039	16051	16093	16126	16162	16223	16224	16232	16249	16263	16292	16294	16296	16311	16362	16368
CRS	G	A	T	T	A	C	T	C	T	T	C	C	C	T	T	T
Team 1	A	.	.	C	.	.	.	C	.	T	T	T	.	.	.	
Team 2	.	.	C	.	.	.	C	C	.	.	
Team 3	C	
Team 4	.	.	C	.	.	T	.	T	C	.	
Team 5	.	G	.	.	G	

Table 7. Polymorphic positions of the mitochondrial DNA sequence of all the persons involved in processing sample.

* numbered according to Cambridge Reference Sequence (Anderson et al., 1981).

Clone	Polymorphic positions*									
	16033	16067	16071	16220	16223	16234	16273	16344	16362	16368
B5	T	.	.	.	C	C
B7	T	.	.	.	C	C
C2	.	.	.	G	T	.	.	.	C	C
C10	T	.	.	.	C	C
D11	T	.	.	.	C	C
E2	T	.	.	.	C	C
E5	T	.	.	.	C	C
F1	T	.	.	.	C	C
F2	T	.	.	.	C	C
F5	T	.	.	.	C	C
F9	.	.	T	.	T	T	.	.	C	C
G3	T	.	.	T	C	C
G5	T	.	.	.	C	C
G8	A	.	.	.	T	.	A	.	C	C
G9	.	T	.	.	T	.	.	.	C	C
H5	T	.	.	.	C	C

Table 8. Polymorphic positions of the sequenced clones derived from the HVI PCR product of the bone extract.

* numbered according to Cambridge Reference Sequence (Anderson et al., 1981).



Multi-substrata analysis on Siberian mummies: A different way for validation in ancient DNA studies?

S. Amory ^{a,b,*}, C. Keyser-Tracqui ^{a,b}, E. Crubézy ^b, B. Ludes ^{a,b}

^a Institute of Legal Medicine, Strasbourg Cedex, France

^b Laboratory of Anthropology, Paul Sabatier University, CNRS, UMR 8555, Toulouse, France

Abstract. Ancient DNA results are always submitted to caution due to the technical difficulties induced by the minute amounts, the degraded nature of the template and the high risk of contamination. A list of criteria of validation has been published as a guideline for ancient DNA researchers [1] A. Cooper, H.N. Poinar, Ancient DNA: do it right or not at all, *Science* 289 (5482) (2000) 1139.]. In addition to these criteria, the analysis of different substrates, i.e. bone, teeth and hairs of the same individual, could be another way to ensure the reliability of the results. This study presents the data obtained on bones, molar teeth and hairs from two Siberian specimens dated from the 18th century. These two subjects excavated from frozen graves were mummified. Their exceptional state of preservation allowed us to test the amplification of autosomal and Y chromosomal STRs and of the HV1 region on the three types of substrate. This approach allowed the identification of artefacts on STR profiles, common when working with Low Copy Number DNA. Indeed, the comparison of the profiles obtained from bones and teeth highlights allelic dropouts and spurious alleles for the bone samples. The possibility to compare results from different substrates, in spite of the limited number of possible cases, represents another, and interesting, criterion to confirm the authenticity of ancient DNA results. © 2006 Published by Elsevier B.V.

Keywords: Ancient DNA; Bone; Hair; Tooth; Validation criterion

1. Introduction

Due to the technical difficulties induced by the minute amounts, the degraded nature of the templates and the high risk of contamination, the validation of the results in the

* Corresponding author. Postal address: 11 rue Humann, 67085 Strasbourg Cedex, France. Tel.: +33 390 243348; fax: +33 390 243362.

E-mail address: sylvain.amory@iml-ulp.u-strabg.fr (S. Amory).

field of ancient DNA became of major interest. The aim of this study was to determine in which way a multiple substrates analysis could be considered as a new criterion of validation. The discovery of two mummies in northeastern Siberia gave us the opportunity to study nuclear and mitochondrial DNA extracted from bones, molar teeth and hair shafts.

2. Materials and method

2.1. Materials

The two individuals were discovered in Yakutia, an autonomous Republic of the Russian Federation, located in the Northeast of Siberia. The extremely harsh and dry climate combined with inhumations directly on permafrost induced good DNA conservation.

The first grave called Munur Urekh (MU) dated from the late 17th century considering the archaeological material associated with the body. The coffin was filled with 5 cm ice and the body was partially frozen.

The Shamanic Tree (ST) grave was dated from the same period. This multiple grave contained four mummified bodies: two females and two males. The samples were taken from the best preserved specimen.

For the two subjects we collected a femur, some hairs and teeth without decay or slit.

2.2. Methods

Two grams of bone powder were extracted by a classic phenol chloroform protocol according to Keyser-Tracqui and Ludes [2].

Tooth samples were deeply decontaminated with DNA away, MilliQ water, absolute alcohol and UV light. The cleaned teeth were pulverised in liquid nitrogen using a 6800 Freezer Mill. DNA extraction was performed with the same protocol as bones.

Sections of 3 cm taken from the root side of 5 hair shafts were extracted as described by Gilbert et al. [3]. For each sample parallel extractions with or without cleaning were tested.

Nuclear DNA was quantified by Real Time PCR with the Quantifiler™ Human DNA Quantification Kit on an ABI PRISM™ 7000 SDS (Applied Biosystems).

Autosomal and Y chromosome STRs were analysed respectively with the AmpFISTR® Profiler Plus™ Kit (Applied Biosystems) and PowerPlex® Y System (Promega). BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) was used for the sequencing of the HV1 region.

3. Results

As DNA concentrations were all inferior to $0.5 \text{ ng } \mu\text{l}^{-1}$ and most of the extracts yielded DNA quantity lower than 100 pg, 34 cycles PCR amplifications were performed [4].

For autosomal STR analyses, tooth extracts gave the best results yielding complete profiles without amplification artefacts (Fig. 1). Moreover, beyond all expectations, we obtained better results with hair than with bone extracts. Indeed, allelic dropouts and spurious alleles were more frequent on the 2 different bone extractions. We have observed the same kind of results for the MU1 individual as well as for Y chromosome STRs.

Sample	Substratum	Allele(s) at marker									
		¹ Amel	D13	D18	D21	D3	D5	D7	D8	FGA	vWA
ST1	Tooth	XY	10/11	13/15	29/29	15/16	10/12	13/16	13/16	21/25	17/19
		XY	10/11	13/15	29/29	15/16	10/12	13/16	13/16	21/25	17/19
	Hair nW ²	XY	10/11	.	29/29	15/16	10/12	13/16	13/16	21/25	17/19
	Hair W ³	XY	10/11	15/15	29/29	15/16	10/12	13/16	13/16	25/25	17/19
	Bone (⁴ Ext ¹)	XY	11/11	15/15	29/29	15/16	10/12	13/13	13/13	21/24/25	17/18/19
	Bone (⁴ Ext ²)	XY	.	.	.	15/16	.	16/16	16/16	.	.
Consensus		XY	10/11	13/15	29/29	15/16	10/12	13/16	13/16	21/25	17/19

Fig. 1. Results of autosomal STR amplification with the Profiler Plus™ Kit (Applied Biosystems) for Shamanic Tree specimen. Allelic dropouts and spurious alleles are shown in bold font (¹amelogenin, ²not washed, ³washed, ⁴extraction).

Mitochondrial DNA sequencing revealed contrasted results. For the ST1 subject, the HV1 sequences obtained for the different substrates presented the same polymorphic positions. But the MU1 sequences highlight well known characteristics for the different substrates. For MU1, heteroplasmies were found on hair and tooth DNA extracts for the 16093 position, known as a mutation hotspot [5], whereas for the bone sample a T/C transition was observed.

4. Conclusion

As far as we know, this work illustrates for the first time the possibility to use hairs as a potential source of nuclear DNA in paleo-anthropological research. Moreover it demonstrates that this approach could represent a new criterion for the authentication of STRs profiles. Further analyses will be realised by this method in order to validate these preliminary results.

Considering mtDNA results, heteroplasmies remain difficult to interpret. This is probably due to the specificities of each substrate and to the sensitivity of the sequencing method. But other techniques, such as PNA probes, could be used to solve this problem.

For anthropological studies it seems obvious that the potential number of cases will be limited due to the rare number of cases with availability of the three types of substrates. Nevertheless this approach could be valuable for forensic identifications of degraded remains.

References

- [1] A. Cooper, H.N. Poinar, Ancient DNA: do it right or not at all, *Science* 289 (5482) (2000) 1139.
- [2] C. Keyser-Tracqui, B. Ludes, Methods for the study of ancient DNA, *Methods Mol. Biol.* 297 (2005) 253–264.
- [3] M.T. Gilbert, A.S. Wilson, M. Bunce, et al., Ancient mitochondrial DNA from hairs, *Curr. Biol.* 14 (2004) 12.
- [4] P. Gill, et al., An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Sci. Int.* 112 (2000) 17–40.
- [5] M. Stoneking, Hypervariable sites in the mtDNA control region are mutational hotspots, *Am. J. Hum. Genet.* 67 (2000) 1029–1032.