

Table des matières

Liste des figures	7
Liste des tableaux	8
I INTRODUCTION	9
I.1 Introduction générale	11
I.2 La paléogénétique	13
I.2.1 Historique	13
I.2.2 L'ADN ancien, une molécule fragile	13
I.2.3 Critères d'authenticité	22
I.2.4 Champs d'applications des études paléogénétiques	23
I.2.5 Evolution de la discipline : d'un petit fragment au génome complet	26
I.3 Contexte général de l'étude	29
I.3.1 Contexte géographique	29
I.3.2 Peuplement de la région	30
I.3.3 Le peuplement actuel	35
I.4 L'apport des premières études paléogénétiques à la compréhension du peuplement du sud de la Sibérie et de l'Altai	36
I.4.1 Les données mitochondriales	36
I.4.2 Les données du chromosome Y	37
I.4.3 Les données autosomales	38
I.5 Problématique et objectifs de l'étude	39
II MATÉRIEL ET MÉTHODES	41
II.1 Prélèvements biologiques anciens	43
II.1.1 Traitement des échantillons	47
II.1.2 Les analyses génétiques réalisées	49

II.1.3	Analyses des données génétiques	67
II.1.4	Précautions prises contre les contaminations	68
III	RÉSULTATS	71
III.1	Résultats des analyses génétiques	73
III.1.1	Taux de réussite des différentes analyses génétiques menées	73
III.1.2	Résultats de l'analyse des STR autosomaux et de la détermination du sexe	74
III.1.3	Les données mitochondriales	77
III.1.4	Les données du chromosome Y	86
III.1.5	Résultats du typage des SNP autosomaux	93
III.1.6	Synthèse des résultats obtenus pour chaque échantillon	97
III.1.7	Etude des liens de parentés	99
III.2	Authenticité des résultats	101
IV	DISCUSSION	103
IV.1	Apport de ce travail à la compréhension de la dynamique de peuplement du sud de la Sibérie et de l'Altaï à l'âge du Bronze	106
IV.1.1	Evolution du pool génique	106
IV.1.2	Dynamique de peuplement de la région, premières hypothèses	117
IV.2	Discussion sur la stratégie expérimentale employée et perspectives	120
V	CONCLUSION	125
VI	BIBLIOGRAPHIE	131
VII	ANNEXES	163
VIII	PUBLICATIONS	181

Liste des figures

Figure 1 : Les dégradations post-mortem de la molécule d'ADN	14
Figure 2 : Les principaux sites de lésions au sein de la molécule d'ADN.....	16
Figure 3 : Motif de dégradations de l'ADN obtenu à partir du séquençage de l'ADN d'un cheval du pléistocène	18
Figure 4 : Evolution des études d'ADN ancien.....	28
Figure 5 : Carte géographique d'une partie de l'Eurasie	30
Figure 6 : Répartition chronologique et géographique des cultures mentionnées dans ce travail entre le IV ^{ème} et I ^{er} millénaire avant J.-C.....	33
Figure 7 : Localisation des sites pour lesquels des échantillons ont été analysés.....	43
Figure 8 : Tombe de l'individu TA10 du site de Tsagaan Asga	46
Figure 9 : Répartition chronologique et géographique des prélèvements analysés	46
Figure 10 : Représentation schématique de l'ADN mitochondrial humain	53
Figure 11 : SNP de la région codante de l'ADN mitochondrial	56
Figure 12: Arbre phylogénétique présentant les SNP du chromosome Y analysés	58
Figure 13 : Principe de la spectrométrie de masse MALDI-TOF	60
Figure 14: Principe de la méthode iPLEX Gold de Sequenom.....	62
Figure 15 : Fréquence des haplotypes mitochondriaux au sein des populations eurasiennes actuelles.....	83
Figure 16 : <i>Median Joining</i> Network réalisé à partir des haplotypes R1b.....	92
Figure 17 : Graphiques triangulaires générés à partir du logiciel STRUCTURE représentant l'affinité génétique des individus anciens avec les 3 populations de référence	97
Figure 18 : Corrélation inverse entre la taille des fragments amplifiés et l'intensité du signal	101
Figure 19 : Répartition géographique et temporelle des haplogroupes mitochondriaux et Y retrouvés dans les échantillons étudiés.....	116

Liste des tableaux

Tableau 1 : Liste des prélèvements disponibles pour ce travail 1/2.....	44
Tableau 2 : Liste des SNP autosomaux sélectionnés	63
Tableau 3: Fréquences alléliques et génotypiques des AIM sélectionnés	66
Tableau 4 : Taux de réussite des analyses génétiques réalisées sur les échantillons d'ADN ancien	73
Tableau 5 : profils génétiques consensus obtenus à partir des échantillons d'ADN ancien (1/2)	75
Tableau 6 : Haplotypes mitochondriaux partagés au sein de l'ensemble des sujets anciens étudiés.....	77
Tableau 7 : Haplotypes consensus et haplogroupes mitochondriaux (1/2).....	78
Tableau 8 : Distribution des haplogroupes obtenus au sein de chaque groupe.....	80
Tableau 9 : Haplotypes mitochondriaux partagés entre les échantillons analysés et les données d'ADN ancien publiées dans la littérature	80
Tableau 10 : Haplotypes Y consensus.....	87
Tableau 11 : Résultats du typage des SNP de la région non recombinante du chromosome Y88	
Tableau 12 : Haplotypes Y partagés entre les individus anciens étudiés et les populations modernes	91
Tableau 13 : Bilan par groupe de la prédiction des caractères pigmentaires pour les individus anciens.....	96
Tableau 14 : Tableau récapitulatif des données obtenues pour chaque échantillon ancien	98
Tableau 15 : Echantillons présentant les mêmes lignées maternelles ou paternelles.....	99

I INTRODUCTION

I.1 Introduction générale

Au milieu des années 80, soit il y a tout juste trente ans, une nouvelle discipline émergeait et allait connaître une révolution importante grâce aux développements des techniques de biologie moléculaire : la paléogénétique. Cette discipline s'intéresse à l'étude d'un ADN dit « ancien » (ADNa), c'est-à-dire un ADN extrait de restes humains, d'animaux, de pathogènes ou encore de végétaux du passé. Cela, dans le but de répondre à de nombreuses problématiques telles que l'évolution des génomes au cours du temps, la dispersion de l'Homme sur la planète ou encore son adaptation à l'environnement. Les avancées technologiques sont telles que nous pouvons aujourd'hui parler de paléogénomique. En effet, longtemps limitées à l'étude de courts fragments du génome, les analyses d'ADNa permettent aujourd'hui, dans certaines conditions, d'analyser des génomes anciens totaux. Si les premières études ont pu être remises en question, les critères stricts de travail mis en place par la communauté scientifique ainsi que les progrès techniques réalisés dans ce domaine permettent aujourd'hui d'aborder plus fidèlement de nombreuses problématiques.

Parmi les divers champs d'application des études d'ADN ancien, plusieurs travaux se sont intéressés à retracer les mouvements migratoires de populations du passé. En effet, si les études génétiques menées sur les populations actuelles peuvent apporter des informations concernant l'évolution récente des populations ou proposer des schémas de peuplement, seul l'accès direct aux génomes des populations du passé permettra d'étendre ces connaissances dans le passé et de s'affranchir des événements postérieurs qui ont pu modifier fortement le pool génique des populations anciennes.

Dans ce travail de thèse nous allons plus particulièrement nous intéresser à la dynamique de peuplement du sud de la Sibérie et du massif de l'Altaï durant l'âge du Bronze. De par sa situation géographique le long du couloir migratoire des steppes eurasiennes et à la rencontre des mondes européens et asiatiques, cette région a connu au cours du temps une succession et un métissage de populations issues d'Europe ou d'Asie. Pourquoi s'intéresser à l'âge du Bronze, c'est-à-dire la période s'étendant du milieu du IV^{ème} au I^{er} millénaire avant J.-C ? Dans cette région, de nombreux vestiges archéologiques couvrant cette période ont été découverts qui, d'après les données archéo- et anthropologiques, semblent correspondre à différentes cultures successives ou contemporaines, aux origines parfois diverses. Certaines d'entre elles ont notamment été associées à des mouvements migratoires d'Ouest en Est rattachés aux populations indo-européennes.

Ainsi l'objectif général de ce travail a été de caractériser génétiquement ces populations de l'âge du Bronze, d'étudier les relations génétiques entre elles mais aussi, dans un contexte plus large, d'essayer de déterminer leur origine biogéographique afin d'apporter de nouveaux éléments pour la compréhension des événements migratoires complexes survenus dans cette région.

La première partie de ce mémoire sera donc consacrée à la présentation de la discipline, ses champs d'applications, son évolution mais aussi les difficultés liées à l'étude d'un ADN dégradé. Ensuite nous aborderons plus en détail le contexte historico-culturel de ce travail, ainsi que l'apport des premières études de paléogénétique. Dans la deuxième partie, le matériel biologique utilisé ainsi que les méthodes d'analyses employées seront présentés. Les résultats obtenus pour les différents sites seront ensuite exposés, puis ils seront discutés au regard des données actuelles afin de dégager l'apport de ces analyses à la compréhension de la dynamique de peuplement de cette région. Enfin, des perspectives seront proposées à ce travail notamment au vu de l'évolution actuelle de la discipline.

I.2 La paléogénétique

I.2.1 Historique

Les premières études de paléogénétique sont apparues dès le milieu des années 80 avec la publication de séquences mitochondriales d'un quagga taxidermisé de 150 ans, espèce éteinte du genre *Equus* (Higuchi et al., 1984), rapidement suivie par la publication de la première séquence humaine provenant d'une momie égyptienne datée du V^{ème} siècle avant J.-C (Pääbo, 1985). Ces résultats amenèrent ainsi la preuve que l'ADN pouvait survivre au sein des organismes, même plusieurs années après leur mort, lançant ainsi la discipline. L'invention quasi simultanément de la PCR (*Polymerase Chain Reaction*) (Mullis and Faloona, 1987; Saiki et al., 1985) permettant d'amplifier en plusieurs centaines de milliers de copies une séquence d'ADN, se présenta comme un outil de choix et ne pouvait mieux tomber pour les paléogénéticiens. Raisonnable au début, les études ont voulu assez rapidement défier la limite de conservation de l'ADN et se sont dirigées vers des échantillons vieux de plusieurs milliers voire millions d'années, s'intéressant aux spécimens contenus dans l'ambre (Cano et al., 1993; DeSalle et al., 1992) et même aux dinosaures (Woodward et al., 1994). Un grand pas en arrière a été fait quelques années plus tard lorsque de nombreux résultats se sont avérés être des contaminations. La PCR, par sa sensibilité, avait donné accès aux informations du passé mais son plus grand avantage s'avéra aussi être le pire ennemi des études d'ADN ancien : l'amplification d'ADN exogène même présent en très faible quantité ! La remise en questions des résultats publiés a permis de mettre en évidence les difficultés inhérentes aux études d'ADN ancien. La meilleure compréhension des mécanismes de dégradation et de conservation de l'ADN, la mise en place de « bonnes pratiques », et les nombreux progrès méthodologiques réalisés ont finalement permis aux études d'ADN ancien, quelques années plus tard, de tenir leurs promesses.

I.2.2 L'ADN ancien, une molécule fragile

I.2.2.1 Les processus de dégradations

Dans une cellule vivante, les dommages subis par la molécule d'ADN peuvent être réparés grâce à la machinerie cellulaire qui permet de maintenir l'intégrité du génome. Lors de la

mort cellulaire, ces mécanismes de réparation cessent, exposant ainsi le génome à de nombreux facteurs pouvant l'endommager. Post-mortem, l'ADN est donc soumis à toute une diversité de processus de dégradation aboutissant dans la plupart des cas à l'obtention d'un ADN fragmenté, chimiquement modifié et présent en faible quantité (Lindahl, 1993) (Figure 1).

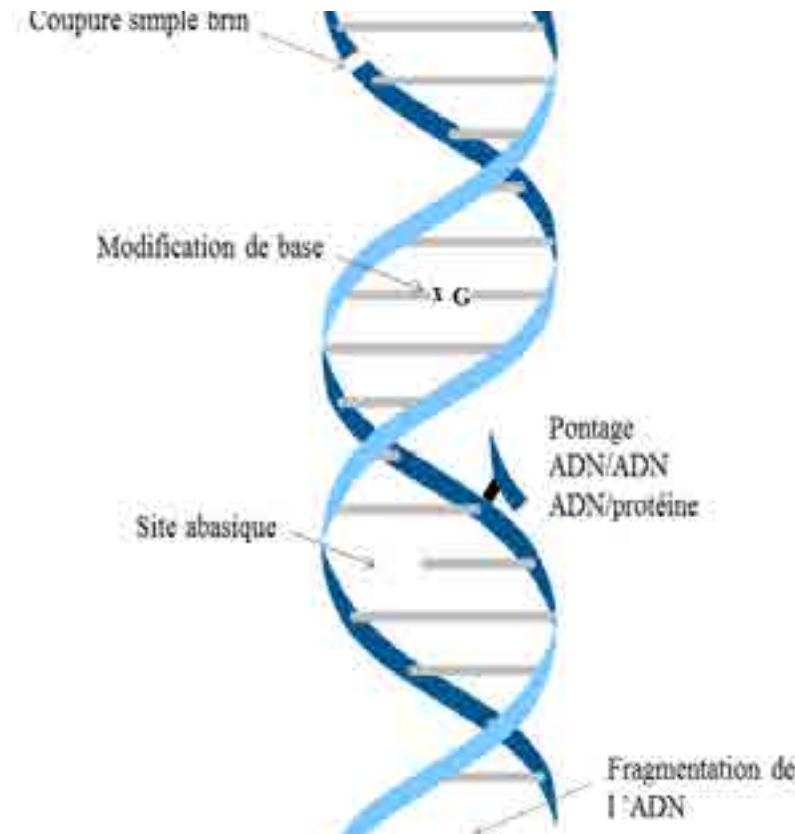


Figure 1 : Les dégradations post-mortem de la molécule d'ADN

I.2.2.1.1 La fragmentation

Plusieurs facteurs sont à l'origine de la fragmentation massive de l'ADN post mortem, aboutissant très rapidement à l'obtention de fragments de petites tailles (100pb) (Sawyer et al., 2012). Tout d'abord, des facteurs enzymatiques. Lors de la mort cellulaire, les **nucléases** qui étaient séquestrées dans la cellule vont être libérées et avoir accès à l'ADN, pouvant ainsi rompre les liaisons phosphodiester entre les nucléotides. Lors de la décomposition du corps, l'activité enzymatique provenant des micro-organismes saprophages agira également sur la fragmentation des molécules d'ADN. Ces processus enzymatiques, vont être complétés par

des phénomènes chimiques principalement sous la forme de réactions hydrolytiques et oxydatives (Figure 2).

L'**hydrolyse** va provoquer des ruptures simple-brin en s'attaquant aux liaisons phosphodiester (entre 2 nucléotides), mais également provoquer la formation de sites abasiques par la rupture des liaisons N-glycosidiques (liaison entre la base azotée et le désoxyribose). Ces sites abasiques vont fragiliser la structure de la molécule d'ADN et favoriser sa fragmentation.

Les réactions oxydatives agiront quant à elles au niveau des liaisons chimiques des résidus ribosiques. Les causes de ces dommages oxydatifs sont entre autres les espèces réactives de l'oxygène et les radicaux libres, tels que le radical hydroxyle, le radical peroxyde et le peroxyde d'hydrogène (Pääbo et al., 2004).

Enfin, d'autres facteurs, tels que les **rayonnements ionisants**, peuvent fragiliser la molécule par une action directe ou indirecte. En effet, ils peuvent interagir avec des molécules d'eau, provoquant la création de radicaux libres qui peuvent dégrader la molécule d'ADN.

1.2.2.1.2 Les modifications de bases

Outre la fragmentation, la molécule d'ADN va subir des dommages provoquant des modifications de la nature chimique des bases (Pääbo, 1989). Parmi les causes de ces lésions, on retrouvera de nouveau les phénomènes d'hydrolyse et d'oxydation (Figure 2)

L'**hydrolyse** va entraîner une perte de groupement amine au niveau des bases azotées induisant la transformation de cytosine en uracile, et de l'adénine en hypoxanthine, un analogue de la guanine. Cela va provoquer lors de l'amplification par PCR l'incorporation de bases incorrectes sous la forme de substitutions $C \rightarrow T$ et $A \rightarrow G$.

Des **conditions oxydantes** provoqueront quant à elles la formation de 8-hydroxyguanine, qui a la particularité de s'hybrider préférentiellement avec une adénine plutôt qu'avec une cytosine, créant ainsi des transversions $C \rightarrow A$ (Lamers et al., 2009).

Ces phénomènes sont particulièrement problématiques dans les études d'ADN ancien puisqu'ils ne vont pas forcément entraver la réplication de l'ADN lors de la PCR mais modifier l'information contenue dans la séquence primaire. Ils vont ainsi aboutir à

l'incorporation de bases erronées et donc à de possibles erreurs lors de l'analyse et de l'interprétation des résultats (Ho et al., 2007) (Figure 2).

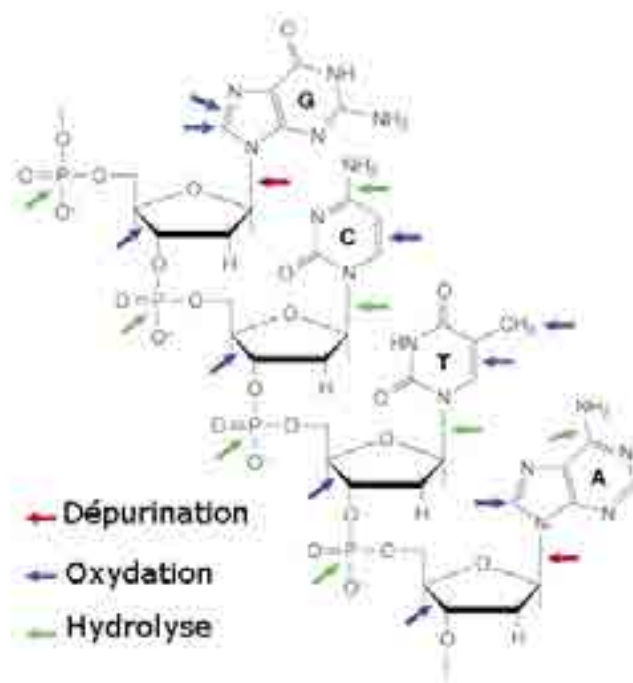


Figure 2 : Les principaux sites de lésions au sein de la molécule d'ADN
(Figure adaptée de (Hofreiter et al., 2001))

I.2.2.1.3 Les lésions bloquantes

La taille des fragments d'ADN amplifiés peut également être limitée par l'apparition de lésions bloquantes, empêchant le passage de la polymérase le long du brin d'ADN et donc son amplification. Plusieurs facteurs peuvent empêcher le passage de la *Taq* polymérase, que ce soit des modifications de bases telle que la formation d'hydantoïnes (produit d'oxydation des pyrimidines), mais aussi l'apparition de pontages intra-chaîne, inter-chaînes ou entre la molécule d'ADN et une protéine. Ces liaisons peuvent être dues à différents phénomènes tels que l'alkylation ou encore la formation de complexes de Maillard, provoqués par la liaison entre les sucres de la molécule d'ADN et les groupements amines des acides nucléiques ou des protéines. Décrit comme étant l'un des facteurs les plus limitant dans les études d'ADN ancien (Hansen et al., 2006), ces lésions bloquantes n'affecteraient pas plus de 40% des molécules d'ADNa (Heyn et al., 2010). Un traitement au bromure de *N*-phenacylthiazolium

(*N*-PTB), qui cliverait les produits de la réaction de Maillard, a été proposé afin d'augmenter la quantité d'ADN extrait (Poinar et al., 1998).

I.2.2.1.4 L'apport des séquenceurs de nouvelles générations à la compréhension des mécanismes de dégradation.

Les séquenceurs de nouvelles générations (ou NGS pour Next Generation Sequencing), de par leur principe, c'est-à-dire un séquençage massif en parallèle, ont permis d'apporter de nouveaux éléments quant au profil de dégradation de l'ADN ancien. Aujourd'hui, plusieurs génomes anciens entiers (mitochondriaux et nucléaires) ont été séquencés, générant ainsi une quantité importante de données obtenues à partir de différentes plateformes de séquençage et à partir de prélèvements provenant de différents environnements et périodes. Outre le défi technique et l'apport de connaissances sur le sujet ancien étudié, les résultats obtenus ont permis d'en savoir un peu plus sur les phénomènes de dégradation. Ainsi, ont pu être mesurés directement : les taux de dépurination, de fragmentation et de désamination des cytosines. Au vu des résultats, un profil de dégradation similaire semble se dégager de toutes les études (Figure 3) dont les éléments majeurs sont : (i) les fragments séquencés débutent pour la plupart après une guanine, faisant ainsi de la dépurination (Figure 2) l'une des causes majeures du processus de fragmentation (Briggs et al., 2007). Cela peut s'expliquer par le fait que la guanine présente une structure nécessitant moins d'énergie pour cliver sa liaison N-glycosidique par rapport à l'adénine (Overballe-Petersen et al., 2012) ;(ii) le phénomène de désamination de la cytosine en uracile est augmenté par rapport aux autres substitutions et est majoritairement présent aux extrémités 5' et 3' des fragments d'ADN lus, dû à la présence d'extrémités simple-brin (Briggs et al., 2007; Orlando et al., 2011). Ce motif de dégradation peut être ainsi utilisé parmi d'autres critères pour assurer le caractère « ancien » des séquences amplifiées sur ces plateformes.

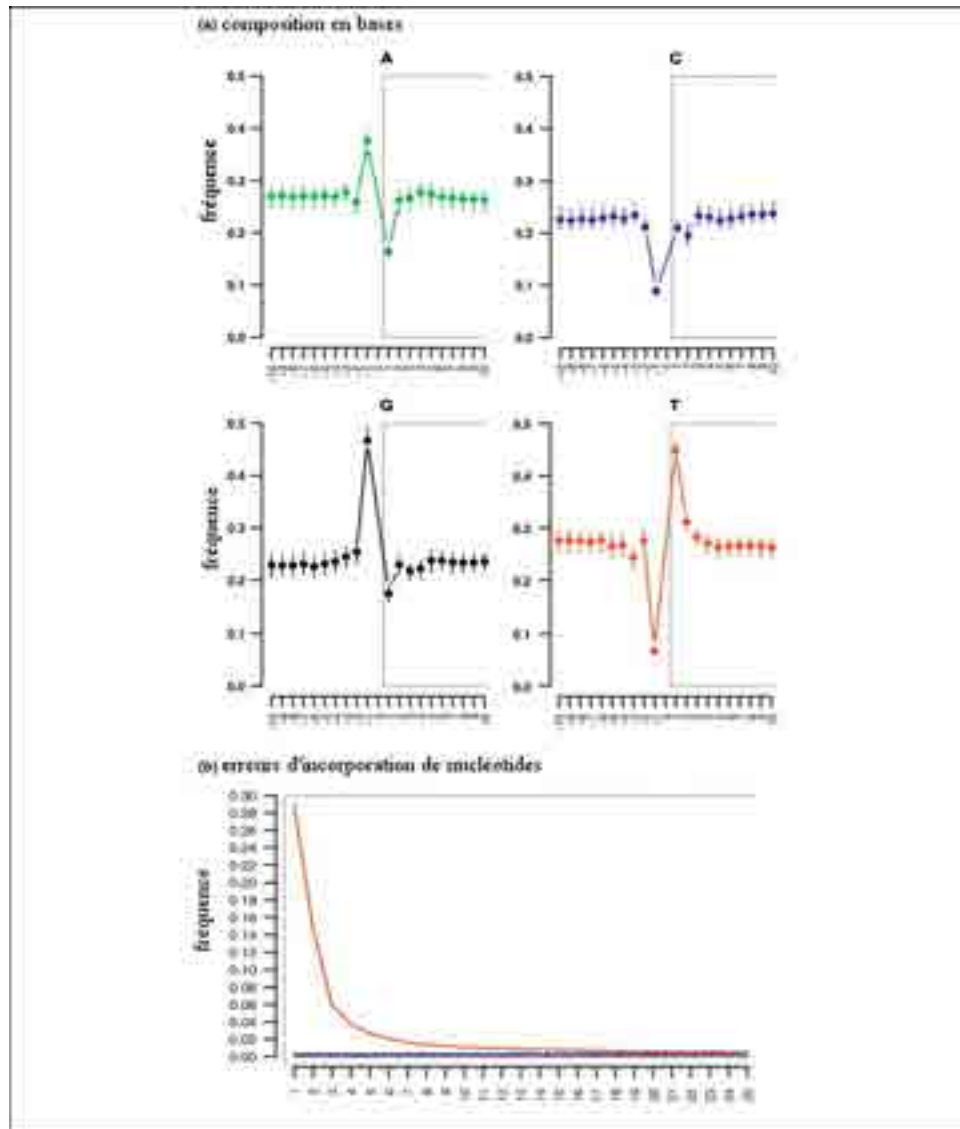


Figure 3 : Motif de dégradations de l'ADN obtenu à partir du séquençage de l'ADN d'un cheval du pléistocène

(a) la composition en nucléotides des séquences lues est reportée pour les 10 premières bases (1 à 10) ainsi que les 10 bases précédant ces séquences (-10 à -1) par alignement sur le génome. La fréquence moyenne d'apparition de chacun des nucléotides sur ces 20 positions est indiquée dans les 4 graphiques. (b) fréquence des erreurs d'incorporation de nucléotides observée sur les 25 nucléotides adjacents aux extrémités. La fréquence de toutes les erreurs d'incorporations est notée en gris à l'exception des transitions C→T et G→A qui sont indiquées respectivement en rouge et en bleu. Le séquençage a été réalisé sur Illumina. Modifiée de (Overballe-Petersen et al., 2012).

I.2.2.2 Préservation de l'ADN ancien

Tous les phénomènes de dégradation vus précédemment vont entraver la conservation de l'ADN au sein de l'organisme *post mortem*. Bien que ces mécanismes soient identiques pour tous les individus, leur survenue est dépendante de nombreux facteurs. En effet, les facteurs environnementaux tels que l'humidité, la température, la salinité ou encore le pH peuvent

influer sur la préservation de l'ADN au sein des organismes. Pour une meilleure préservation de l'ADN, il serait donc préférable que le corps soit conservé dans un environnement avec des températures basses et sans grandes variations (idéalement dans le permafrost), un milieu sec, anaérobie et à l'abri des rayonnements UV afin de limiter les réactions enzymatiques et chimiques. Les premiers tests *in vitro* avaient fixé cette limite à quelques centaines de milliers d'années (Lindahl, 1993). Néanmoins, certains milieux comme les milieux froids, pourraient permettre à l'ADN d'être conservé beaucoup plus longtemps (Allentoft et al., 2012), bien qu'il soit impossible aujourd'hui de prédire, au sein d'un organisme, l'état de préservation du génome. Cependant, les dernières publications attestent d'ADN conservé dans des restes vieux de 700000 ans.

I.2.2.3 Les sources d'ADN ancien

La préservation de l'ADN va également varier en fonction de sa source. Potentiellement, l'ADN peut être retrouvé dans chaque prélèvement biologique, or tous ne subissent pas les dommages du temps de la même façon. En effet, les tissus mous seront dans la plupart des cas rapidement dégradés et seuls les tissus durs tels que les os et les dents persisteront, faisant d'eux les substrats les plus abondamment retrouvés sur les terrains de fouille et les plus étudiés par les paléogénéticiens. La préservation de l'ADN au sein des tissus durs serait due à l'adsorption de l'ADN sur la matrice minérale composée principalement d'hydroxyapatite de calcium (Campos et al., 2012; Lindahl, 1993).

Néanmoins, des conditions particulières telles que le permafrost ou la momification permettront de conserver plus de substrats biologiques et sources d'ADN potentielles. Parmi ces derniers, les cheveux ou les poils se sont révélés être des substrats intéressants du fait de leur structure hydrophobe, principalement composée de kératine, qui les rend moins sensibles aux contaminations extérieures et plus faciles à nettoyer (Gilbert et al., 2006; Gilbert et al., 2004). Ainsi, le premier génome humain ancien a été obtenu à partir de cheveux conservés dans le permafrost (Rasmussen et al., 2010).

D'autres substrats tels que les coprolithes peuvent également être utilisés pour étudier le sujet lui-même, sa flore intestinale et les pathogènes associés ou encore les habitudes alimentaires. Le plus vieux coprolithe humain étudié est daté d'environ 12500 ans (Jenkins et al., 2012).

Enfin, les sédiments eux-mêmes peuvent contenir des molécules d'ADN ancien et révéler la présence de certaines espèces malgré l'absence de restes macroscopiques (Willerslev et al.,

2003). L'ADN mitochondrial dérivé des urines, des fèces ou des cellules épithéliales par exemple persisterait dans les sédiments sur de longues périodes dans des conditions environnementales optimales, c'est-à-dire des sols très froids. Ces molécules d'ADN sont généralement appelées *sedaDNA* (sedimentary ancient DNA). La découverte de ces molécules d'ADN ancien signifie que le sujet auquel elles appartiennent se trouvait à cet endroit à cette période. Cela permettrait par exemple de dater la disparition de certaines espèces (Haile et al., 2009). Une difficulté majeure dans l'utilisation de *sedaDNA* est qu'il est impossible de réaliser une datation directe, donc la fidélité des études dépendra de l'étude du site et de la datation des restes associés.

I.2.2.4 Les inhibiteurs de l'amplification

Une fois l'ADN extrait, la plupart des techniques utilisées sont basées sur une amplification par PCR, or, cette étape peut être limitée par des inhibiteurs généralement co-extraits avec les molécules d'ADN ancien. Peuvent être considérées comme des inhibiteurs toutes les substances qui ont un effet néfaste sur la PCR, amenant parfois à des faux négatifs. Parmi les substances inhibitrices on retrouvera par exemple des composants organiques du sol (acides humiques, acides fulviques, tannins...) ou encore des constituants même de l'organisme (l'hème, le collagène de type I, l'urée, la mélanine, l'eumélanine ...). Une liste plus complète est présentée dans la revue de (Schrader et al., 2012). Les modes d'action de ces inhibiteurs sont variables : certains d'entre eux peuvent agir sur l'ADN polymérase soit en la dégradant (l'urée), en l'inhibant (le collagène), en s'y complexant (la mélanine) ou en dérivant son activité en mimant la structure des acides nucléiques (les polysaccharides). D'autres substances inhibitrices interagissent directement avec la molécule d'ADN (les acides humiques). Enfin, certaines substances peuvent interagir avec les cofacteurs nécessaires à la polymérase.

La présence de ces inhibiteurs peut cependant être facilement détectée par PCR multiplexe en temps réel. En effet, lors de la quantification de l'ADN, un contrôle interne de PCR (IPC : Internal PCR Control), correspondant à une séquence synthétique non retrouvée dans le génome humain, est utilisé afin de détecter la présence d'inhibiteurs. La concentration d'IPC dans la réaction étant connue, toute diminution de sa concentration lors de la quantification sera le signe d'une inhibition de la réaction. Cette méthode est utilisée dans le kit QuantifilerTM commercialisé par la société Life Technologies, couramment utilisé dans les

études d'ADN ancien ainsi qu'en criminalistique où il est fréquent de faire face à ces inhibiteurs de PCR du fait des substrats biologiques de nature variée remis par les enquêteurs.

Des moyens peuvent être mis en place pour limiter l'action et la présence de ces inhibiteurs. Aujourd'hui, les kits d'extractions et de purifications commercialisés sont développés afin de limiter la présence de ces inhibiteurs par l'utilisation de colonnes de silice, qui ne retiendront que les acides nucléiques, ou par une étape de micro-concentration afin d'éliminer les très petits fragments qui pourraient limiter l'efficacité de la PCR. Si les inhibiteurs persistent après extraction, le choix de la polymérase ou l'ajout de BSA (Bovine Serum Albumine) dans les mélanges réactionnels aide à contenir leurs effets. Enfin, la dilution des extraits d'ADN est un moyen de diminuer la présence des inhibiteurs, même s'il convient peu à l'analyse d'extraits très faiblement concentrés en ADN.

I.2.2.5 Les contaminations

L'état dégradé de l'ADN ancien va introduire des difficultés supplémentaires lors de son analyse. L'une de ces difficultés est d'éviter les contaminations par des molécules d'ADN exogènes tout au long du traitement des échantillons. En effet, des molécules d'ADN intactes provenant d'organismes modernes, même en très petites quantités, seront préférées comme matrice par la *Taq* polymérase au moment de l'amplification, aux dépens de la molécule d'ADN ancien qui présente de nombreuses dégradations empêchant ainsi son amplification. Ce phénomène est d'autant plus dangereux lorsque les études portent sur des restes humains anciens car contrairement à de l'animal ou du végétal, il est beaucoup plus difficile de différencier une contamination par un membre des équipes ayant été en contact avec l'échantillon de l'ADN endogène. Cette contamination peut apparaître à tout moment, lors des fouilles archéologiques, de l'échantillonnage, et durant les différentes étapes de traitement au sein du laboratoire. Cela est particulièrement compliqué lorsque l'on travaille à partir de collections anciennes de musées où il est souvent impossible de retrouver toutes les personnes ayant été en contact avec les prélèvements. C'est pourquoi, l'un des défis dans les études paléogénétiques est d'adopter une stratégie et une méthode permettant d'assurer l'authenticité des résultats.

I.2.3 Critères d'authenticité

Dans l'objectif d'assurer l'authenticité des résultats, des premières mesures, se résumant en trois points (analyse de plusieurs extraits par échantillons, utilisation de contrôles négatifs et observation d'une corrélation inverse entre l'efficacité de la PCR et la taille des fragments d'ADN), ont été mises en place en 1989 (Pääbo, 1989). Ces mesures ont évolué au cours du temps pour aboutir à la publication en 2000 par Cooper et Poinar de la liste des « 9 critères d'authenticité » bien connue des paléogénéticiens et présentée ci-dessous (Cooper and Poinar, 2000).

1) Zones de travail physiquement isolées

Toutes les étapes en amont de l'amplification doivent être réalisées dans un laboratoire dédié à l'ADN ancien. Ce laboratoire doit être physiquement isolé des laboratoires où se déroulent les étapes post amplification afin d'éviter les contaminations.

2) Utilisation de contrôles négatifs

Des contrôles négatifs doivent être inclus à chaque étape du protocole (extraction, amplification). Il est préférable de ne pas utiliser de contrôle positif afin d'éviter les risques de contamination.

3) Cohérence dans les résultats obtenus

Une corrélation inverse entre l'efficacité de la PCR et la taille des fragments amplifiée doit être observée. Les séquences obtenues doivent avoir un sens phylogénétique.

4) Résultats reproductibles

Les résultats doivent être répétables à partir du même extrait mais aussi à partir d'extraits différents pour chaque individu. Il est essentiel d'utiliser plusieurs couples d'amorces afin d'obtenir des séquences chevauchantes qui doivent être identiques au niveau de la partie commune.

5) Clonage

Les séquences obtenues doivent être vérifiées par clonage des produits amplifiés puis séquencées afin de déterminer le ratio de séquences endogènes/exogènes, de détecter les erreurs de réplication dues aux possibles dommages subis par l'ADN ou encore la présence de séquences recombinantes.

6) Réplifications indépendantes

Les analyses doivent être répliquées dans des laboratoires indépendants afin de s'assurer qu'il n'y ait pas de contaminations intra-laboratoire.

7) Evaluer l'état de préservation biochimique de l'échantillon

La préservation de l'ADN peut être corrélée à la préservation d'autres biomolécules comme le collagène ou les acides aminés.

8) Quantification de l'ADN

L'ADN extrait doit être quantifié afin d'assurer une certaine fiabilité aux résultats. En effet, en dessous de 1000 copies d'ADN il est alors impossible d'exclure une contamination sporadique.

9) Etude des restes associés

Lorsqu'il existe des restes associés aux prélèvements humains, tels que des restes de faune, leur analyse est importante pour s'assurer qu'il n'y a pas de contaminations humaines au sein de ces échantillons.

Ces critères vont permettre de guider les paléogénéticiens, mais néanmoins rares sont les études qui utilisent tous ces critères. En effet, le respect de l'ensemble de ces critères n'est pas un garant absolu de l'authenticité des résultats. Une démarche réfléchie et critique vis-à-vis des travaux réalisés et des résultats obtenus est préférée afin d'assurer au maximum la fiabilité des résultats (Gilbert et al., 2005). La démarche mise en place au sein du laboratoire sera détaillée dans les sections II.2.4 et III.2.

I.2.4 Champs d'applications des études paléogénétiques

Les études d'ADN ancien se sont intéressées à de nombreuses problématiques. Elles ne se limitent pas à l'étude d'ADN humain mais touchent aussi l'étude de l'ADN extrait de nombreuses espèces animales, végétales ou encore de pathogènes. Nous en donnons ici un aperçu très succinct.

I.2.4.1.1 ADN non humain

Les études d'ADN non-humain s'intéressent parfois à l'étude d'espèces éteintes telles que le quagga (Higuchi et al., 1984), l'ours des cavernes (Loreille et al., 2001), le moa (Cooper et al., 2001) ou encore le mammouth (Poinar et al., 2006). Des travaux se sont également intéressés à la domestication des espèces animales, c'est-à-dire à l'adaptation ou transformation de caractères héréditaires au contact de l'Homme, telles que le chien (Ovodov et al., 2011; Thalmann et al., 2013), le bétail (Cieslak et al., 2010; Edwards et al., 2007; Larson et al., 2010) ou encore à la domestication des végétaux et leur évolution (Palmer et al., 2012).

I.2.4.1.2 Paléo-pathologie

Il est également possible, à partir de restes humains anciens, d'extraire l'ADN des micro-organismes qu'ils hébergent. L'étude des pathogènes anciens, appelée paléo-pathologie, a ainsi permis d'apporter de nombreux éléments quant à l'état sanitaire des populations du passé et l'évolution des maladies infectieuses. L'accès à ces génomes anciens apporte également de précieuses informations en termes d'évolution moléculaire. Ainsi, plusieurs génomes ont été analysés appartenant aux bactéries responsables de la peste (Bianucci et al., 2008; Bos et al., 2011; Drancourt et al., 1998; Drancourt et al., 2007; Haensch et al., 2010; Tran et al., 2011a; Tran et al., 2011b; Wiechmann et al., 2010), de la tuberculose (Hershkovitz et al., 2008), et de la coqueluche (Theves et al., 2011), ou aux virus de la grippe espagnole (Taubenberger et al., 2005) et de la variole (Biagini et al., 2012).

I.2.4.1.3 ADN humain

Grâce à l'étude des restes humains il a été possible d'apporter de nombreux éléments de réponse à des questions débattues depuis des décennies. Les progrès techniques, notamment l'utilisation des séquenceurs de nouvelles générations ont permis d'étudier les relations génétiques entre les différents hominins, dont par exemple la possible interaction entre l'Homme de Néandertal et l'Homme moderne (Green et al., 2010; Sankararaman et al., 2014; Vernot and Akey, 2014). Cela a permis également la découverte d'un nouvel hominidé, l'Homme de Denisova (Meyer et al., 2012; Reich et al., 2010), proposant pour la première fois de caractériser une population uniquement sur la base de son génome.

A l'échelle des populations, il est possible d'étudier leurs dynamiques, leurs mouvements et ainsi d'apporter des éléments quant à la compréhension du peuplement de l'Amérique (Garcia-Bour et al., 2004; Lalueza et al., 1997), de l'Europe (Bramanti et al., 2009; Deguilloux et al., 2011; Der Sarkissian et al., 2013; Gamba et al., 2012; Haak et al., 2010; Haak et al., 2005; Lacan et al., 2011a; Lacan et al., 2011b; Malmstrom et al., 2009; Sampietro et al., 2007), des steppes eurasiennes (Keyser et al., 2009; Lalueza-Fox et al., 2004; Molodin et al., 2012) ou de l'Océanie (Adcock et al., 2001). Les études paléogénétiques peuvent également apporter des éléments pour la compréhension du recrutement funéraire que ce soit au sein de grands ensembles tels que la nécropole Xiongnu d'Egyin Gol en Mongolie (Keyser-Tracqui et al., 2003a) ou à l'échelle d'une sépulture collective (Clisson et al., 2002; Deguilloux et al., 2014). L'étude des SNP des génomes anciens permet également de s'intéresser à des adaptations fonctionnelles comme la capacité à digérer le lait frais (Burger et al., 2007; Lacan et al., 2011a; Malmstrom et al., 2010) ou la capacité de disposer d'un langage articulé (Krause et al., 2007a). Certains caractères phénotypiques des individus anciens peuvent également être prédits, tels que des caractères pigmentaires (couleur des yeux, cheveux, peau) (Bouakaze et al., 2009; Draus-Barini et al., 2013; Lalueza-Fox et al., 2007; Olalde et al., 2014). Enfin à l'échelle de l'individu, le sexage permet également d'apporter un élément essentiel aux anthropologues, notamment pour la caractérisation des juvéniles pour lesquels les méthodes physiques de détermination du sexe restent peu fiables. L'identification de personnages historiques est également possible grâce aux études d'ADN ancien comme par exemple pour les restes de la famille Romanov (Coble et al., 2009; Gill et al., 1994; Rogaev et al., 2009), de Louis XVI (Lalueza-Fox et al., 2011), de Richard III (www.le.ac.uk/richardiii), ou encore d'Henri IV (Charlier et al., 2010). Notons que ces identifications font souvent l'objet de polémiques et contestations.

Ainsi, aujourd'hui la capacité à séquencer des génomes entiers ou à typer un très grand nombre de polymorphismes à partir d'ADN extrait de spécimens anciens permet d'avoir accès à de très nombreuses informations. Notons que l'étude de certains caractères reste néanmoins liée à l'avancement des travaux menés sur les génomes actuels (étude de gènes candidats ou études d'associations).

I.2.5 Evolution de la discipline : d'un petit fragment au génome complet

Dans un premier temps, les études paléogénétiques s'intéressaient à des petits fragments d'ADN, généralement localisés sur l'ADN mitochondrial qui a l'avantage d'être présent en un nombre élevé de copies au sein d'une cellule contrairement à l'ADN nucléaire. L'évolution de la discipline a accompagné l'évolution des technologies. Ces développements technologiques et méthodologiques ont permis, dans un premier temps, l'étude de fragments d'ADN nucléaire, notamment grâce aux progrès réalisés dans le domaine criminalistique. Il était alors possible d'établir par exemple des empreintes génétiques à partir de faible quantité d'ADN ou encore de typer un nombre de plus en plus grand de SNP (Single Nucleotide Polymorphisms). Puis le premier génome mitochondrial ancien a été séquencé au début des années 2000, il s'agissait d'un génome de moa, un oiseau non volant (Cooper et al., 2001). Le séquençage a été réalisé à partir de PCR classique en amplifiant des séquences chevauchantes de 400 à 600 pb, afin de couvrir l'ensemble du génome de plus de 16000 paires de bases. Au milieu des années 2000, la discipline a connu une nouvelle révolution avec l'apparition des technologies de séquençage à haut débit ou NGS (Next Generation sequencing). Sous ce terme sont regroupées plusieurs technologies développées par différentes sociétés permettant la lecture de plusieurs milliers de séquences en parallèles. Bien qu'elles présentent toutes des caractéristiques propres et par conséquent chacune leurs avantages et leurs inconvénients, globalement ces méthodes sont assez proches. Elles peuvent être décomposées en 3 étapes principales : (i) la préparation et l'amplification des molécules d'ADN, (ii) l'incorporation des nucléotides, et (iii) la lecture des séquences. On parlera alors de séquenceurs de seconde génération par opposition au séquençage classique par la méthode Sanger.

Le premier succès obtenu avec l'utilisation de ces plateformes à partir d'ADNa a consisté au séquençage d'une partie du génome d'un mammouth (Poinar et al., 2006). Concernant, l'Homme, le premier génome mitochondrial humain ancien a été publié en 2008 (Gilbert et al., 2008) et le premier génome total humain en 2010 (Rasmussen et al., 2010). S'ensuivit de près la publication du premier génome de Neandertal en 2010 (Green et al., 2010). Depuis, le nombre de génomes anciens séquencés ne cesse d'augmenter, repoussant toujours un peu plus les limites de la paléogénomique. Parmi ces génomes, on relèvera notamment le génome d'un aborigène australien d'une centaine d'années (Rasmussen et al., 2011) ou celui de l'homme des glaces Ötzi (Keller et al., 2012). Le plus vieux génome d'un Homme moderne daté de

24000 ans et retrouvé dans le sud de la Sibérie centrale a été publié très récemment (Raghavan et al., 2014). Il y a quelques mois, c'est la parution du génome mitochondrial d'un Homininé archaïque de plus de 300000 ans au nord de l'Espagne qui a intéressé la communauté scientifique (Meyer et al., 2014). Le génome d'un Homme de Neandertal, cette fois-ci retrouvé dans l'Altai a également été analysé récemment (Prufer et al., 2014). Enfin, parmi toutes ces avancées, c'est la parution du séquençage du génome d'un cheval retrouvé dans le permafrost daté d'environ 700000 ans (Orlando et al., 2013) qui a repoussé les limites de préservation de l'ADN au sein des organismes laissant ainsi envisager l'étude d'ADN datant de près d'un million d'années.

Les développements étant constants, une nouvelle génération de séquenceurs (troisième génération) permettant de faire du séquençage en molécule unique, ont été développés et semblent également prometteurs pour les études d'ADN ancien (Ginolhac et al., 2012; Orlando et al., 2011). En effet, cette méthode s'affranchit des étapes de préparations et d'amplifications des molécules d'ADN, en amont du séquençage, qui peuvent générer des incorporations de bases erronées, (surtout lors de l'analyse d'ADN ancien qui a généralement subi de nombreux dommages post-mortem). De plus, le matériel de départ requis est moindre en comparaison des séquenceurs de seconde génération, permettant ainsi d'envisager l'étude de prélèvements ne renfermant que d'infimes quantités d'ADN. Néanmoins, leur utilisation n'est encore que très anecdotique.

Ainsi, en quelques années, la discipline a connu une accélération importante (Figure 4) passant de l'analyse d'une centaine de paires de bases du génome mitochondrial avec une limite de retour dans le passé posée à 100000 ans, à aujourd'hui des génomes entiers ou partiels allant jusqu'à 700000 ans.

L'obtention des génomes totaux permet actuellement d'envisager l'exploration d'autres informations telles que les modifications épigénétiques (Llamas et al., 2012), afin d'obtenir, par exemple, des informations sur l'âge au décès des individus (Pedersen et al., 2013).

Néanmoins, ces développements technologiques ne s'appliquent pour le moment qu'à un nombre limité de spécimens et ne permettent pas d'étudier des « populations » anciennes. Or, l'étude de la variabilité génique au sein de ces populations est un élément clé pour comprendre les processus de peuplement de certaines périodes et/ou régions.

	Extraction	Amplification	Séquençage	Analyse	Données générées
Pre NGS	Phénol-chloroforme Silice	PCR	Sanger Clonage moléculaire	Évaluation des dommages de l'ADN et des contaminations à l'œil	Courtes séquences
NGS	Silice (optimisation des protocoles)	Préparation de librairies Enrichissement	Séquençage haut débit	Bioinformatique profils de dégradation de l'ADN	Génomes
Futur	Augmentation de l'efficacité Récupération de très courts fragments d'ADN		Molécule unique	Détection des sites méthyles Meilleure caractérisation des dommages	Augmentation du nombre de génomes anciens, diversité des espèces...

Figure 4 : Evolution des études d'ADN ancien

L'utilisation des séquenceurs de nouvelles générations (NGS) a considérablement augmenté la portion de génome analysée. De la même façon, les améliorations des techniques d'extraction et de préparation des échantillons ont également permis d'augmenter la quantité d'ADN disponible pour l'analyse. Les améliorations futures permettront sans doute d'augmenter la quantité d'ADN récupérée par l'optimisation des méthodes d'extraction, le séquençage en molécule unique permettra de s'affranchir des étapes d'amplification en amont et ainsi il sera certainement possible d'obtenir l'analyse de génomes anciens encore plus divers et d'augmenter leur nombre afin d'avoir des éléments de comparaison. L'évolution porte également sur les informations obtenues à partir de ces génomes, telles que l'étude des modifications épigénétiques. Adaptée de (Shapiro and Hofreiter, 2014).

I.3 Contexte général de l'étude

Après avoir présenté la discipline dans son ensemble, nous nous intéresserons dans cette partie au contexte géographique et culturel de l'étude. Les sujets étudiés proviennent des régions du sud de la Sibérie et de l'Altaï et sont datés de l'âge du Bronze (cf. section II.1). Ainsi, nous aborderons l'intérêt d'étudier cette région et plus particulièrement cette période.

I.3.1 Contexte géographique

Couvrant une région large de plus de 8000 km, de l'Ukraine à la Mongolie, les steppes eurasiennes connectent l'Europe, l'Asie centrale, la Chine, l'Asie du Sud et le Moyen Orient (Figure 5). Cet immense territoire peut être divisé en plusieurs zones en fonction du climat et de la végétation. En effet, il y a un gradient de terres de plus en plus arides entre le Nord et le Sud. Trois zones sont généralement distinguées du nord au sud : la steppe boisée, la steppe herbeuse, puis la steppe semi-désertique qui débouche sur les déserts du Taklamakan et de Gobi à l'Est. La formation de ces steppes est le résultat d'un climat continental et d'un déficit d'humidité ayant dicté l'écosystème de cette région. De grands fleuves rythment ces steppes qui ne sont que très peu entrecoupées. En effet, seul le massif montagneux de l'Altaï, joint à ceux du Tianshan et du Pamir, forment une barrière géologique haute de plus de 4000 mètres mais qui reste néanmoins franchissable en de nombreux points.

Le massif de l'Altaï présente un étagement habituel de climats depuis les forêts de base jusqu'à la végétation raréfiée des sommets. La chaîne montagneuse s'étale sur plus de 2000 km et est comprise entre la Russie, le Kazakhstan, la Mongolie et la Chine, ou plus précisément sa province du Xinjiang. De nombreux grands fleuves sibériens y prennent leur source en particulier l'Ob et l'Irtych. Dans le sud de la Sibérie, en république de Touva, le fleuve Iénisseï prend sa source et traverse le bassin du Minoussinsk connu pour les nombreux objets en bronze, « les bronzes du Minoussinsk », extraits des tombes de cette région.



Figure 5 : Carte géographique d'une partie de l'Eurasie

I.3.2 Peuplement de la région

I.3.2.1 Les premiers peuplements

Le sud de la Sibérie et l'Altai hébergent de nombreux vestiges archéologiques qui ont révélé la présence de populations humaines très diverses. L'homme moderne y aurait vécu depuis plus de 40000 ans (Goebel, 1999). La découverte de restes appartenant à l'Homme de Neandertal dans le site d'Okladnikov, dans les montagnes de l'Altai, a étendu encore plus à l'est l'aire de répartition des Néandertaliens (Krause et al., 2007b) qui était jusqu'il y a quelques années limitée au site de Teshik-Tash en Ouzbékistan. Récemment, un nouvel homininé a été découvert dans l'Altai : l'Homme de Denisova, (Meyer et al., 2012; Reich et al., 2010).

I.3.2.2 Les peuplements préhistoriques

Depuis l'âge du Bronze, se sont développés, sur une zone allant de l'Europe de l'Est à l'Altaï, de nombreux peuples des steppes qui ont étendu leurs influences vers l'Europe et l'Asie. A certains moments, des flux géniques l'ont emporté sur d'autres. En effet, des influences européennes ont pu s'exercer jusqu'au cœur de l'Asie centrale aux âges des métaux alors que quelques siècles plus tard, cette région allait connaître des influences asiatiques. De nombreuses cultures contemporaines ou successives ont été décrites tout le long du couloir des steppes. Nous présenterons ici principalement les cultures de la région étudiée.

I.3.2.2.1 L'âge du Bronze

Le Néolithique est une période encore assez peu connue dans la région du sud de la Sibérie et de l'Altaï. La séquence culturelle connue commence au milieu du IV^{ème} millénaire avant J.C avec l'apparition de la culture d'**Afanasievo** (Figure 6). Les études archéologiques et anthropologiques ont, dans un premier temps, suggéré un lien entre cette culture et la culture de Yamna (tombes à fosses) des steppes d'Europe de l'Est. Cette hypothèse est basée sur la découverte des mêmes rituels funéraires, du même matériel culturel et de caractères physiques communs entre les individus mis au jour. Ce lien direct entre les cultures de Yamna et d'Afanasievo a été nuancé par David W. Anthony qui propose un lien avec une culture pre-Yamna, la culture Repin, localisée entre la Volga et l'Oural, qui aurait migré à travers les steppes Kazakhes vers 3700-3500 avant J.-C et se serait établie dans l'ouest de l'Altaï (Anthony, 2007). Cette culture se serait étendue du bassin du Minoussinsk au nord du Bassin du Tarim, en passant par l'Altaï et aurait perduré jusqu'au milieu du III^{ème} millénaire avant J.-C (Svyatko et al., 2009). Certaines pratiques Yamna, retrouvées dans l'Altaï, suggèrent des échanges continus entre l'Altaï et l'Oural pendant cette période. Peu de sépultures sont retrouvées dans le nord des steppes kazakhes. Les porteurs de la culture d'Afanasievo auraient, selon certains auteurs, été les locuteurs de la branche Tokharienne des langues indo-européennes, branche uniquement présente dans le Bassin du Tarim (Mallory and Mair, 2000). Si la culture Afanasievo est la plus connue et la plus étudiée, durant le IV^{ème} millénaire, se trouvait également au pied de l'Altaï, une culture proche de la culture d'Afanasievo : la culture **Bol'shemyskaya**, actuellement très peu documentée.

A la culture d'Afanasievo a succédé la culture d'**Okunevo** (2500-1800 avant J.-C), dans le bassin du Minoussinsk plus que dans l'Altaï. Bien que les cultures d'Okunevo et

d'Afanasievo présentent des similarités au niveau du matériel culturel, la documentation anthropologique indique que la population d'Okunevo était plus « mongoloïde » que la population d'Afanasievo (Blyakharchuk and Chernova, 2013). Cette culture a coexisté avec la culture d'**Elunino**, présente sur les contreforts de l'Altai, ainsi qu'avec la culture **Chemurchek** dans le nord-ouest de la Mongolie. C'est également à cette période, avec le commerce du métal, qu'un grand système d'interactions entre les cultures des steppes commence.

Dans un troisième temps, ces cultures ont été supplantées par la culture d'**Andronovo** (1800-1200 avant J.-C). On parle de culture d'Andronovo mais il s'agit plutôt d'un ensemble culturel qui se serait étendu sur un vaste territoire englobant le sud de la Sibérie et l'Asie Centrale. Cette culture serait apparentée à la culture de **Srubnaya** (tombes à rondins), dans l'ouest des steppes formant ainsi une chaîne de cultures similaires des frontières de l'Europe à la Chine, et faisant des steppes un corridor de communications. La fin de l'âge du Bronze est marquée par la culture de **Karasouk** (1500-1000 avant J.-C), qui s'est développée sur la partie orientale du territoire occupé par la culture d'Andronovo.

Durant cette période, la domestication du cheval (Outram et al., 2009) ainsi que l'utilisation des chars, permettant d'être plus mobile, participeront au développement et à la diffusion de ces cultures le long des steppes eurasiennes.

Contrairement au sud de la Sibérie, les connaissances sur la période de **l'âge du Bronze en Mongolie** étaient, jusqu'à la fin du 20^{ème} siècle, très faibles. Les études archéologiques réalisées ces dernières années ont permis d'apporter de nouvelles connaissances et ont révélé qu'il existait des différences importantes entre l'ouest et l'est de la Mongolie. Dans la partie ouest, les cultures de l'âge du Bronze et du début de l'âge du Fer semblent être proches des cultures du sud de la Sibérie. En effet, des monuments ont été découverts et attribués aux cultures d'**Afanasievo** (première moitié du III^{ème} millénaire avant J.-C) et de **Chemurchek** (2500-1800 avant J.-C), ainsi qu'à une culture locale appelée **Munkh Khairkhan** (1800-1500 avant J.-C). Enfin, des tombes qui n'ont pu être rattachées à aucune culture connue, ont été regroupées sous le terme de culture de **Baitag** (1400-1100 BC) (Kovalev and Erdenebaatar, 2009).

	Steppes de l'Ouest	Bassin du Minoussinsk	Altaï, Russie	Altaï, Mongolie
IV	Repin		Bol'shemyskaya	
	Yamna	Afanasievo	Afanasievo	Afanasievo
III	Katakombnaya	Okunevo	Elunino	Chemurchek
II	Srubnaya	Andronovo	Andronovo	Munkh-Khairkhan
I				

Millénaires avant J.-C.

Figure 6 : Répartition chronologique et géographique des cultures mentionnées dans ce travail entre le IV^{ème} et I^{er} millénaire avant J.-C

I.3.2.2.2 Depuis l'âge du Fer

L'âge du Fer dans les steppes eurasiennes est marqué par le développement de la civilisation Scythe. Cette civilisation regroupe des peuples partageant des caractéristiques culturelles communes, appelées « triade scythique », qui sont l'art animalier, un armement typique et le harnachement du cheval. Elle s'est étendue sur plus de 8000 km, des rives de la mer Noire à la Mongolie (Lebedynsky, 2011). Deux pôles majeurs de cette civilisation sont connus : les Scythes d'Europe, localisés sur les bords de la mer Noire, et les Scytho-Sibériens, retrouvés au sud de la Sibérie. Ces peuples ont laissé de nombreux kourganes¹ dont des kourganes royaux très imposants (exemple du kourgane d'Arzhan-1 dans la république de Touva, au sud de la Sibérie, qui faisait 120 mètres de diamètre ou celui de Berel au Kazakhstan oriental (Clisson et al., 2002)). Les données archéologiques ont permis d'apporter de nombreux éléments relatifs à cette culture mais plusieurs points restent aujourd'hui débattus notamment l'origine et la diffusion de cette culture ainsi que les relations génétiques entre ces deux foyers.

¹ Un kourgane est le terme russe pour tumulus. Cela correspond à une tombe creusée, surmontée d'un monticule de pierre.

A cette civilisation Scythe ont succédé d'autres grands mouvements de populations, majoritairement mongoloïdes. Tout d'abord les populations hunniques, suivies des peuples Turques au 11^{ème} siècle, qui eurent un impact culturel majeur sur la région avec la mise en place des langues Turques de la famille Altaïque. Enfin, vers le 13^{ème} siècle après J.-C, c'est l'un des plus grands empires, l'empire mongol de Gengis Khan s'étendant de la mer Caspienne à la Chine, qui aurait eu, un impact à la fois culturel et génétique lors de sa diffusion. En effet, une lignée paternelle mongoloïde, actuellement présente chez 8% des hommes dans une large région d'Asie et une partie de l'Europe aurait diffusé, selon certains auteurs, avec les mouvements migratoires de Gengis Khan et ses descendants (Zerjal et al., 2003).

I.3.2.3 Les steppes eurasiennes et la diffusion des indo-européens

Les cultures décrites précédemment sont parfois associées à la diffusion des langues indo-européennes. En effet, un ensemble de langues parlées de l'Europe à l'Inde, en passant par l'Iran, ont entre elles plus d'affinités qu'avec les autres groupes linguistiques. Ces langues ont donc été rassemblées sous le terme de langues indo-européennes. Du moment où l'on parle d'une famille linguistique se pose la question de son origine et de son histoire. Ainsi est venue l'idée qu'un peuple parlant la langue originelle (proto-indo-européen) aurait diffusé à partir d'une région de l'Eurasie et aurait répandu cette langue. Deux grandes théories coexistent actuellement pour expliquer cette diffusion. La première, défendue par l'archéologue Colin Renfrew, propose une diffusion des langues indo-européennes à partir du Proche-Orient en l'associant à la diffusion de l'agriculture et de l'élevage du Néolithique. La seconde, défendue par Marija Gimbutas et reprise par James Mallory, propose un foyer originel au niveau des steppes européennes du nord de la mer Noire puis une diffusion associée à une série de cultures steppiques réunies au sein d'un même ensemble culturel appelé «culture des Kourganes», caractérisé par ses rites funéraires (tombes sous tumulus ou kourganes), le rôle du cheval et du charriot, et une économie basée sur l'élevage. Ainsi, selon la théorie de Maria Gimbutas, cet ensemble culturel serait la traduction archéologique de l'ensemble des locuteurs de l'indo-européen et de leur expansion.

I.3.3 Le peuplement actuel

La population actuelle du sud de la Sibérie et de l'Altaï est composée de nombreuses populations issues des mêmes formations historiques et culturelles. Néanmoins, aujourd'hui ces populations sont plus ou moins isolées et présentent parfois des différences à la fois sur des bases anthropologiques, linguistiques et génétiques (Dulik et al., 2012).

Afin de tenter de comprendre l'origine et l'évolution des populations locales actuelles, des études génétiques ont été menées sur ces populations contemporaines. Deux systèmes sont principalement étudiés pour retracer l'histoire des populations : l'ADN mitochondrial et le chromosome Y. En effet l'ADN mitochondrial et une grande partie du chromosome Y (environ 95%) sont transmis de façon uni-parentale. L'ADN mitochondrial est transmis de la mère à ses enfants. Son étude permet donc de retracer les lignées maternelles de l'ensemble des individus. Le chromosome Y est quant à lui transmis uniquement de père en fils, son étude permet donc de suivre les lignées paternelles des individus masculins.

Les études des lignées maternelles, menées sur l'ensemble des populations modernes du sud de la Sibérie, ont toutes révélé un mélange de lignées ouest- (H, HV, J, T et U, X, N1a) et est-eurasiennes (C, D, F, A, B, G, M, Y et Z) sur la base de leur distribution actuelle. Ces populations présentent des lignées mitochondriales très hétérogènes, ce qui pourrait être dû à leur situation géographique à la rencontre des mondes européens et asiatiques. Néanmoins, la distribution et la fréquence de ces lignées varient entre les différents groupes culturels (Derenko et al., 2003; Dulik et al., 2012; Gubina et al., 2013).

L'analyse de la région non recombinante du chromosome Y révèle également la présence d'haplogroupes ouest- (R1a1 majoritairement, R1b1b1) et est-eurasiens (N, C, Q, D). Par comparaison avec les lignées mitochondriales, les lignées paternelles partagées entre les différentes populations locales sont plus rares (Dulik et al., 2012; Khar'kov et al., 2011). Ces résultats indiquent un flux génique entre ces populations différent selon les hommes et les femmes. L'organisation patrilineaire souvent observée au sein de ces groupes humains pourrait expliquer ces différences.

Ainsi, ce qui est observé aujourd'hui est le résultat d'un ensemble complexe de flux géniques et d'évolution *in situ*. Comprendre les modalités de peuplement de cette région en étudiant les génomes des populations modernes est donc particulièrement difficile. Afin de s'affranchir des modifications du pool génique intervenues lors des différents mouvements migratoires,

seul l'accès direct aux génomes des populations du passé permet d'aborder efficacement l'histoire du peuplement de cette région.

I.4 L'apport des premières études paléogénétiques à la compréhension du peuplement du sud de la Sibérie et de l'Altai

Depuis une dizaine d'années, plusieurs études paléogénétiques ont été menées dans les régions du sud de la Sibérie, de l'Altai ainsi que du Bassin du Tarim (Chikisheva et al., 2007; Clisson et al., 2002; Gao et al., 2008; Gonzalez-Ruiz et al., 2012; Keyser et al., 2009; Lalueza-Fox et al., 2004; Li et al., 2010; Molodin et al., 2012; Mooder et al., 2006; Pilipenko et al., 2010; Ricaut et al., 2004a; Ricaut et al., 2004b). Ces études se sont principalement intéressées aux populations des âges du Bronze et du Fer. Le climat aride et froid, parfois présent dans ces régions, a certainement contribué à une relativement bonne conservation de l'ADN au sein des échantillons biologiques étudiés. Ainsi, l'ensemble des résultats obtenus a permis d'avoir une première vision de la structuration temporelle et spatiale des lignées génétiques aux âges du Bronze et du Fer dans ces régions.

I.4.1 Les données mitochondriales

La plupart des études se sont uniquement intéressées aux données mitochondriales. Les résultats obtenus montrent qu'à l'âge du Bronze, des individus du sud de la Sibérie (culture d'Andronovo et de Karasouk) (Keyser et al., 2009), de l'Altai russe (Chikisheva et al., 2007) ainsi que du Kazakhstan (Lalueza-Fox et al., 2004) étaient porteurs de lignées ouest-eurasiennes. Parmi ces lignées, figure la lignée e U5, également retrouvée chez des individus néolithiques à l'ouest du lac Baïkal (Mooder et al., 2006). Au contraire, dans l'Altai mongol, les individus de l'âge du Bronze semblaient être majoritairement porteurs de lignées est-eurasiennes, bien que ce résultat soit à considérer avec prudence puisque seuls 3 individus ont été étudiés (Gonzalez-Ruiz et al., 2012). Cette structure génétique semble avoir ensuite évolué vers un mélange de lignées ouest- et est-eurasiennes à l'âge du Fer, dans l'ensemble des régions (Chikisheva et al., 2007; Gonzalez-Ruiz et al., 2012; Keyser et al., 2009; Lalueza-Fox et al., 2004; Ricaut et al., 2004a; Ricaut et al., 2004b).

Néanmoins, des haplogroupes est-eurasiens semblent avoir diffusé au-delà de ces régions avant l'âge du Bronze puisque l'haplogroupe C a été retrouvé en Ukraine au néolithique et à l'âge du Bronze (Newton, 2011; Nikitin et al., 2012), les haplogroupes A, C, D, et Z ont été retrouvés dans l'ouest de la Sibérie dès le début de l'âge du Bronze (Molodin et al., 2012) et les haplogroupes C, D et Z dans le nord-est de l'Europe il y a environ 3500 ans (Der Sarkissian et al., 2013). Ces résultats suggèrent donc au moins un flux génique d'est en ouest avant l'âge du Fer.

Dans le Bassin du Tarim, la population présentait dès l'âge du Bronze un mélange de lignées mitochondriales est- et ouest-eurasiennes (Li et al., 2010).

I.4.2 Les données du chromosome Y

Bien que très informatives de l'origine biogéographique des populations, les lignées paternelles ont été beaucoup moins étudiées du fait de la plus grande difficulté à retrouver de l'ADN nucléaire au sein des organismes anciens, contrairement à l'ADN mitochondrial qui est présent en un nombre plus élevé de copies. Les deux études majeures ont révélé la présence de façon quasi exclusive de l'haplogroupe R1a1a (90% pour les individus sud-sibériens et 100% dans le bassin du Tarim) (Keyser et al., 2009; Li et al., 2010). Cet haplogroupe a également été retrouvé chez un individu Xiongnu au nord-est de la Mongolie (Kim et al., 2010).

La distribution actuelle de l'haplogroupe R1a1a, en Europe centrale, dans le sud de la Sibérie ainsi qu'en Inde, amène certains auteurs à proposer l'haplogroupe R1a1a comme un marqueur ou signature génétique de l'expansion des Indo-Européens (Semino et al., 2000; Wells et al., 2001). Néanmoins cette théorie est remise en question par d'autres auteurs qui proposent, pour cet haplogroupe, un foyer d'origine en Asie du Sud (Mirabal et al., 2009; Sharma et al., 2009). Des études plus récentes ont affiné la phylogénie de cet haplogroupe et mis en évidence la présence de sous-haplogroupes qui sépareraient des lignées européennes et asiatiques actuelles (Pamjav et al., 2012; Underhill et al., 2010). Ainsi, l'étude de ces marqueurs pourra surement apporter des éléments supplémentaires et aider à préciser l'origine de ces lignées masculines.

I.4.3 Les données autosomales

Peu de données génétiques proviennent de marqueurs autosomaux pour ces populations anciennes. A quelques exceptions près, seule l'étude de sujets sud-sibériens des âges du Bronze et du Fer a livré ce type d'informations. Les caractères phénotypiques tels que la couleur de la peau, des cheveux et des yeux ont été étudiés et les résultats obtenus suggèrent que la plupart des individus avaient probablement le teint pâle ainsi que les cheveux et les yeux clairs. Des marqueurs appelés AIM (Ancestry Informative Markers) ont également été étudiés. Ils ont été choisis du fait des différences importantes de fréquence allélique entre les populations (ici Africaine, Asiatique et Européenne). L'analyse de ces marqueurs a révélé que les sujets sud-sibériens anciens étaient génétiquement proches des populations européennes actuelles. Ainsi, l'ensemble de ces données suggère une probable origine européenne pour ces sujets anciens (Bouakaze et al., 2009; Keyser et al., 2009).

Ces données ont également permis de s'intéresser à la question du peuplement du bassin du Tarim. En effet, les données phénotypiques obtenues ainsi que la présence de l'haplogroupe R1a1a à la fois chez des individus du bassin du Tarim et chez des individus sud-sibériens de l'âge du Bronze ont renforcé l'hypothèse d'un lien entre les populations des steppes et les populations du bassin du Tarim. La découverte, au début du XX^{ème} siècle, dans le bassin du Tarim, de momies parfaitement conservées au phénotype plutôt européen, datant pour les plus anciennes du début du II^{ème} millénaire avant J.-C, avait amené de nombreuses interrogations quant à leur origine et au peuplement de cette région. Deux hypothèses ont alors été proposées (Hemphill and Mallory, 2004). La première, correspondant à un peuplement de la région par des agriculteurs sédentaires à partir des oasis de la Bactriane, région à cheval sur les régions actuelles de l'Afghanistan, l'Ouzbékistan et le Turkménistan. La seconde, proposant un peuplement à partir des populations des steppes du Nord par les porteurs de la culture d'Afanasievo puis de celle d'Andronovo. Les données génétiques obtenues sur les deux groupes anciens de l'âge du Bronze, issus du sud-sibérien et du bassin du Tarim vont donc dans le sens d'un peuplement par des populations venues du Nord des steppes (Keyser et al., 2009).

Ainsi, bien que les premières études aient permis d'apporter des éléments essentiels à la compréhension du peuplement de cette région, le nombre de données disponibles reste encore

relativement faible. Seules deux études d'ADN ancien se sont intéressées aux données du chromosome Y. De la même façon, très peu de données sont disponibles concernant les individus inhumés dans l'Altaï mongol. C'est dans ce contexte qu'a donc été initié ce travail de thèse.

I.5 Problématique et objectifs de l'étude

Le peuplement de la région du sud de la Sibérie et du massif de l'Altaï est un processus complexe. Nous avons vu que de nombreuses cultures aux origines parfois diverses s'y sont succédées. Si un consensus apparaît pour dire que des mouvements migratoires d'Ouest en Est ont eu lieu, de nombreuses questions restent en suspens. De quelles régions venaient ces populations ? Y a-t-il une homogénéité et une continuité génétique entre elles ? D'où venaient les individus présents dans l'Altaï mongol à l'âge du Bronze ? Existe-il un lien génétique entre les populations du sud de la Sibérie et de la Mongolie ?

Pour tenter d'apporter des éléments de réponse à toutes ces questions, nous avons étudié l'ADN d'individus issus de différentes cultures de l'âge du Bronze provenant à la fois du sud de la Sibérie et du massif de l'Altaï.

Ainsi, à partir des données moléculaires obtenues, la diversité génétique de chaque groupe culturel a pu être caractérisée. L'analyse comparative des données anciennes obtenues par notre équipe, mais aussi par d'autres, nous a permis d'appréhender les relations génétiques et les interactions possibles entre ces populations anciennes. De la même façon, la comparaison des différentes lignées maternelles et paternelles retrouvées chez ces individus anciens avec celles des populations eurasiennes actuelles, nous a permis d'émettre des hypothèses quant à l'origine biogéographique de ces individus.

Grâce à l'ensemble de ces résultats il a été possible d'apporter de nouveaux éléments quant à la compréhension de la dynamique de peuplement de la région ciblée.

Comme présenté dans la première partie, les études d'ADN ancien requièrent des précautions et une méthodologie particulières. Dans l'objectif d'étudier de façon « globale » la diversité de cette population, il a été nécessaire, au cours de ce travail, de mettre en place une stratégie d'analyse adaptée aux caractéristiques d'un ADN datant de plusieurs milliers d'années et

permettant l'étude de nombreux marqueurs (mitochondriaux, du chromosome Y et autosomaux) à mode de transmission et taux de mutation variables.

II MATÉRIEL ET MÉTHODES

II.1 Prélèvements biologiques anciens

Lors de ce travail de thèse, 69 échantillons (Tableau 1) provenant de différents sites couvrant une région allant de la Khakassie (sud de la Sibérie) au sud de l'Altai mongol ont été testés (Figure 7). Ces échantillons sont majoritairement datés de l'âge du Bronze. Ils ont été exhumés, pour la plupart d'entre eux, de kourganes (Figure 8). A l'exception de l'individu TA4 du site de Tsagaan Asga pour lequel nous disposons d'un fragment d'os long, tous les prélèvements biologiques mis à notre disposition étaient des dents.

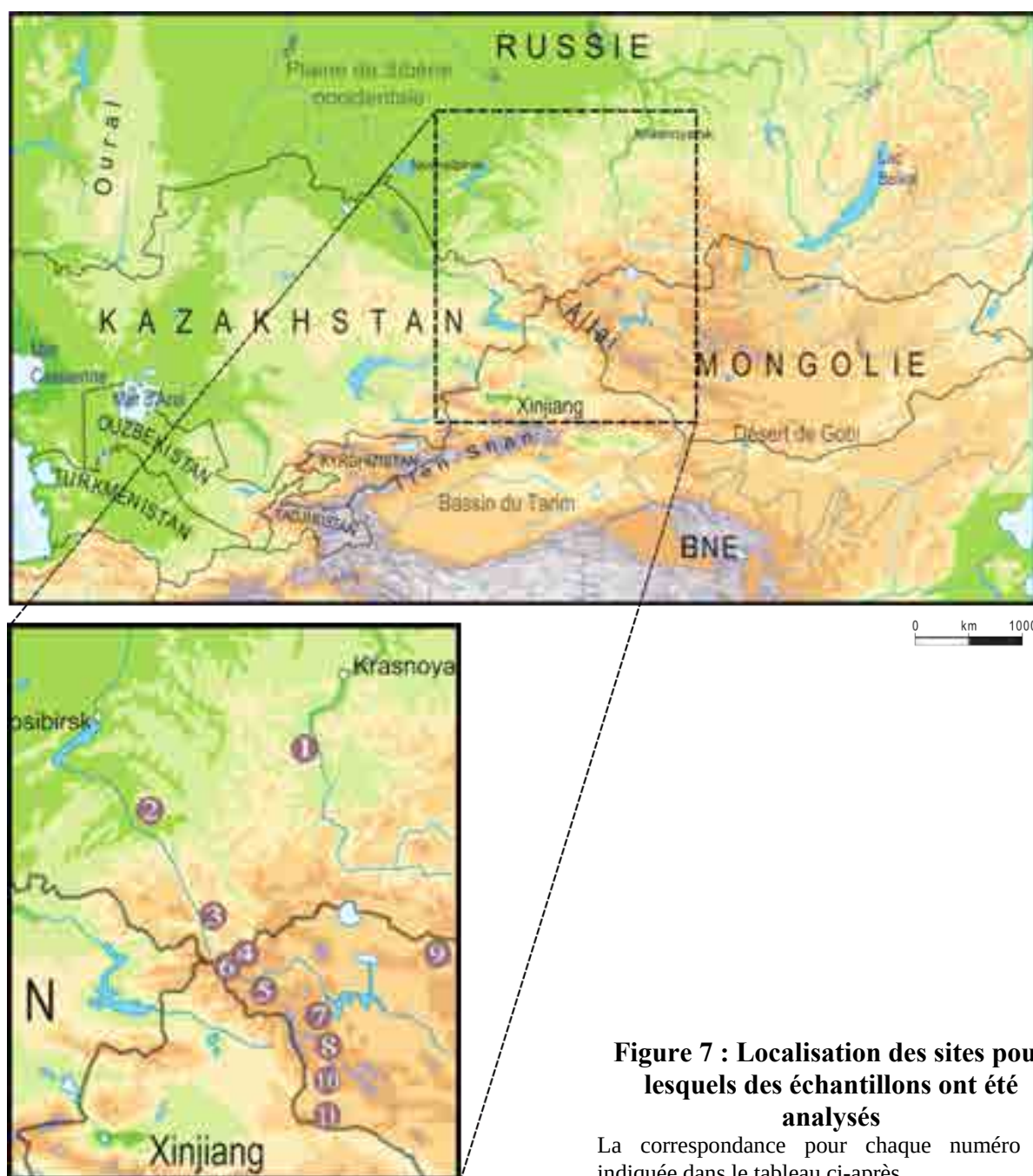


Figure 7 : Localisation des sites pour lesquels des échantillons ont été analysés

La correspondance pour chaque numéro est indiquée dans le tableau ci-après.

Tableau 1 : Liste des prélèvements disponibles pour ce travail 1/2

Localisation	N° Carte	Culture	Datation	Code labo	N	Nom échantillon	informations complémentaires
Khakasie, Sud Sibérie, Russie	1	Afanasievo	3500-2600 avant J.-C	Kaf	4	Kh19	Itkol'II, kourgane 24, tombe 1
						Kh20	Itkol'II, kourgane 23, tombe 2
						Kh21	Itkol'II, kourgane 27, tombe 1
						Kh22	Itkol'II, kourgane 27, tombe 1
		Okunevo	2300-1800 avant J.-C	Ok	18	Kh1	Krasny Kamen, kourgane 1, tombe 1
						Kh2	Krasny Kamen, kourgane 1, tombe 3
						Kh3	Krasny Kamen, kourgane 1, tombe 6
						Kh4	Uybat-Charkov, kourgane 1, tombe 1
						Kh5	Uybat-Charkov, kourgane 1, tombe 2
						Kh6	Uybat-Charkov, kourgane 1, tombe 3
						Kh7	Uybat-Charkov, kourgane 1, tombe 3
						Kh8	Uybat-Charkov, kourgane 1, tombe 6
						Kh9	Uybat-Charkov, kourgane 1, tombe 9
						Kh10	Uybat-Charkov, kourgane 1, tombe 11
						Kh11	Uybat-Charkov, kourgane 1, tombe 11
						Kh12	Uybat-Charkov, kourgane 1, tombe 12
						Kh13	Itkol'II, kourgane 14, tombe 2
						Kh14	Itkol'II, kourgane 14, tombe 3
						Kh15	Itkol'II, kourgane 14, tombe 5
						Kh16	Itkol'II, kourgane 14, tombe 7
						Kh17	Itkol'II, kourgane 14, tombe 8
						Kh18	Itkol'II, kourgane 14, tombe 9
Région de l'Altaï, Russie	2	Elunino	2300-1700 avant J.-C	El	5	6	Teleytskiy Vzvoz-I, tombe 7
						7	Teleytskiy Vzvoz-I, tombe 18
						8	Teleytskiy Vzvoz-I, tombe 21
						9	Teleytskiy Vzvoz-I, tombe 41
						10	Staroaleyka-2, tombe 64
République de l'Altaï, Russie	3	Bol'shemyskaya	Enéolithique	Bol	1	11	Tytkesken-VI, kourgane 61
		Afanasievo	3500-2600 avant J.-C	Aaf	5	1	Bersyukta-I, kourgane 1
						2	Tytkesken-VI, kourgane 95
						3	Saldyar-I, kourgane 25
						4	Saldyar-I, kourgane 36
						5	Choburak-III, kourgane 1

Tableau 1 : Liste des prélèvements disponibles pour ce travail 2/2

Localisation	N° Carte	Culture	Datation	Code labo	N	Nom échantillon	informations complémentaires
Nord de l'Altaï, Mongolie	4	Chemurcek	2500-2300 avant J.-C	ChemN	3	AT_626	Kulala ula, kourgane
						AT_628	Kumdi goovi, kourgane, tombe 2
						AT_635	Kurgak govi 2, kourgane, tombe 2
Nord de l'Altaï, Mongolie	5	Sagsai	1400-1100 avant J.-C	Sag	7	TA4	Tsagaan Asga, tombe 4
						TA8	Tsagaan Asga, tombe 8
						TA9	Tsagaan Asga, tombe 9
						TA10	Tsagaan Asga, tombe 10
						TA11	Tsagaan Asga, tombe 11
						TA12	Tsagaan Asga, tombe 12
						TA14	Tsagaan Asga, tombe 14
	6		1400- 900 avant J.-C		6	TU15	Takhilgat Uzuur-5, tombe 15
						TU17	Takhilgat Uzuur-5, tombe 17
						TU18	Takhilgat Uzuur-5, tombe 18, enfant
						TU25	Takhilgat Uzuur-5, tombe 25
						TU32	Takhilgat Uzuur-5, tombe 32
						TU34	Takhilgat Uzuur-5, tombe 34
			Afanasievo	2742 ± 92 avant J.-C	Maf	1	TU31
Khovd aimag, Mongolie	7	Chemurcek	2300-1800 avant J.-C	Chem	5	12	Ulaan Hudag-I, kourgane 12
						13	Ulaan Hudag-II, kourgane 3
						14	Ulaan Hudag-I, kourgane 12
						15	Halzan Uzuur-II, kourgane 2
						16	Halzan Uzuur-II, kourgane 4
Khovd aimag, Mongolie	8	Munkh-Khairkhan	1700-1400 avant J.-C	MK	1	AT_614	Ulaan goviin uzuur 2, kourgane
Zabkhan aimag, Mongolie	9	Munkh-Khairkhan	1700-1400 avant J.-C	MK	2	AT_861	Khukh khushoony bom 1, kourgane
						AT_862	Khukh khushoony bom 2, kourgane
Sud de l'Altaï, Mongolie	10	Chemurcek	2400-2000 avant J.-C	ChemS	8	AT_592 (1,2,3,4,5)	Yagshiin khodoo, kourgane 3
						AT_642	Keviin am, kourgane 2
						AT_640	Buural kharyn ar, kourgane
						ATNEW01	kheviin am, kourgane 1
Entre l'Altaï et Tianshan, Mongolie	11	Baitag	1200-900 avant J.-C	Bai	3	AT_630	Kheviin am, kourgane 1, tombe 3
						AT_672	Uliastain gol III, kourgane 7
						AT_681	Uliastain gol III, kourgane 2
nombre d'échantillons testés						69	



Photo E. Crubézy

Figure 8 : Tombe de l'individu TA10 du site de Tsagaan Asga

La répartition chronologique et géographique des 69 échantillons testés est présentée dans la Figure 9. Suivant une distribution spatiale, le groupe étudié se décompose de la façon suivante : 22 échantillons proviennent de la Khakassie, 11 de l'Altaï russe et 36 de Mongolie. Selon une distribution chronologique, 11 individus sont datés du début de l'âge du bronze ou avant, 39 sont datés du milieu de l'âge du Bronze (env. 2000 ans avant J.-C) et enfin 19 individus sont attribués à des cultures datant de la fin de l'âge du Bronze.

**Figure 9 : Répartition chronologique et géographique des prélèvements analysés**

II.1.1 Traitement des échantillons

Les dents et les os sont généralement les seuls restes disponibles pour retrouver de l'ADN. Nous avons choisi de travailler principalement à partir de prélèvements dentaires car ils présentent plusieurs avantages : (i) les dents ne sont pas totalement résistantes aux contaminations mais l'émail dentaire, du fait de sa faible porosité, est une bonne barrière protectrice contre d'éventuelles dégradations (UV, moisissures, microbes...), (ii) les prélèvements sont relativement faciles à décontaminer, et (iii) chaque dent peut être traitée de façon indépendante facilitant la réplication des analyses. Lorsque plusieurs prélèvements sont à notre disposition par individu, une observation macroscopique nous permet dans un premier temps de choisir le prélèvement qui paraît le mieux préservé. Pour les prélèvements dentaires par exemple, le choix se portera dans l'idéal, sur une dent non cariée, sans fissure et avec des racines fermées afin de limiter l'entrée de contaminants.

II.1.1.1 Préparation des échantillons

La première étape du traitement des échantillons consiste en une décontamination de surface. Cette étape est très importante surtout lorsque les conditions dans lesquelles ont été prélevés et stockés les échantillons ne sont pas connues.

Les prélèvements dentaires sont dans un premier temps débarrassés des éventuels résidus de terre à l'aide d'un scalpel stérile. Puis ils sont ensuite nettoyés à l'aide de compresses stériles légèrement imbibées d'eau de javel à 30% puis rincés avec de l'eau ultrapure et enfin exposés aux UV sur chaque face pendant 20 min. Cette étape est très délicate, la dent ne doit pas être trop imbibée car bien que l'émail procure une barrière protectrice, les prélèvements étant anciens, la dent s'est fragilisée. Il est possible que des fissures se soient formées ou que les racines ne soient pas totalement fermées, notamment chez les immatures, entraînant l'infiltration de contaminants ou des solutions de lavage pouvant ainsi léser l'ADN. Une fois nettoyée, la dent est prête à être réduite en poudre.

Pour les prélèvements osseux, la méthode est différente. Les contaminants pouvant s'infiltrer plus profondément dans l'os, jusqu'au moins 1 à 2 millimètres d'épaisseur (Bouwman et al., 2006), il ne suffit pas uniquement de nettoyer sa surface mais d'en enlever une partie. Pour cela, à l'aide d'une dremel, l'os ou le fragment osseux est abrasé mécaniquement sur toute sa surface sur quelques millimètres de profondeur. Cette étape est réalisée sous une hotte à

extraction dédiée à cet effet. La partie ainsi nettoyée est ensuite utilisée pour être réduite en poudre.

Afin d'être réduits en poudre, les prélèvements décontaminés sont placés dans un cryobroyeur à azote liquide (6870 FREEZER/MILL[®], SPEX Sample Prep[®]). Le froid permet d'éviter les échauffements qui peuvent induire une dégradation des molécules d'ADN durant cette étape. Les prélèvements sont ensuite stockés dans une salle dédiée aux échantillons anciens à l'abri de la lumière et dans des conditions stables de température. Ainsi, il est possible de collecter à partir d'une dent, de quelques centaines de milligrammes de poudre (dent d'immature), à 2 grammes (molaire d'adulte). A partir d'un prélèvement osseux, la quantité récupérée est souvent plus importante, de l'ordre de plusieurs grammes.

II.1.1.2 Extraction de l'ADN

II.1.1.2.1 Extraction de l'ADN à partir d'échantillons modernes

Les échantillons modernes utilisés dans ce travail sont ceux des individus ayant participé à l'étude de la détermination des traits phénotypiques ou aux fouilleurs et manipulateurs ayant été en contact avec les échantillons. Les prélèvements d'ADN sont réalisés par cytobrossage buccal. Conformément à la législation française, un consentement éclairé a été signé par toutes les personnes impliquées.

A partir de la cytobrosse, l'extraction est réalisée à l'aide du kit QiAamp[®] DNA Mini kit de Qiagen. Le protocole utilisé est celui décrit par le fabricant pour l'extraction d'ADN à partir de cellules buccales. Cette extraction est réalisée dans des locaux totalement séparés de ceux dédiés aux analyse de l'ADN ancien (dans le même bâtiment mais à un autre étage).

II.1.1.2.2 Extraction de l'ADN à partir des échantillons anciens

A partir de la poudre obtenue après cryobroyage des tissus durs anciens nous pouvons procéder à l'extraction d'ADN. Cette étape est critique dans le processus de traitement des échantillons car il s'agit d'obtenir un rendement maximal d'extraction puisque l'ADN n'est parfois présent qu'en très faible quantité au sein du prélèvement. L'étape de purification associée à l'extraction doit également permettre d'éliminer les éventuels inhibiteurs présents dans l'échantillon.

La méthode utilisée dans cette étude est la suivante. La première étape du protocole consiste en une décalcification et une lyse des échantillons. Pour cela 150 à 200 mg de poudre sont placés dans un tampon constitué de : 500 µl d'EDTA à 0,5M (chélateur des ions calcium), 50 mg de protéinase K à 20 mg.ml⁻¹ et 5 µl de DTT à 1M (agent réducteur). L'ensemble est placé une nuit sous agitation à 50°C. La seconde étape consiste à purifier l'ADN. Pour cela le kit NucleoSpin® Gel and PCR Clean-up commercialisé par la société Macherey Nagel a été utilisé. Ce kit est destiné à la purification directe des produits de PCR ou à partir de gels d'agarose. Il a la propriété de purifier des petits fragments d'ADN (jusqu'à 50 pb), idéal donc pour la purification d'ADN dégradé comme c'est souvent le cas avec l'ADN extrait des échantillons anciens. Le principe de la purification est basé sur l'utilisation d'une colonne avec une matrice de silice. L'ADN est mélangé à un tampon contenant des sels chaotropiques permettant la fixation de l'ADN sur la membrane de silice puis les éléments indésirables issus de la lyse sont éliminés par des lavages successifs à base d'éthanol. Enfin, l'ADN est élué dans 250 µl d'eau ultrapure. Les extraits sont ensuite purifiés et concentrés par ultrafiltration grâce aux colonnes Amicon Ultra 0.5 ml 30 KDa, (Millipore). Le volume final d'extrait est ainsi de 50 µl.

II.1.2 Les analyses génétiques réalisées

A partir des extraits d'ADN, nous avons pu réaliser différentes analyses génétiques. Une stratégie expérimentale particulière a été mise en place afin de répondre aux différentes problématiques de recherche tout en consommant le moins d'ADN possible puisque les substrats anciens disponibles étaient limités. Plusieurs marqueurs ont été étudiés, des marqueurs à transmission uniparentale c'est-à-dire hérités uniquement de la mère ou du père localisés sur l'ADN mitochondrial et la région non recombinante du chromosome Y ainsi que des marqueurs à transmission biparentale localisés sur les autosomes.

II.1.2.1 La détermination du sexe de l'individu

La détermination génétique ou moléculaire du sexe des individus va permettre de compléter et de confirmer les données anthropologiques quand celles-ci sont disponibles. En effet, la diagnose sexuelle est parfois difficile à établir lorsque les restes osseux sont très fragmentés et mélangés, ou lorsque les individus sont immatures. Dans ces cas précis, la détermination du sexe apportera des éléments essentiels à l'étude du recrutement funéraire

La technique choisie est celle utilisée en médecine légale qui consiste à amplifier une région du gène de l'amélogénine. L'amélogénine intervient dans la synthèse de l'émail dentaire. Le gène codant cette protéine est présent sur les chromosomes sexuels (AMGX et AMGY) avec, sur le chromosome X une délétion de 6pb par rapport au chromosome Y (Sullivan et al., 1993). Ainsi l'amplification de cette région donnera des profils électrophorétiques différents pour un homme et une femme. Les amorces utilisées sont proches de la région délétée et permettent l'amplification d'un fragment de 112 pb sur le chromosome Y et de 106 pb sur le chromosome X. La détection de deux fragments à 106 pb indique que l'individu testé est féminin (XX) alors que la détection d'un fragment à 106 pb et d'un second à 112 pb indique qu'il est masculin (XY). Ces amorces sont présentes dans tous les kits d'identification génétique commercialisés et permettent ainsi de disposer d'une empreinte ou profil génétique avec le sexe de l'individu analysé.

Les résultats erronés sont rares avec cette méthode mais néanmoins parfois observés notamment dans certaines populations, du fait de délétions (Cadenas et al., 2007) ou lorsque l'ADN analysé est présent en trop faible quantité ou est très dégradé (Quincey et al., 2013), amenant ainsi des pertes d'allèles (dropout en anglais). C'est pourquoi, afin d'augmenter la fiabilité de la détermination du sexe, des marqueurs spécifiques du chromosome Y sont inclus dans les nouveaux kits d'identification, par exemple dans le kit GlobalfilerTM commercialisé par la société Life technologies.

II.1.2.2 Etablissement du profil génétique des individus

Un profil génétique est généralement établi à partir de l'amplification simultanée de STR (Short Tandem Repeats) autosomaux. Les STR sont des polymorphismes de longueurs, c'est-à-dire des répétitions de motifs de quelques paires de bases. Ce nombre de répétitions peut varier d'un individu à l'autre. L'allèle obtenu correspondra ainsi au nombre de répétitions. C'est le nombre de STR analysés qui va permettre d'obtenir un pouvoir de discrimination plus ou moins fort assurant l'unicité de chaque profil. Dans ce travail, le kit AmpFLSTR® Identifiler® Plus (Life technologies), employé dans les investigations médico-légales, a été utilisé pour l'établissement des profils génétiques. Il permet l'amplification simultanée de 15 STR autosomaux (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, THO1, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA) ainsi que du locus de l'amélogénine conférant un pouvoir de discrimination de 5.01×10^{-18} quand l'ensemble des marqueurs a pu être analysé.

Ces marqueurs étant localisés sur les autosomes, chaque individu possède pour chaque locus 2 allèles, l'un provenant de son père et l'autre de sa mère. L'étude de ces marqueurs permet donc également d'étudier les liens de proches parentés.

Dans ce travail, la réaction d'amplification a été réalisée à partir de 5 µl d'ADN dans un volume final de 12.5 µl. Le protocole suivi est celui recommandé par le fabricant à l'exception du nombre de cycles qui est augmenté de 28 à 32. L'analyse des produits PCR est réalisée par électrophorèse capillaire sur un séquenceur automatique 3500 Genetic Analyser (Life technologies). Les résultats sont analysés à l'aide du logiciel Gene Mapper v4.1 (Life technologies).

II.1.2.3 Détermination des haplotypes et des haplogroupes mitochondriaux et du chromosome Y

Contrairement aux marqueurs autosomaux, les marqueurs localisés sur l'ADN mitochondrial et sur la région non recombinante du chromosome Y sont à transmission uniparentale. L'ADN mitochondrial se transmet de la mère à ses enfants (filles et garçons) et le chromosome Y se transmet uniquement de père en fils. Ainsi l'analyse de ces marqueurs permet d'étudier les lignées maternelles et paternelles des individus. L'étude de ces lignées se fera par la détermination des haplotypes ainsi que des haplogroupes Y et mitochondriaux, deux notions employées fréquemment dans ce manuscrit et définies ci-après.

L'haplotype

Un haplotype correspond à une combinaison de mutations présentes sur une même région d'intérêt (l'ADN mitochondrial) ou sur un même chromosome (chromosome Y). Les polymorphismes concernés peuvent être aussi bien des polymorphismes ponctuels de séquence (SNP : Single Nucléotide Polymorphism), des insertions/délétions ou encore des polymorphismes de longueurs (STR). Les individus partageant la même signature génétique sur l'ensemble de ces sites polymorphes portent alors le même haplotype. L'haplotype mitochondrial est défini par des SNP alors que l'haplotype du chromosome Y est défini par des STR. Ces haplotypes peuvent être regroupés au sein de grands groupes appelés haplogroupes.

L'haplogroupe

Un haplogroupe regroupe les haplotypes dérivant d'un même ancêtre commun. Les haplotypes appartenant au même haplogroupe partagent ainsi souvent des mutations en commun dans le cas de l'ADN mitochondrial. Concernant le chromosome Y, si l'haplogroupe est ici défini par les allèles présents au niveau de différents loci STR, l'haplogroupe lui est généralement défini par un SNP. Ces haplogroupes ayant le plus souvent des répartitions géographiques spécifiques, leur étude rend possible la détermination de l'origine biogéographique d'un individu. Aujourd'hui, les avancées technologiques ont permis d'affiner considérablement la phylogénie, entraînant des modifications fréquentes des noms des haplogroupes (principalement des sous-haplogroupes, la désignation des macro-haplogroupes restant constante). Ces noms sont construits par l'enchaînement de lettres et de chiffres. Les haplogroupes du chromosome Y étant défini par des SNP, nous associerons toujours le nom du SNP à l'haplogroupe pour faciliter la compréhension en dépit de l'évolution constante de la nomenclature.

II.1.2.3.1 Analyse de l'ADN mitochondrial

L'ADN mitochondrial (ADNmt) est un ADN extranucléaire circulaire de 16569 pb contenu au sein des mitochondries, organites cellulaires impliqués dans la respiration cellulaire. Il peut être séparé en deux régions distinctes : (i) une région non codante d'environ 1100 pb, appelée aussi région de contrôle et (ii) une région codante constituée de gènes codants principalement pour des ARN ribosomiaux et de transfert ainsi que des protéines impliquées dans la phosphorylation oxydative (Figure 10).

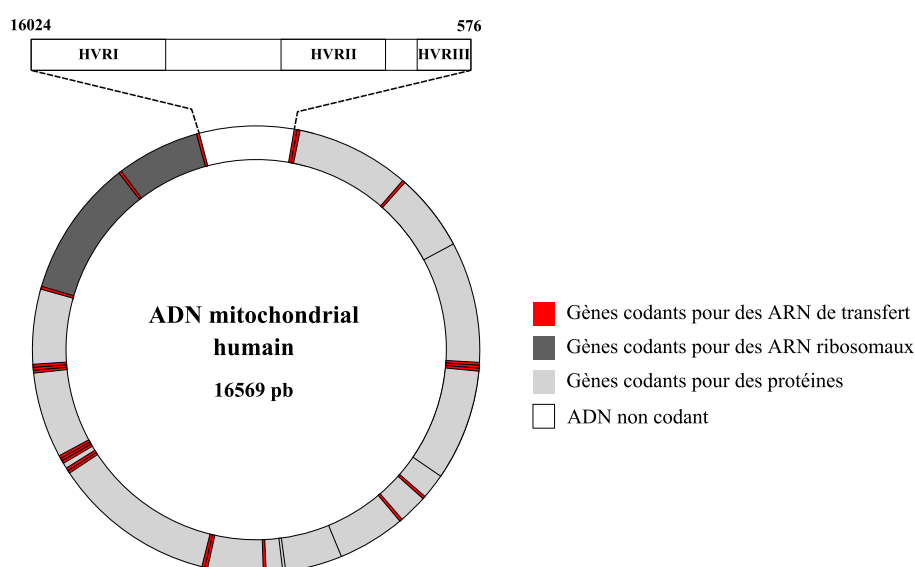


Figure 10 : Représentation schématique de l'ADN mitochondrial humain

Plusieurs caractéristiques font de cet ADN un support de choix pour les études d'ADN ancien. L'ADNmt est présent en un nombre de copies élevé, il se transmet uniquement par voie maternelle *via* l'ovocyte, et passe de la mère à l'enfant sans recombinaison. Les variations de séquences apparues au niveau de la séquence d'ADNmt sont donc le fruit d'événements aléatoires qui apparaissent au cours du temps et qui sont transmis de la mère à l'enfant. Le taux de mutation n'est pas uniforme sur toute la séquence de l'ADNmt. La région non codante contient trois régions hypervariables appelées HV1, HV2 et HV3. Ce sont généralement ces régions, notamment les régions HV1 et HV2, qui sont classiquement analysées dans les études de génétique des populations puisque leur fort taux de mutations permettra d'observer des différences entre les populations.

L'ADNmt a été séquencé entièrement en 1981 (Anderson et al., 1981). Cette séquence appelée CRS (Cambridge Reference Sequence) a ainsi été utilisée comme séquence de référence. Une version corrigée (rCRS = revised Cambridge Reference Sequence) a été publiée quelques années plus tard (Andrews et al., 1999). Les études de comparaison se font donc par rapport à cette séquence de référence révisée. Récemment des auteurs ont proposé de remplacer cette séquence de référence par une séquence appelée « Reconstructed Sapiens Reference Sequence » (RSRS) située à la racine de l'arbre phylogénétique des *Homo sapiens* (Behar et al., 2012). En effet, les événements mutationnels sont généralement reportés par rapport à la séquence rCRS or cette séquence est celle d'un européen moderne. La séquence RSRS correspondrait elle à une séquence ancestrale plutôt qu'à une séquence « périphérique » de la phylogénie. Cependant, l'ensemble des données disponibles que ce soit dans la littérature ou dans les bases de données telles que EMPOP (<http://empop.org>) sont annotées en fonction de la séquence rCRS. Afin de garder une cohérence et de faciliter les analyses, nous travaillerons donc toujours par comparaison avec la séquence rCRS.

Le séquençage de la région HV1 de l'ADN mitochondrial

Pour déterminer les haplotypes mitochondriaux de nos échantillons, nous avons séquencé une région de 381pb localisée sur la région hypervariable 1 (HV1) de l'ADN mitochondrial. L'état généralement fragmenté de l'ADN étudié ne permet pas d'amplifier directement ce fragment en une seule fois, nous avons donc procédé au séquençage de deux fragments (HV1-A et HV1-B) chevauchants de 250 et 220 pb correspondant aux positions nucléotidiques 15989 à 16239 et 16190 à 16410. Le fait d'avoir des séquences chevauchantes permet

également d'assurer un contrôle au niveau de la zone de chevauchement. Les amorces utilisées sont les suivantes :

HV1-A_F15989 : CCCAAAGCTAAGATTCTAAT

HV1-A_R16239 : TGGCTTTGGAGTTGCAGTTG

HV1-B_F16190 : CCCCATGCTTACAAGCAAGT

HV1-B_R16410 : GAGGATGGTGGTCAAGGGAC

Les réactions d'amplifications ont été effectuées de manière indépendante dans un volume final de 50 μl contenant 10 μl d'extrait d'ADN, 1 μl de chaque amorce (10 μM), 10 μl de pré-mélange réactionnel 5X (5 μl de tampon HGS, 3 μl de MgCl_2 25 mM, 0.5 μl de BSA 20 mg/ml, 0.1 μl de chaque dNTP 100 mM, 1,1 μl d'eau ultra pure), 0.5 μl de HGS Diamond Taq® (Eurogentec) à 5 $\text{U} \cdot \mu\text{l}^{-1}$ et de l'eau ultrapure QSP 5 μl .

Les réactions d'amplifications ont ensuite été réalisées avec les paramètres suivants : une dénaturation initiale/activation à 94°C pendant 13 min, suivie de 38 cycles composés d'une étape de dénaturation à 94°C pendant 30 sec, d'une étape d'hybridation à 48°C pendant 30 sec et d'une étape d'élongation à 72°C pendant 45 sec. La réaction est finalisée par une étape d'élongation à 72°C pendant 5min. Pour le fragment HV1-B le programme utilisé est le même à l'exception de la température d'hybridation qui est de 51°C.

Afin de vérifier la taille des produits amplifiés et d'estimer leur concentration, 10 μl de produits PCR ont été déposés sur gel d'agarose 2,5%. La concentration des produits d'amplification a été estimée par l'intermédiaire du marqueur de taille MassRuler™ DNA Ladder, low range (Fermentas). Les produits PCR sont ensuite purifiés à l'aide du kit NucleoSpin® Gel and PCR Clean-up, Macherey Nagel. Cette étape permet d'une part, d'éliminer les composants résiduels provenant du milieu réactionnel de la PCR autre que l'ADN (amorces, sels, dNTP), et d'autre part, de concentrer les produits PCR à la concentration requise (2 $\text{ng} \cdot \mu\text{l}^{-1}$) pour la PCR de séquençage par ajustement du volume d'élution.

La PCR de séquençage est réalisée à l'aide du kit BigDye® Terminator V1.1 Cycle Sequencing (Life technologies). Le milieu réactionnel est composé de 2 μl de produits PCR purifiés et concentrés à 2 $\text{ng} \cdot \mu\text{l}^{-1}$, 1 μl d'amorces à 0.8 μM , 1 μl de mix 5X de 2 μl de BDT et d'eau ultrapure QSP 10 μl .

Les produits sont ensuite purifiés par une précipitation à l'éthanol. Les 10 µl de la réaction de séquence sont déposés dans 40 µl d'une solution d'éthanol à 80% préalablement placée au congélateur pendant 2 heures minimum. L'ensemble est ensuite déposé à -20°C pendant 15 min. Puis les tubes sont centrifugés 30 min à 12000 g. Le surnageant est prélevé et le culot est rincé dans une solution d'éthanol à 70% puis agité légèrement et centrifugé 5 min à 12000 g. Enfin, le surnageant est éliminé et le culot est laissé à l'air libre pendant plusieurs heures pour éliminer toutes traces d'éthanol qui pourraient entraver la lecture des séquences d'ADN. Avant séquençage, le culot est resuspendu dans 10 µl d'eau ultrapure puis vortexé et centrifugé 15 min après.

Afin de réaliser l'électrophorèse capillaire 4 µl de chaque échantillon sont ajoutés à 16 µl d'eau ultrapure. L'électrophorèse est réalisée sur un séquenceur automatique 3500 Genetic Analyser (Life technologies). Les séquences sont analysées grâce au logiciel SequencherTM 4.8 (Gene Codes Corporation). Une séquence consensus HV1 est construite pour chaque échantillon après un minimum de deux extractions et deux amplifications. Les haplotypes HV1 sont ensuite obtenus par comparaison à la séquence de référence rCRS. A partir de ces haplotypes, les haplogroupes sont déterminés à l'aide du logiciel haplogrep (<http://haplogrep.uibk.ac.at/>), des données de la littérature et du site phylotree reportant la phylogénie actuelle (<http://www.phylotree.org/>).

Typage des SNP de la région codante de l'ADN mitochondrial

Des positions polymorphes de la région codante de l'ADN mitochondrial ont été sélectionnées à partir du site « phylotree » et des données de la littérature afin de confirmer l'appartenance aux haplogroupes préalablement déduits du séquençage de la région HV1.

Au vu du nombre important d'haplotypes obtenus au sein parmi les échantillons anciens étudiés, nous avons choisi de typer uniquement les SNP spécifiques des macro-haplogroupes. Les 13 SNP sélectionnés (Figure 11) ont été analysés dans une réaction multiplexe. Le typage a été réalisé par spectrométrie de masse MALDI-TOF dont le principe est présenté dans la section II.2.2.4. Les amorces utilisées ont été dessinées à l'aide du logiciel Assay Design 4.0 (Sequenom). Les amorces ainsi que la taille des amplicons attendus sont présentés en Annexe 1.

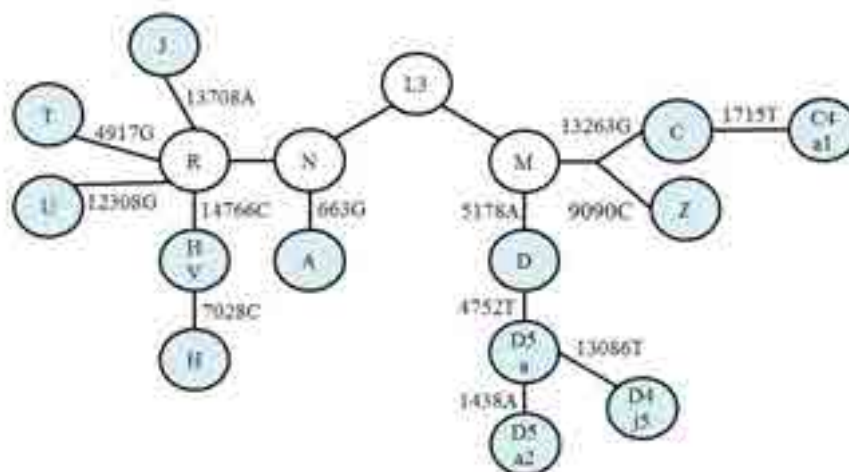


Figure 11 : SNP de la région codante de l'ADN mitochondrial

Arbre phylogénétique représentant les 13 SNP de la région codante de l'ADN mitochondrial analysés en une seule réaction multiplexe par spectrométrie de masse MALDI-TOF. Les haplogroupes testés sont indiqués dans les ronds bleus. Les positions des SNP sur l'ADN mitochondrial ainsi que l'allèle muté par rapport à la séquence de référence rCRS sont indiqués sur chaque branche.

II.1.2.3.2 Détermination des haplotypes et haplogroupes Y

Les STR du chromosome Y

Le chromosome Y est la marque du sexe masculin. Il est transmis de façon uniparentale de père en fils. Environ 95% échappe à la recombinaison méiotique; cette région est appelée NRY (Non Recombining Y). Les marqueurs étudiés se trouvent sur cette région, ils sont ainsi liés et se transmettent d'un seul bloc de générations en générations permettant ainsi de retracer les lignées paternelles.

La détermination des haplotypes Y pour les individus masculins a été réalisée grâce à l'amplification simultanée de 17 STR (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATAH4, DYS437, DYS438, DYS448) localisés sur la partie non recombinante du chromosome Y à l'aide du kit AmpFLSTR® Yfiler® (Life technologies).

Les réactions d'amplifications ont été réalisées à partir de 5 µl d'extrait d'ADN dans un volume final de 12,5 µl. Le protocole suivi est celui donné par le fabricant à l'exception du nombre de cycles qui a été augmenté de 30 à 34. L'électrophorèse capillaire est réalisée sur un

séquenceur automatique 3500 Genetic Analyser (Life technologies) et l'analyse est réalisée à partir du logiciel GeneMapper[®] 4.1 (Life technologies).

Ainsi, pour chaque individu masculin, l'ensemble des allèles obtenus pour les 17 loci STR correspond à l'haplotype Y. A partir de ces haplotypes, les haplogroupes sont déduits à l'aide de l'outil informatique Haplogroup Predictor disponible en ligne (<http://www.hprg.com>). Ce programme permet de prédire l'haplogroupe à partir de l'haplotype et d'y attribuer un score de confiance. Néanmoins, seul le typage des SNP d'intérêt permet d'assurer l'appartenance à un haplogroupe (Muzzio et al., 2011).

Les SNP du chromosome Y

Les SNP analysés ont été sélectionnés en fonction des prédictions obtenues à partir des haplotypes. Afin d'affiner les résultats, nous avons également sélectionné des SNP spécifiques des sous-haplogroupes d'intérêts (Figure 12).

Les 27 SNP sélectionnés n'ayant pu être analysés au sein d'une seule réaction multiplexe, deux groupes ont été constitués de la façon suivante : au sein d'une réaction ont été regroupés les 13 SNP spécifiques de l'haplogroupe R et de ses sous-haplogroupes, et au sein de la seconde réaction, les 14 SNP restant, spécifiques des haplogroupes C, D, N et Q ainsi que leur sous-haplogroupes. De cette façon, sur la base des haplotypes déduits des STR nous n'avions plus qu'à utiliser l'une ou l'autre des réactions multiplexes pour confirmer l'haplogroupe et essayer d'affiner cette affiliation.

Les SNP ont été typés par spectrométrie de masse MALDI-TOF. Le principe de la méthode est présenté dans la section II.2.2.4. Les amorces utilisées ont été dessinées à l'aide du logiciel Assay Design 4.0 (Sequenom). Les amorces, ainsi que la taille des amplicons attendue, sont présentés en Annexe 1.

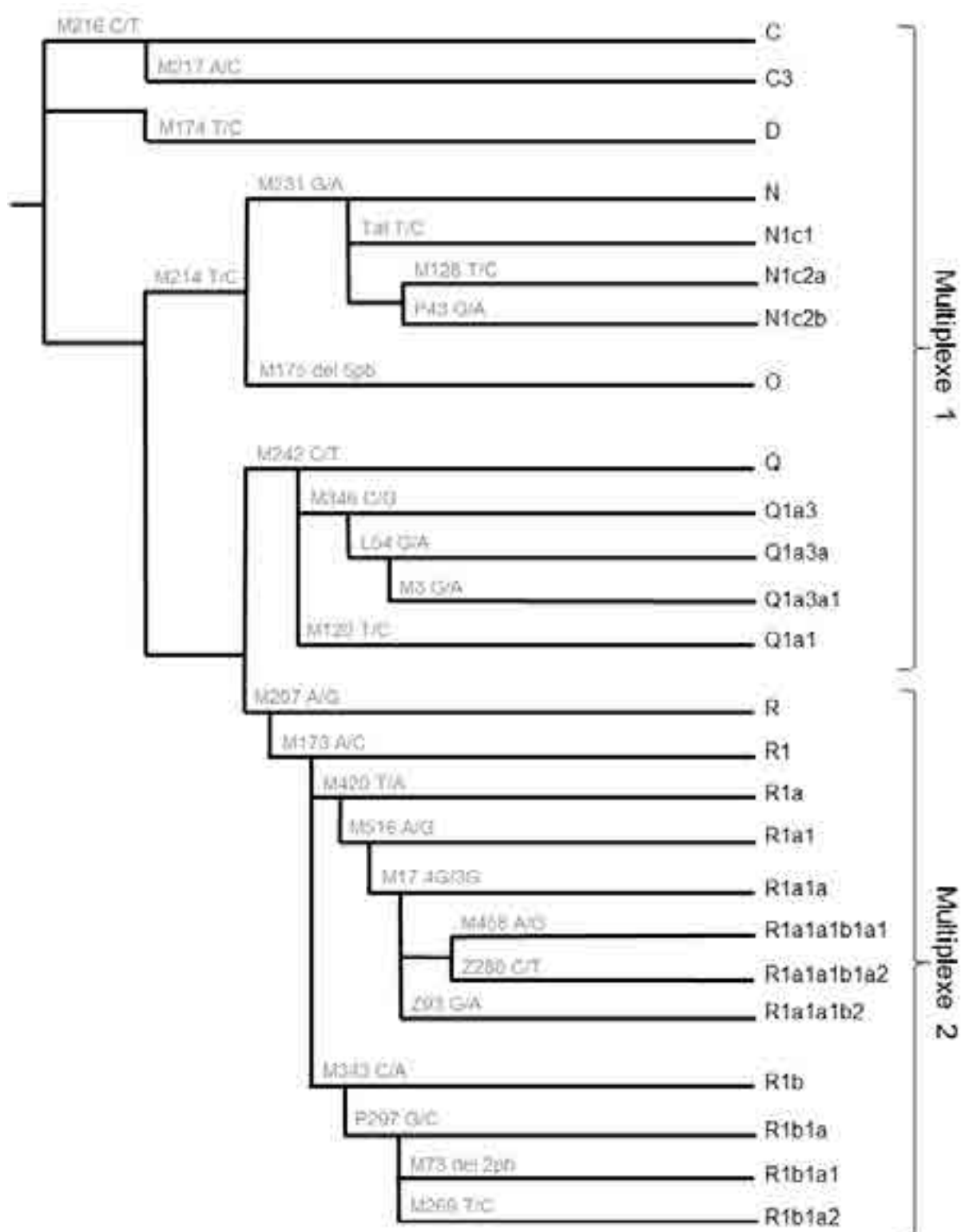


Figure 12: Arbre phylogénétique présentant les SNP du chromosome Y analysés

Les haplogroupes sont indiqués au bout de chaque branche ainsi que les marqueurs et les polymorphismes associés. La longueur des branches n'est pas significative. Le marqueur M434 (G/A), n'apparaissant pas dans l'arbre phylogénétique, mais décrit comme étant spécifique de certaines populations d'Asie du Sud, a également été analysé et inclus dans la réaction multiplexe 2.

II.1.2.4 Typage des SNP par spectrométrie de masse MALDI-TOF

De nombreuses techniques de génotypage des SNP sont aujourd'hui disponibles. Certaines permettent de génotyper quelques SNP et/ou échantillons tandis que d'autres sont utilisées pour le génotypage à très haut débit. La plupart de ces méthodes suivent le même principe c'est-à-dire une réaction de discrimination allélique suivie d'une détection des variants alléliques. Parmi toutes les méthodes disponibles, nous avons utilisé une méthode basée sur une discrimination allélique par extension d'amorces suivie d'une détection par spectrométrie de masse. Cette stratégie a été choisie car elle répondait à plusieurs critères nécessaires aux études d'ADN ancien : (i) le volume d'ADN nécessaire pour établir un génotype devait être faible (quelques μl) (ii) la méthode devait être sensible (de l'ordre d'une centaine de pg), reproductible et permettre le génotypage simultané de plusieurs SNP.

II.1.2.4.1 Principe de la spectrométrie de masse MALDI-TOF (Matrix Assisted Laser Desorption/Ionisation - Time Of Flight).

La technique MALDI-TOF utilise une irradiation par un faisceau laser UV à 337 nm. L'irradiation provoque un échauffement local qui permet le transfert des charges positives de la matrice à l'échantillon (ionisation). Une micro-explosion désorbe le mélange et provoque son expansion en phase gazeuse. Une fois formés, les ions sont accélérés par un champ électrostatique appliqué dans la zone d'accélération et envoyés dans un tube sous vide dans lequel ils sont séparés en fonction de leur rapport masse/nombre de charge (m/z). Comme le temps de vol est proportionnel à la masse, les ions de masse faible volent plus vite que les ions de masse plus importante, et sont ainsi détectés les premiers. Le détecteur collecte les ions et quantifie leurs intensités qui seront amplifiées et visualisées sous forme d'un spectre (Figure 13). La masse moléculaire est ainsi calculée à partir du rapport m/z . Il s'agit donc d'une méthode directe de mesure d'une propriété intrinsèque des molécules : la masse.

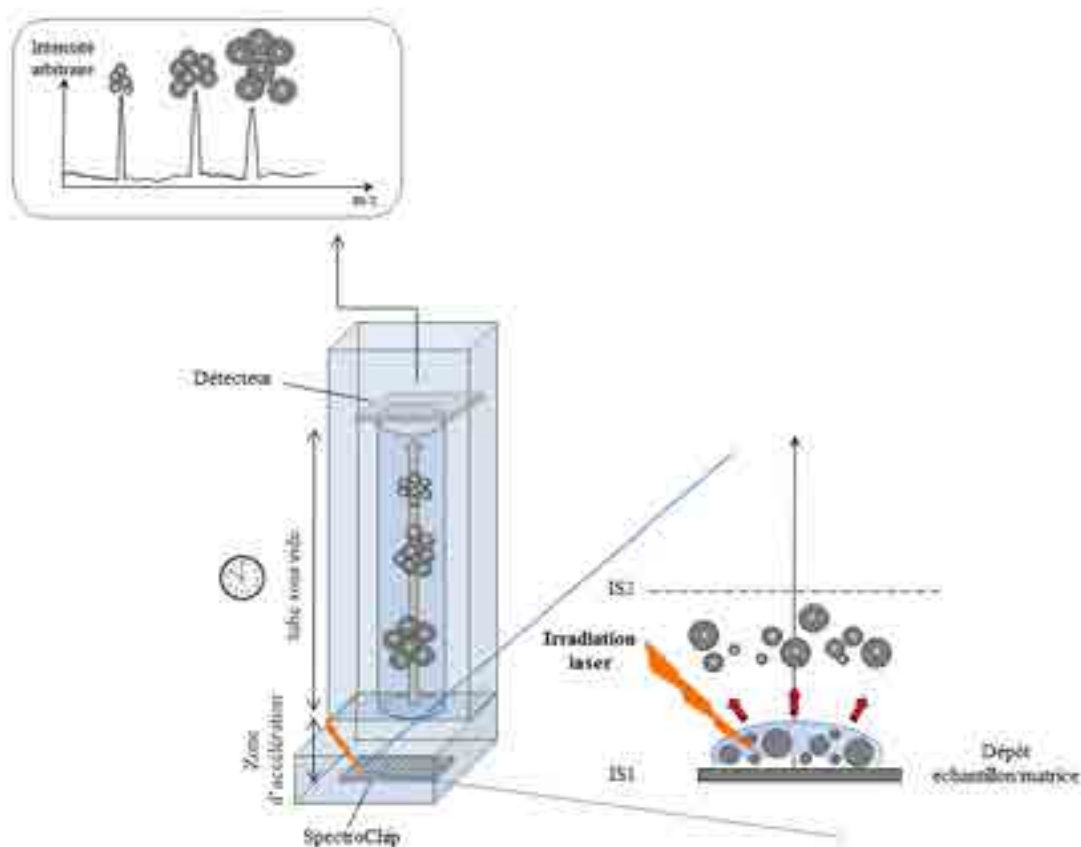


Figure 13 : Principe de la spectrométrie de masse MALDI-TOF

L'irradiation du faisceau laser incident va désorber et ioniser le mélange échantillon/matrice co-cristallisé sur la SpectroChip. Les produits d'extensions vont être entraînés en phase gazeuse dans la zone d'accélération avec la matrice, puis accélérés grâce à une différence de potentiel entre la cible IS1 et la grille d'extraction IS2 avant de pénétrer dans un tube de vol libre de champ où ils vont voler vers le détecteur. Le temps de vol sera d'autant plus court que les molécules sont légères. Leur analyse sera ensuite visible sur un spectre où les molécules apparaîtront en fonction de leur rapport masse/charge.

II.1.2.4.2 Avantages de la spectrométrie de masse MALDI-TOF et du système Mass Array de Sequenom.

La mesure directe de la masse des molécules représente un avantage par rapport à d'autres techniques de mesure indirecte utilisant par exemple, des marqueurs fluorescents. En effet, la détection d'une propriété intrinsèque des molécules permet d'assurer des mesures indépendantes des conditions réactionnelles et ainsi d'augmenter la fiabilité des résultats. De plus, comme la technique utilise des amorces non marquées, le coût est diminué. La spectrométrie de masse présente d'autres avantages : elle utilise un instrument de haute précision capable de détecter des produits de masses ne différant que de quelques daltons

(jusqu'à 9 Da), et sa large fenêtre de détection rend possible l'analyse simultanée de différents marqueurs (multiplexage).

La spectrométrie de masse MALDI-TOF étant une méthode développée principalement pour l'analyse protéique, le laboratoire s'est orienté vers une méthode conçue spécialement pour l'analyse des acides nucléiques, la plateforme Mass Array commercialisée par la société Sequenom. Plusieurs applications sont possibles sur cette plateforme dont le génotypage de SNP à l'aide de la technologie iPLEX[®] Gold (Sequenom) (Gabriel et al., 2009). Cette méthode s'est révélée particulièrement intéressante dans les analyses d'ADN ancien puisqu'à partir de quelques microlitres d'extrait, un nombre important de SNP (jusqu'à 40) peut être analysé à partir de court fragment d'ADN (environ 100 pb). Même si cette méthode est optimale pour une quantité d'ADN de l'ordre du nanogramme, il semble possible d'obtenir des génotypes complets jusqu'à 100pg d'ADN (Mendisco et al., 2011).

II.1.2.4.3 Protocole de typage des SNP

Le protocole utilisé pour les différentes analyses réalisées dans ce travail est celui recommandé par le fabricant. Il est composé des étapes suivantes (Figure 14) : (i) l'amplification des régions contenant les SNP d'intérêt est réalisée par PCR multiplexe à partir de 3.4 µl d'ADN, (ii) les dNTP non incorporés durant la PCR sont déphosphorylés et neutralisés par la SAP (Shrimp Alkaline Phosphatase), (iii) la réaction de discrimination allélique est réalisée à l'aide d'une amorce s'hybridant en amont du SNP d'intérêt (amorce SBE pour Single Base Extension), l'amorce est étendue par un ddNTP de masse modifiée en fonction de l'allèle présent sur la séquence cible, (iv) les échantillons sont dilués dans 41 µl d'eau ultrapure afin d'éliminer les surfactants, (v) les produits sont dessalés à l'aide d'une résine, (vi) une dizaine de nanolitres des produits sont déposés à l'aide d'un robot (Nanodispenser, Sequenom) sur une micropuce 96 puits (SpectroCHIP, Sequenom) contenant la matrice constitué d'acide 3-hydroxypicolinique (3-HPA) parfaitement adaptée aux études d'ADN (Tang et al., 1993), (vii) la micropuce est ensuite introduite dans le spectromètre de masse MALDI-TOF (Mass Array Analyser Compact 96, Sequenom) afin d'être analysée et enfin, (viii) la mesure de la masse de chaque produit permet d'attribuer les génotypes puisqu'une série d'applications spécifiques (Mass Array Workstation, Sequenom) permet de préparer l'analyse en assimilant à chaque expérience : les échantillons traités, les réactions de génotypage ainsi que toutes les masses attendues avec, pour chacune d'entre elles la correspondance en terme de génotypes. Dans la fenêtre d'analyse n'apparaissent que les

amorces SBE et les produits d'extension. Les amorces de l'amplification initiale sont dessinées avec une extension de 10 nucléotides augmentant ainsi leur masse et les fait de ce fait sortir de la fenêtre de détection lors de l'analyse par spectrométrie de masse.

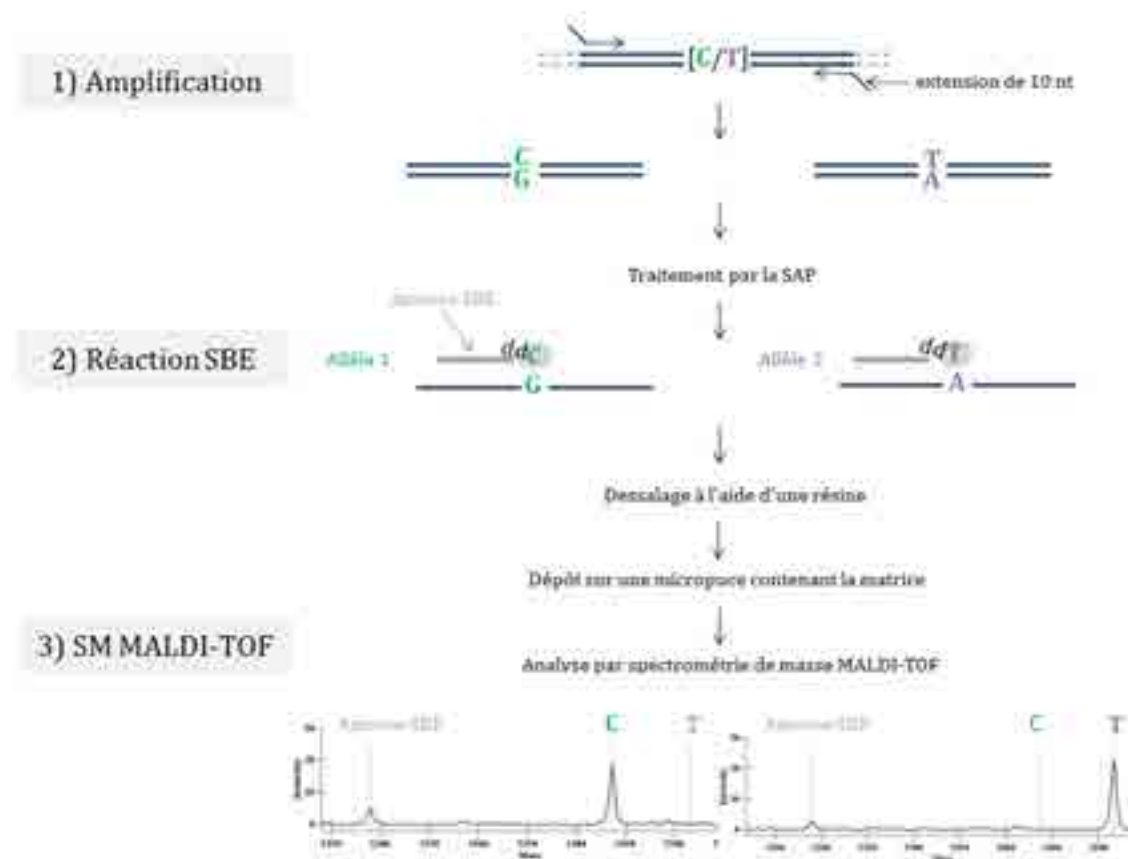


Figure 14: Principe de la méthode iPLEX Gold de Sequenom

II.1.2.5 Analyse des SNP autosomaux : détermination des caractères phénotypiques et de l'origine biogéographique des individus

Afin d'apporter des informations sur l'apparence physique des individus (couleurs des cheveux et des yeux) ainsi que sur leur origine biogéographique, des SNP autosomaux ont été sélectionnés à partir de la littérature (Tableau 2).

	SNP	Chromosome	Gène	Allèle
Hiris Plex	N29insA	16	MC1R	C/insertion A
	rs11547464	16	MC1R	G/A
	rs885479	16	MC1R	C/T
	rs1805008	16	MC1R	C/T
	rs1805005	16	MC1R	G/T
	rs1805006	16	MC1R	C/A
	rs1805007	16	MC1R	C/T
	rs1805009	16	MC1R	G/C
	Y1520CH	16	MC1R	C/A
	rs2228479	16	MC1R	G/A
	rs1110400	16	MC1R	T/C
	rs28777	5	SLC45A2	G/C
	rs16891982*	5	SLC45A2	G/C
	rs12821256	12	KITLG	A/G
	rs4959270	6	EXOC2	C/A
	rs12203592	6	IRF4	C/T
	rs1042602	11	TYR	G/T
	rs1393350	11	TYR	C/T
	rs1800407	15	OCA2	G/A
	rs2402130	14	SLC24A4	A/G
	rs12896399	14	SLC24A4	T/G
	rs12913832	15	HERC2	C/T
	rs2378249	20	ASIP/PIGU	T/C
	rs683	9	TYRP1	T/G
AIM	rs3827760	2	EDAR	C/T
	rs1426654	15	SLC24A5	G/A
	rs2814778	1	DARC	G/A
	rs1545397	15	OCA2	T/A

Tableau 2 : Liste des SNP autosomaux sélectionnés

Les SNP surlignés correspondent aux 6 SNP inclus dans le test Iris Plex (pour la détermination de la couleur des yeux). * ce SNP est également inclus dans le panel AIM.

II.1.2.5.1 Détermination des caractères phénotypiques

Mécanisme de la pigmentation

La couleur de la peau, des yeux ou des cheveux résulte principalement de la quantité, du type ainsi que de la distribution du pigment mélanique. La mélanine est produite par les mélanocytes à l'intérieur d'organites spécifiques: les mélanosomes. Cette synthèse et l'ensemble des processus qui la contrôlent constituent la mélanogenèse. Deux types de mélanines sont produits dans des proportions différentes : l'**eumélanine** (de couleur brune ou noire) et la **phaeomélanine** (de couleur jaune orangée). Les deux types de mélanines proviennent de la transformation enzymatique de la tyrosine en dihydroxyphénylalanine (dopa) puis, sous l'action de la tyrosinase en dopaquinone. Ensuite, les voies de synthèse sont différentes pour les deux types de mélanines. La synthèse d'eumélanine va impliquer les enzymes tyrosinase-related protein (TRP) alors que la synthèse de phaeomélanine va impliquer des dérivés soufrés. D'autres protéines vont également intervenir dans la synthèse de ces pigments. Certaines de ces protéines sont localisées sur les gènes OCA2, SLC24A5 ou SLC45A2. La pigmentation est génétiquement prédéterminée, néanmoins des facteurs environnementaux tels que les UV peuvent venir influencer sur cette régulation.

La détermination des traits phénotypiques tels que la couleur des yeux, des cheveux et de la peau intéresse plusieurs domaines notamment la criminalistique et l'anthropologie. Les nombreuses études de gènes candidats et les études d'association (Eiberg et al., 2008; Han et al., 2008; Kayser et al., 2008; Sulem et al., 2007; Valverde et al., 1995) ont permis depuis plusieurs années d'identifier des variants génétiques impliqués dans les différents phénotypes pigmentaires. Ces SNP ont ensuite été analysés afin d'estimer si leur impact sur le phénotype était assez significatif pour être intégrés dans des futurs tests de prédictions phénotypiques (Branicki et al., 2007; Branicki et al., 2011; Liu et al., 2009).

Certains de ces marqueurs génétiques sont présents sur des gènes impliqués directement dans le processus de pigmentation tandis que d'autres impactent sur ces gènes par des mécanismes de régulation. Parmi ces gènes, on peut noter particulièrement le gène MC1R (MelanoCortin 1 Receptor) qui est fortement impliqué dans la couleur rousse des cheveux et la peau claire (Valverde et al., 1995). Les gènes HERC2 et OCA2, deux gènes voisins, ont quant à eux un impact important sur la couleur des yeux et de la peau. Le gène HERC2, bien que non

impliqué directement dans le processus de régulation, porte un SNP (rs12913832), localisé dans une région intronique, fortement impliqué dans la détermination de la couleur des yeux. Celui-ci va agir comme un activateur de la transcription du gène OCA2 en modulant la conformation de la chromatine et en induisant le recrutement de facteurs de transcription augmentant ainsi la production de mélanine (Visser et al., 2012).

Plusieurs tests ont été proposés afin de déterminer des caractères phénotypiques des individus (Allwood and Harbison, 2013; Ruiz et al., 2013; Spichenok et al., 2011; Walsh et al., 2011) dont l'un développé au laboratoire en 2009 par Caroline Bouakaze (Bouakaze et al., 2009). Des études postérieures à ce travail ont mis à jour des SNP présentant des valeurs prédictives plus importantes. Dans ce contexte, un test appelé HIrisPlex a été mis au point par l'équipe du Professeur M. Kayser aux Pays-Bas, permettant de déterminer la couleur des cheveux et des yeux des individus (Walsh et al., 2013). Ce test, dédié aux investigations médico-légales, se base sur l'analyse de 24 SNP autosomaux impliqués dans les différents phénotypes pigmentaires. Avec ce test, un logiciel de prédiction basé sur une importante base de données a été développé et permet, à partir des génotypes obtenus pour chaque individu, de réaliser aisément une estimation de la couleur des yeux (brun, intermédiaire ou bleu), des cheveux (noir, roux, brun ou blond) mais aussi de la nuance (clair ou foncé). Un panel mondial de populations a également été analysé. Les résultats obtenus ont montré que l'efficacité de ce test était indépendante de l'origine biogéographique des individus testés. Nous avons donc choisi dans cette étude de travailler à partir de cet ensemble de 24 SNP.

II.1.2.5.2 Détermination de l'origine biogéographique

Afin d'estimer l'origine biogéographique d'un individu, il est également possible de s'appuyer sur l'étude de marqueurs autosomaux appelés **AIM** (Ancestry Informative Markers). Ces marqueurs sont particulièrement intéressants car ils présentent des fréquences alléliques très différentes entre certaines populations. Plusieurs panels d'AIM ont été proposés dans la littérature permettant de différencier les populations à l'échelle des continents (Halder et al., 2008; Kosoy et al., 2009; Nassir et al., 2009; Nievergelt et al., 2013; Phillips et al., 2007), ou bien au sein d'une même population (Qin et al., 2014).

Dans cette étude, cinq SNP AIM ont été sélectionnés pour déterminer l'origine biogéographique des individus anciens (certains d'entre eux étant également impliqués dans les processus pigmentaires). En effet, ces SNP présentent des fréquences alléliques clairement

différentes entre les 3 populations majeures : européenne (résidents américains originaires du nord et de l'ouest de l'Europe), asiatique (Chine et Japon) et africaine (Nigeria et Kenya) (Tableau 3). Leur pertinence a été démontrée au sein de panels AIM récemment publiés (Fondevila et al., 2013; Phillips et al., 2007).

Les données génotypiques des 3 populations de référence sont issues du 1000Genomes project_phase1 et obtenues via la base de données du projet ensembl (<http://www.ensembl.org>).

rs3827760	allèle A	allèle G	génotype A/A	génotype A/G	génotype G/G
ASN	0.122	0.878	0.021	0.203	0.776
EUR	0.985	0.015	0.971	0.029	-
AFR	0.994	0.006	0.992	0.004	0.004

rs1426654	allèle A	allèle G	génotype A/A	génotype A/G	génotype G/G
ASN	0.014	0.986	-	0.028	0.972
EUR	0.996	0.004	0.992	0.008	-
AFR	0.077	0.923	0.004	0.146	0.850

rs16891982	allèle C	allèle G	génotype C/C	génotype G/C	génotype C/C
ASN	0.993	0.007	0.986	0.014	-
EUR	0.030	0.970	0.003	0.055	0.942
AFR	0.949	0.051	0.902	0.093	0.004

rs2814778	allèle C	allèle T	génotype C/C	génotype C/T	génotype T/T
ASN	-	1.000	-	-	1.000
EUR	0.003	0.997	-	0.005	0.995
AFR	0.943	0.057	0.898	0.089	0.012

rs1545397	allèle A	allèle T	génotype A/A	génotype A/T	génotype T/T
ASN	0.082	0.918	0.007	0.150	0.843
EUR	0.925	0.075	0.855	0.140	0.005
AFR	0.972	0.028	0.947	0.049	0.004

Tableau 3: Fréquences alléliques et génotypiques des AIM sélectionnés

Les données sont issues du 1000Genomes project_phase1 d'après la base de données du projet Ensembl. ASN : asiatique, EUR : européenne, AFR : africaine.

II.1.2.5.3 Méthode d'analyse des SNP autosomaux

L'analyse des SNP autosomaux impliqués dans la détermination des caractères phénotypiques et de l'origine biogéographique a été réalisée de façon combinée. Six SNP ont été analysés par séquençage direct du fait de leur proximité sur le gène MC1R (rs11547464, rs885479, rs1805008, rs1805007, Y1520CH, rs1110400). Les 22 autres SNP ont été regroupés au sein d'une réaction multiplexe et analysés par spectrométrie de masse MALDI-TOF suivant le

protocole décrit dans la section II.2.2.4. Les amorces utilisées ainsi que la taille des amplicons sont présentés dans l'Annexe 2.

Ainsi, à partir des génotypes, nous avons pu prédire la couleur des yeux et des cheveux à l'aide de l'outil informatique associé au test HIRIS Plex (Walsh et al. 2013). Les phénotypes sont donnés en fonction des valeurs les plus élevées obtenues pour chaque catégorie et d'un guide d'interprétation présenté dans la publication.

L'origine biogéographique des individus a été estimée à partir des génotypes obtenus pour les 5 SNP AIM à l'aide du programme STRUCTURE 2.3.4 (Pritchard et al., 2000). Les génotypes des 556 individus appartenant aux 3 populations de référence (européenne, asiatique et africaine) ont été utilisés et analysés afin de vérifier la cohérence entre les clusters formés et l'origine biogéographique de chaque individu. Ensuite les génotypes des individus typés dans ce travail ont été analysés afin d'estimer la proportion de leur génome dérivée de chaque population de référence. La cohérence entre les génotypes obtenus et l'origine biogéographique des échantillons modernes analysés au laboratoire a également été vérifiée.

II.1.3 Analyses des données génétiques

Les échantillons anciens étudiés ne peuvent pas être définis dans leur ensemble comme une « population ». En effet, les échantillons proviennent de différentes périodes, de différentes régions et de différentes cultures. De la même façon, l'effectif de chaque sous ensemble est bien trop petit (< 15 individus) pour être traité comme une population. Néanmoins, si les calculs des paramètres génétiques n'ont pas été réalisés afin d'éviter des biais trop importants, nous avons souhaité évaluer l'affinité génétique de chaque individu avec les individus anciens déjà analysés ainsi qu'avec les populations eurasiennes actuelles.

II.1.3.1 Recherche d'haplotypes partagés

Afin d'effectuer les recherches d'haplotypes partagés nous avons constitué des bases de données en compilant les données de la littérature. Ainsi, nous avons construit plusieurs bases de données : (i) une base de données regroupant 16081 haplotypes mitochondriaux issus des populations modernes eurasiennes (

Annexe 3), (ii) une base de données regroupant 520 haplotypes mitochondriaux anciens (

Annexe 4), (iii) une base de données regroupant 29662 haplotype Y issus des populations européennes et asiatiques actuelles (Annexe 5). Le nombre de données disponibles concernant des haplotypes Y anciens étant très faible, nous avons compilé 85 haplogroupes Y anciens disponibles dans la littérature (Annexe 4). Les données du chromosome Y ont été compilées dans une base de données réalisées avec le logiciel Access (Microsoft Office) alors que la recherche des haplotypes partagés est réalisée avec l'outil windows Grep 2.3 (<http://www.wingrep.com>) pour les données mitochondriales.

II.1.3.2 Analyses Network

Les réseaux phylogénétiques (ou Network) ont été réalisés à l'aide du logiciel Network 4.6 (Fluxus Engineering) utilisant l'algorithme Median Joining (MJ) (Bandelt et al., 1999). Pour les données STR du chromosome Y, le poids attribué à chaque locus a été pondéré en fonction de la variance allélique observée selon les valeurs données dans la littérature (Qamar et al., 2002). Le détail des options utilisées est présenté dans la légende des figures.

II.1.4 Précautions prises contre les contaminations

Afin d'éviter toute contamination par de l'ADN exogène de nombreuses précautions ont été prises lors du traitement des échantillons d'ADN ancien.

1) Les étapes pré-PCR et post-PCR ont été réalisées dans des laboratoires séparés et isolés. Le cheminement des personnes allant toujours de la partie « pré- » vers la partie « post-PCR » et jamais dans le sens inverse. Ainsi, toutes les étapes en amont de l'amplification par PCR sont réalisées dans un laboratoire dédié qui répond à des critères stricts. Celui-ci est séparé en deux parties distinctes. La première partie est destinée au prétraitement des échantillons (nettoyage et cryobroyage). La seconde partie est dédiée à l'extraction d'ADN et à la préparation des réactions d'amplification. Les étapes d'extraction d'ADN et de préparation des milieux réactionnels sont réalisées sous hotte à flux laminaires dans deux pièces différentes. Ce laboratoire est en surpression et sous irradiation UV. Seules quelques personnes sont habilitées à y rentrer et celles-ci doivent porter un équipement spécifique composé, d'un pyjama, d'une blouse, d'un masque, d'une charlotte, de sur-chaussures ainsi que d'une double paire de gants. Les paillasses et le matériel utilisés sont nettoyés à la javel, à l'eau ultra-pure, ainsi qu'avec une solution de dégradation d'ADN (DNA-ExitusPlus, Macherey Nagel), puis irradiés sous des rayons ultra-violet à chaque utilisation. Les pointes ainsi que les divers plastiques utilisés sont toujours stériles.

- 2) Les profils génétiques des personnes ayant été en contact avec les échantillons (manipulateurs et fouilleurs quand cela est possible) ainsi que celles ayant eu accès au laboratoire d'ADN ancien ont été réalisés. Ils ont ensuite été comparés aux profils génétiques établis à partir des prélèvements anciens pour s'assurer qu'il n'y ait pas eu de contamination des échantillons par ces personnes.
- 3) Des témoins négatifs (sans ADN) ont été réalisés lors de chaque extraction et chaque analyse génétique, afin d'assurer l'absence de contamination lors des manipulations soit par le manipulateur lui-même, soit inter-échantillons ou encore par le matériel utilisé.
- 4) Les analyses réalisées dans ce travail ont été répliquées à partir de plusieurs extractions indépendantes et parfois à partir de plusieurs prélèvements.

III RÉSULTATS

III.1 Résultats des analyses génétiques

III.1.1 Taux de réussite des différentes analyses génétiques menées

Les analyses génétiques ont été menées sur un total de 69 échantillons. Les taux de réussite pour chaque groupe d'échantillons sont présentés dans le tableau suivant.

Echantillons	N	Haplogroupe mitochondrial	Sexe	STR autosomaux*	Haplogroupe Y	
					N hommes	N hg
Kaf	4	4	4	4	3	3
Ok	18	13	14	10	7	6
Bol	1	1	1	0	1	1
Aaf	5	2	2	1	1	0
El	5	4	4	3	3	3
ChemN	3	2	3	3	1	1
Sag	13	11	11	10	8	8
Maf	1	1	1	1	1	1
Chem	5	4	3	3	1	1
MK	3	3	2	2	2	2
ChemS	8	1	1	1	0	0
Bai	3	0	0	0	0	0
Total	69	46 (~67%)	46(~67%)	38(~55%)	28	26

Tableau 4 : Taux de réussite des analyses génétiques réalisées sur les échantillons d'ADN ancien

N : effectif. hg : haplogroupe.

*Seuls les profils génétiques pour lesquels un allèle a été obtenu pour au moins 8 des 15 marqueurs analysés ont été comptabilisés.

Les résultats obtenus montrent que l'ADN est relativement bien conservé au sein des spécimens analysés puisque, près de 67% des échantillons ont permis d'obtenir des données exploitables. Cependant, deux groupes d'échantillons ont livré un ADN trop dégradé pour obtenir des résultats : il s'agit des échantillons ChemS et Bai. Aucun résultat ne sera donc présenté pour ces deux groupes. Il est intéressant de noter que, pour la plupart des individus pour lesquels nous avons obtenu des données mitochondriales, nous avons également pu

obtenir des informations à partir de l'ADN nucléaire, ce qui atteste de la qualité de l'ADN extrait.

III.1.2 Résultats de l'analyse des STR autosomaux et de la détermination du sexe

Un profil génétique au moins partiel a pu être obtenu pour 46 des 69 individus étudiés (Tableau 5). Parmi ces 46 individus, 7 présentent un profil génétique très partiel. En revanche, pour 24 individus, il a été possible d'obtenir au moins un allèle pour chaque locus. L'étude du locus de l'amélogénine inclus dans les kits commerciaux d'identification génétique a permis de déterminer le sexe de 46 individus anciens. Le groupe étudié est ainsi composé de 18 femmes et 28 hommes.

La comparaison des profils génétiques obtenus à partir des échantillons anciens avec ceux des personnes ayant été en contact avec ces échantillons n'a révélé aucune contamination par les manipulateurs. De plus la comparaison deux à deux des profils même partiels ne semble pas indiquer de contamination inter-échantillons.

Echantillons		Amel	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Kaf	Kh19	XX	10/11	29/32.2	(8)/9	(12)	14/17	7/9.3	12/12	9/(10)	(17)/19	13/13	16/16	8/12	12/12	10/12	24/(26)
	Kh20	XY	10/13	29/30	(8)	(10)	15/17	7/9.3	(8/12)	(9/13)	(20)	13/14	15/16	8/12	(14)	11/12	23/26
	Kh21	XY	10/14	30/31	8/11	12/(15)	15/16	9.3	9/10	9/12	19/24	15/15	15/16	11/12	14/20	11/12	20/23
	Kh22	XY	13/13	31.2/32.2	(9)/12	13/13	15/15	7/9.3	8/9	10/13	20/20	14/15.2	18/18	(8)/9	12/16	9/11	(22)/23
Ok	Kh1	XX	14/14	29/(30)	9/(10)	12/12	15/17	7/9.3	(8/9)	11/11	(20)	14.2/(15)	16/(17)	(11)	16/16	12/12	21/21
	Kh2	XX	14/14	30/31.2	8/9	11/13	15/15	9.3/9.3	8/9	10/13	23/23	14.2/14.2	17/19	8/11	12/14	11/12	20/21
	Kh3	XX	(13/14)	30/30	11/11	-	16/16	7/7	(8)	-	-	14.2/14.2	18/(19)	11/11	-	(11)/12	(21/25)
	Kh5	XX	(13)	-	-	-	(15)	-	-	-	-	14.2/(15)	-	-	-	(11)	-
	Kh7	XY	13/13	32.2/32.2	(8/11)	(10)	15/17	(6)/9.3	9/(14)	10/(11)	-	14/15	14/17	12/12	(18/19)	11/12	20/(24)
	Kh8	XY	13/14	(31.2)/32.2	11/11	12/12	16/16	7/9	(11)	(9)/10	-	14/14.2	17/17	8/9	(15)	11/13	23/23
	Kh9	XX	(14)	-	-	-	(15)/18	-	-	-	-	14/14	-	-	-	-	-
	Kh12	XY	14/15	29/33.2	8/8	11/12	15/16	7/9.3	9/10	11/13	21/26	14/14.2	18/18	10/11	14/14	11/13	21/23
	Kh13	XY	10/13	30/31.2	8/8	11/12	16/16	6/8	8/8	10/13	20/22	13.2/14	15/19	9/9	14/14	11/12	21/22
	Kh14	XX	12/13	30.2/32.2	10/12	11/12	15/16	9/9.3	9/11	10/12	23/25	15/15	14/19	12/12	18/18	9/12	22/24
	Kh15	X(Y)	13/15	29/29	(12)	(9/12)	15/15	7/7	11/12	(9)/12	(20)	14/15	17/(19)	9/9	14/14	12/12	21/23
	Kh16	XX	(13)/15	-	-	-	-	(9.3)	(13)	-	-	-	(18)	(11)	-	-	-
	Kh17	X(Y)	13/13	(29)	-	-	16/16	7/(9.3)	-	-	-	(13)	(17)	8/8	-	(9)/13	-
	Kh18	(XY)	(10)	(29)	-	-	(15/17)	(7)	-	(9)	-	(13/14)	-	-	-	-	-
Bol	11	XY	(10/13)	(31.2)	-	-	(16)	(9.3)	-	-	-	(14/15.2)	-	-	-	(11)	(22)
Aaf	2	XX	13/(14)	(27)	(11)	-	16/16	6/(9.3)	(8/12)	(9)	-	15/15.2	15/17	(8)/12	(12/18)	(11)/13	22/22
	3	(XY)	-	-	-	-	(17)	-	-	-	-	(15)	-	-	-	-	-
EI	6	XY	11/16	32.2/32.2	(12)	-	15/17	9.3/9.3	9/9	(9/12)	-	13.2/13.2	16/16	11/(12)	14/14	11/12	22/27
	7	XY	10/15	(29/31.2)	(8/12)	(10)	15/15	6/9.3	(10)	10/(11)	(19)	13/13	18/18	12/12	14/14	10/10	22/(23)
	8	XX	-	-	-	-	-	-	-	-	-	(13/15)	-	(11)	-	(11/12)	(24)
	9	XY	13/13	(31.2)	-	-	16/16	(7)	(10)	(13)	-	(13.2)/14	(16)	(8/11)	-	(13)	(24)

Tableau 5 : profils génétiques consensus obtenus à partir des échantillons d'ADN ancien (1/2)

Les allèles entre parenthèses n'ont été observés qu'une seule fois. Amel : amélogénine.

Les profils consensus ont été obtenus à partir d'au moins 2 amplifications différentes à partir d'extraits d'ADN différents.

Echantillons		Amel	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	THO1	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
ChemN	AT_626	XY	12/15	29/29	-	(10)	15/16	8/9	(9)/11	9/9	-	13.2/15	18/19	(8)/11	19/(21)	12/13	22/(25)
	AT_628	XX	(16)	-	-	-	(15/16)	(9)	(8)	(11)	-	(14/15)	(14/16)	-	(17)	(11/12)	(20/22)
	AT_635	XX	14/(15)	(31/32.2)	(9)/10	-	16/16	(9/9.3)	(8)	(9/11)	-	13/15.2	17/17	-	(13/16)	10/11	21/24
Sag	TA4	XY	12/13	30/33.2	8/11	11/12	15/16	9/9	11/13	10/13	20/25	15/15.2	17/18	8/11	14/14	11/13	19/23
	TA8	XY	12/16	31/31	8/10	11/14	15/16	6/9.3	8/10	10/11	20/23	11/15.2	18/19	8/11	17/20	11/12	22/24
	TA9	XX	12/17	29/31.2	11/11	11/13	15/17	7/9	10/12	12/14	20/23	14/15	16/17	8/9	12/17	11/13	23/24
	TA10	XY	11/12	29/30	8/10	10/11	17/18	8/9	8/11	9/12	20/(22)	12/14	16/18	8/8	19/19	11/13	22/23
	TA11	XY	12/14	30/32.2	8/11	10/11	14/15	7/9.3	10/14	10/12	19/19	13/14.2	14/14	11/12	14/18	11/12	22/26
	TA12	XY	13/13	30/30	9/11	10/11	15/17	9.3/9.3	11/14	11/11	17/19	14/14	17/18	9/12	14/18	12/13	22/24
	TA14	XY	13/14	33.2/33.2	9/10	11/14	15/15	6/9	11/12	12/14	19/25	15/15	18/20	8/8	13/15	9/11	23/24
	TU15	XX	14/16	30/30	11/12	(11)	15/17	6/9.3	10/13	9/11	(19)	13/14	17/17	9/11	17/20	12/13	24/25
	TU17	XY	14/14	32.2/32.2	8/11	10/11	17/18	7/9.3	10/12	10/11	17/18	14.2/15.2	17/18	11/11	16/17	12/13	21/25
	TU18	(XX)	10/10	-	-	-	-	(9.3)	-	-	-	(13)	-	-	-	(12)	-
	TU34	XY	11/13	30.2/32.2	8/11	10/10	16/16	6/7	9/12	11/12	(19/25)	13/13.2	18/20	8/11	13/19	10/11	19/20
Maf	TU31	XY	10/12	29/33.2	12/12	10/11	14/14	8/9.3	11/12	9/12	18/23	14/15.2	14/17	8/8	13/14	11/12	21/25
Chem	13	XY	11/15	30/(31.2)	10/10	(12)	14/15	8/9.3	9/12	(11)/13	(18)/19	13/15.2	17/19	8/11	15/(17)	10/11	23/24
	15	XX	12/13	31.2/33.2	(12)	(12)	15/16	6/9	10/12	(11/13)	(23)	14/14	15/17	8/(11)	(14)	10/12	24/24
	16	XX	12/13	30/31	11/(12)	-	14/16	(7)	(11)/14	10/12	-	14/15	(16)/17	8/8	14/14	11/13	(23)/24
MK	AT_614	XY	12/14	30/33.2	8/10	11/12	15/15	7/7	8/11	9/11	18/20	13/13	17/18	10/11	11/16	11/11	23/23
	AT_861	XY	(13/14)	(30)	-	-	(14/15)	9.3/9.3	(10)	9/9	-	14/14	16/19	8/11	(13/17)	(11)/12	21/21
ChemS	AT 592_3	XX	13/13	30/(31)	-	12/12	16/16	(9)/9.3	11/14	(12)	(23)	14/15	17/17	(8/12)	(13)	12/12	22/25

Tableau 5 : profils génétiques consensus obtenus à partir des échantillons d'ADN ancien (2/2)

Les allèles entre parenthèses n'ont été observés qu'une seule fois. Amel : amélogénine.

Les profils consensus ont été obtenus à partir d'au moins 2 amplifications différentes à partir d'extraits d'ADN différents.

III.1.3 Les données mitochondriales

III.1.3.1 Haplotypes et haplogroupes mitochondriaux

Le séquençage de la région HV1, à partir d'au moins deux extraits d'ADN différents pour chaque échantillon, a permis d'obtenir un haplotype mitochondrial pour 43 individus soit environ 62 % de l'ensemble des échantillons testés (Tableau 7). Trente-deux haplotypes différents ont été observés. Vingt-cinq d'entre eux sont uniques et sept sont partagés par 2 ou 3 individus anciens (Tableau 6). Au vu du nombre important d'haplotypes différents caractérisés, il n'a pas été possible d'établir pour chacun d'eux le sous-haplogroupe exact d'appartenance. En effet, le typage des positions codantes définissant chacun de ces sous-haplogroupes aurait requis un nombre trop important de SNP à typer. C'est pourquoi seule une validation de l'haplogroupe d'appartenance a été effectuée par typage des positions définissant les macro-haplogroupes. L'efficacité de ce typage s'est révélée excellente puisqu'elle est de 100%. Ce typage a permis de confirmer l'appartenance des 32 haplotypes à 7 haplogroupes majoritaires (H, U, J, T, C, D et A). Pour deux haplotypes (16037G 16311C 16319A et 16305G 16311C), le typage des SNP sélectionnés n'a pas permis de les affilier à un haplogroupe.

Echantillons	Haplotype HV1
AT_614 ; Kh13 ; TU34	CRS
Kh3 ; Kh8 ; Kh21	16362C
2 ; 11 ; TU31	16037G 16311C 16319A
Kh19 ; Kh20 ; 7	16223T 16298C 16327T
Kh1 ; Kh2	16093C 16223T 16288C 16298C 16327T
6 ; 9	16223T 16298C 16319A 16325C 16327T 16362C
8 ; TA11	16223T 16362C

Tableau 6 : Haplotypes mitochondriaux partagés au sein de l'ensemble des sujets anciens étudiés

Echantillons		haplotype HVS-1 (16024-16383)	Hg HV1	A- 663	C- 13263	C4a1- 1715	D- 5178	D5a- 4752	D5a2- 1438	HV- 14766	H- 7028	T- 4917	U- 12308	J- 13708	Z- 9090	Hg SNP
Kaf	Kh19	16223T 16298C 16327T	C	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	Kh20	16223T 16298C 16327T	C	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	Kh21	16362C	(H)	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	Kh22	16354T 16356C	(H/U)	A	A	C	C	C	G	T	T	A	<u>G</u>	G	T	U
Ok	Kh1	16093C 16223T 16288C 16298C 16327T	C5	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	Kh2	16093C 16223T 16288C 16298C 16327T	C5	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	Kh3	16362C	(H)	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	Kh7	16069T 16126C	J	A	A	C	C	C	G	T	T	A	A	<u>A</u>	T	J
	Kh8	16362C	(H)	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	Kh9	16223T 16227C 16278T 16290T 16311C 16319A	A10	<u>G</u>	A	C	C	C	G	T	T	A	A	G	T	A
	Kh12	16256T 16270T	U5	A	A	C	C	C	G	T	T	A	<u>G</u>	G	T	U
	Kh13	CRS	H/U	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	Kh14	16126C 16294T 16296T 16324C	T2(a1b)	A	A	C	C	C	G	T	T	<u>G</u>	A	G	T	T
	Kh15	16189C 16223T 16362C	D	A	A	C	<u>A</u>	C	G	T	T	A	A	G	T	D
	Kh16	-	-	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	Kh17	16189C 16223T 16290T 16319A 16362C	A4	<u>G</u>	A	C	C	C	G	T	T	A	A	G	T	A
	Kh18	16223T 16242T 16290T 16319A	A8	<u>G</u>	A	C	C	C	G	T	T	A	A	G	T	A
Bol	11	16037G 16311C 16319A	?	A	A	C	C	C	G	T	T	A	A	G	T	?
Aaf	2	16037G 16311C 16319A	?	A	A	C	C	C	G	T	T	A	A	G	T	?
	3	16192T 16256T 16270T 16304C	U5(a1)	A	A	C	C	C	G	T	T	A	<u>G</u>	G	T	U
El	6	16223T 16298C 16319A 16325C 16327T 16362C	C1	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	7	16223T 16298C 16327T	C	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
	8	16223T 16362C	D	A	A	C	<u>A</u>	C	G	T	T	/	A	G	T	D
	9	16223T 16298C 16319A 16325C 16327T 16362C	C1	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C

Tableau 7 : Haplotypes consensus et haplogroupes mitochondriaux (1/2)

Les positions mutées sont indiquées en fonction de la séquence de référence indiquée CRS. Les allèles informatifs pour la détermination de l'haplogroupe sont indiqués en gras.

Echantillons			haplotype HVS-1 (16024-16383)	Hg HV1	A-	C-	C4a1-	D-	D5a-	D5a2-	HV-	H-	T-	U-	J-	Z-	Hg SNP
					663	13263	1715	5178	4752	1438	14766	7028	4917	12308	13708	9090	
ChemN	AT_626	16223T 16239T 16319A 16362C		D4b1	A	A	C	<u>A</u>	C	G	T	T	A	A	G	T	D
	AT_635	16214T 16223T 16290T 16311G 16319A 16362C		A4	<u>G</u>	A	C	C	G	G	T	T	A	A	G	T	A
Sag	TA4	16093C 16126C 16163G 16186T 16189C 16294T		T1a	A	A	C	C	C	G	T	T	<u>G</u>	A	G	T	T
	TA8	16356C		U4	A	A	C	C	C	G	T	T	A	<u>G</u>	G	T	U
	TA9	16086C 16092C 16164G 16182C 16183C 16185T 16189C 16193del 16223T 16266T 16362C		D5(a2a1)	A	A	C	A	<u>T</u>	<u>A</u>	T	T	A	A	G	T	D5a2
	TA10	16093C 16129A 16223T 16260T 16263C 16298C 16327T		C4a1	A	<u>G</u>	<u>T</u>	C	C	G	T	T	A	A	G	T	C4a1
	TA11	16223T 16362C		D	A	A	C	<u>A</u>	C	G	T	T	/	A	G	T	D
	TA12	16305G 16311C		?	A	A	C	C	C	G	T	T	A	A	G	T	?
	TA14	16092C 16129A 16223T 16362C		D2	A	A	C	<u>A</u>	C	G	T	T	A	A	G	T	D
	TU15	16183C 16189C 16356C		(H/U)	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	TU17	16223T 16290T 16319A 16362C		A4	<u>G</u>	A	C	C	C	G	T	T	A	A	G	T	A
	TU18	16223T 16311C 16362C		D	A	A	C	<u>A</u>	C	G	T	T	A	A	G	T	D
	TU34	CRS		H/U	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
Maf	TU31	16037G 16311C 16319A		?	A	A	C	C	C	G	T	T	A	A	G	T	?
Chem	13	16126C 16294T 16296T		T2	A	A	C	C	C	G	T	T	<u>G</u>	A	G	T	T
	14	-		-	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	15	16051G 16129C 16183C 16189C 16355T 16362C		U2e	A	A	C	C	C	G	T	T	A	<u>G</u>	G	T	U
	16	16093C 16223T 16297C 16298C 16327T		C4	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
MK	AT_614	CRS		H/U	A	A	C	C	C	G	<u>C</u>	<u>C</u>	A	A	G	T	H
	AT_861	16223T 16294T 16362C		D	A	A	C	<u>A</u>	C	G	T	T	A	A	G	T	D
	AT_862	-		-	A	<u>G</u>	C	C	C	G	T	T	A	A	G	T	C
ChemS	AT 592_3	16126C 16292T 16294T		T2c1	A	A	C	C	C	G	T	T	<u>G</u>	A	G	T	T

Tableau 7: Haplotypes consensus et haplogroupes mitochondriaux (2/2)

Les positions mutées sont indiquées en fonction de la séquence de référence indiquée CRS. Les allèles informatifs pour la détermination de l'haplogroupe sont indiqués en gras.

III.1.3.2 Analyse des données mitochondriales

III.1.3.2.1 Distribution des haplogroupes mitochondriaux

Les haplogroupes mitochondriaux obtenus à partir des échantillons anciens étudiés peuvent être classés en 2 catégories d'après leur répartition actuelle (i) les haplogroupes ouest-eurasiens (H, U, J et T), (ii) les haplogroupes est-eurasiens (A, B, C et D). Le Tableau 8 montre la distribution des haplogroupes obtenus au sein de chaque groupe.

Echantillons	Hg ouest eurasiens					Hg est eurasiens				Autre
	H	U	J	T	Total	A	C	D	Total	?
Kaf	1	1	0	0	2	0	2	0	2	0
Ok	3	1	1	1	6	3	3	1	7	0
Bol	0	0	0	0	0	0	0	0	0	1
Aaf	0	1	0	0	1	0	0	0	0	1
El	0	0	0	0	0	0	3	1	4	0
ChemN	0	0	0	0	0	1	0	1	2	0
Sag	2	1	0	1	4	1	1	4	6	1
Maf	0	0	0	0	0	0	0	0	0	1
Chem	1	1	0	1	3	0	1	0	1	0
MK	1	0	0	0	1	0	1	1	2	0
ChemS	0	0	0	1	1	0	0	0	0	0

Tableau 8 : Distribution des haplogroupes obtenus au sein de chaque groupe

III.1.3.2.2 Haplotypes mitochondriaux partagés avec les autres populations anciennes eurasiennes

Les haplotypes caractérisés ont été comparés aux données mitochondriales d'échantillons anciens publiées et incluses dans notre base de données présentée dans la section II.2.3.1. Les résultats sont présentés dans le Tableau 9 page suivante.

Tableau 9 : Haplotypes mitochondriaux partagés entre les échantillons analysés et les données d'ADN ancien publiées dans la littérature

Seuls les individus pour lesquels des séquences partagées ont été retrouvées sont présentés dans le tableau.
BP : Before Present (avant aujourd'hui (1950)) ; AD : Anno Domini (après Jésus-Christ) ; BC : Before Christ (avant Jésus-Christ).

Echantillons	Hg	Haplotypes	Individus partageant le même haplotype		
			N	Datation/culture	Localisation
AT_614 Kh13 TU34	H	CRS	4	Linear Pottery culture (5000 BC)	Allemagne
			6	Néolithique (3000 BC)	France
			3	Néolithique (3500 - 3000 BC)	Espagne
			1	Funnel Beaker culture (3000 BC)	Suède
			1	Chasseurs-cueilleurs	Europe centrale
			1	milieu 2 ^{ème} millénaire BC	Altai, Russie
			1	13 ^{ème} - 11 ^{ème} siècle BC	Kazakhstan
			2	culture de Tagar, 800-100 BC	Sud de la Sibérie
			1	2135 BP	Désert du Taklamakan
			1	2800-2100 BP	Xinjiang, Chine
			2	culture de Pazyryk (4 ^{ème} - 3 ^{ème} siècle BC)	Altai, Russie
			4	8 ^{ème} - 5 ^{ème} siècle BC	Kazakhstan
Kh19 Kh20 7	C	16223T 16298C 16327T	1	3 ^{ème} siècle BC	Kazakhstan
			1	Néolithique (6125-4885 BC)	Lac Baïkal
			2	Néolithique (3710-2020 BC)	Lac Baïkal
			4	culture d'Ust-Tartas (env 4 ^{ème} millénaire BC)	Sibérie de l'Ouest
			3	culture de Krotovo (2 ^{ème} millénaire BC)	Sibérie de l'Ouest
			4	culture d'Andronovo (2 ^{ème} millénaire BC)	Sibérie de l'Ouest
			1	Néolithique (3600BP)	Yakoutie
			6	3500 BP	Nord-est de l'Europe
			1	2800-2100 BP	Xinjiang, Chine
			1	Fin du 2 ^{ème} millénaire BC	Sibérie de l'Ouest
Kh1 Kh2	C	16093C 16223T 16288C 16298C 16327T	1	culture de Sargat 8 ^{ème} - 12 ^{ème} siècle AD	Sibérie de l'Ouest
			1	Age du Fer	Altai, Mongolie
8 TA 11	D	16223T 16362C	1	Xiongnu (2000 BP)	Mongolie
			1	culture d'Ust-Tartas (env 4 ^{ème} millénaire BC)	Sibérie de l'Ouest
			6	culture d'Oudinovo (début du 3 ^{ème} millénaire BC)	Sibérie de l'Ouest
			3	3500 BP	Nord-est de l'Europe
			2	3600 BP	Nord-est de la Chine
			1	2800-2100 BP	Xinjiang, Chine
			5	Xiongnu	Mongolie
13	T	16126C 16294T 16296T	1	770-256 BC	Mongolie Intérieure, Chine
			1	Linear Pottery culture (5000 BC)	Allemagne
TA4	T	16093C 16126C 16163G 16186T 16189C 16294T	1	Age du Fer	Kazakhstan
			1	14 - 11 siècle BC	Kazakhstan
Kh12	U	16256T 16270T	2	Néolithique (6125 - 4885 BC)	Lac Baïkal
			1	Néolithique (3710 - 2020 BC)	Lac Baïkal
			3	Pitted ware culture (3000 BC)	Suède
			2	culture de Krotovo (2 ^{ème} millénaire BC)	Sibérie de l'Ouest
			1	Fin du deuxième millénaire BC	Sibérie de l'Ouest
3	U	16192T 16256T 16270T 16304C	1	culture Late Krotovo (2 ^{ème} millénaire BC)	Sibérie de l'Ouest
			2	culture d'Andronovo (2 ^{ème} millénaire BC)	Sibérie de l'Ouest
TA8	U	16356C	1	7000 BP	Nord-ouest de la Russie
			2	culture d'Andronovo (2 ^{ème} millénaire BC)	Sud de la Sibérie
Kh7	J	16069T 16126C	1	Funnel Beaker culture (3000 BC)	Europe Centrale
			4	Linear Pottery culture (5000 BC)	Allemagne
			6	Néolithique (3000 BC)	France
			2	Néolithique 3500 - 3000 BC	Espagne
			1	Fin de l'âge du Bronze / début de l'âge du Fer	Sibérie de l'Ouest
			1	Age du Fer	Altai, Mongolie
TU17	A	16223T 16290T 16319A 16362C	1	7 ^{ème} - 3 ^{ème} siècle BC	Kazakhstan
			3	Xiongnu	Mongolie
Kh18	A	16223T 16242T 16290T 16319A	1	culture de Pazyryk (4 ^{ème} - 3 ^{ème} cent BC)	Altai, Russie
Kh9	A	16223T 16227C 16278T 16290T 16311C 16319A	1	Fin de l'âge du Bronze / début âge du Fer	Sibérie de l'Ouest
AT_626	D	16223T 16239T 16319A 16362C	1	4 ^{ème} - 3 ^{ème} siècle BC	Altai, Russie
AT_861	D	16223T 16294T 16362C	1	Fin de l'âge du Bronze / début de l'âge du Fer	Sibérie de l'Ouest
Kh15	D	16189C 16223T 16362C	1	3600 BP	Nord-est de la Chine
			2	2800 - 2100 BP	Xinjiang, Chine
TU18	D	16223T 16311C 16362C	2	Xiongnu	Mongolie
			1	770 - 256 BC	Mongolie Intérieure, Chine
TA12	?	16305G 16311C	1	2800 - 2100 BP	Xinjiang, Chine

Parmi les 32 haplotypes recherchés, 18 (soit 56%) sont retrouvés au sein de la base de données anciennes. De manière attendue, parmi les séquences retrouvées, celles qui le sont le plus fréquemment sont les séquences de base pour les haplogroupes C (16223T, 16298C, 16327T), D (16223T, 16362C), U5a (16256T, 16270T) et J (16069T, 16126C), ainsi que la séquence identique à la séquence de référence (CRS). Les résultats obtenus montrent également que les haplogroupes est-asiatiques C et D étaient présents en Sibérie de l'Ouest aux environs du 4^{ème} millénaire avant J.C. chez des sujets de la culture Ust-Tartas. L'haplogroupe U5 présente une répartition déjà très vaste puisqu'il est également présent chez des individus néolithiques mis au jour à la fois en Suède et dans la région du lac Baïkal. Parmi les haplogroupes ouest-asiatiques, les haplotypes affiliés aux haplogroupes J et T sont retrouvés majoritairement dans les populations de l'ouest de l'Eurasie à l'exception d'un individu porteur de l'haplogroupe J retrouvé dans l'Altaï mongol à l'âge du Fer. L'haplogroupe U4 est également retrouvé auparavant au nord-est de l'Europe puis dans le sud-sibérien. Enfin, les haplotypes est-asiatiques sont majoritairement partagés avec des populations asiatiques (Mongolie, Chine) mais aussi avec des populations sud-sibériennes et ouest-sibériennes à la fin de l'âge du Bronze et au début de l'âge du Fer.

III.1.3.2.3 Haplotypes mitochondriaux partagés avec les populations modernes

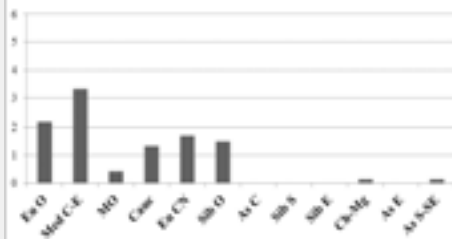
Parmi les 32 haplotypes recherchés dans la base de données contemporaines, 22 haplotypes (soit ~ 69%) sont retrouvés. Huit haplotypes ne sont pas présents au sein de cette base de données. Plusieurs explications sont possibles. Il se peut que ces haplotypes ne soient pas référencés dans notre base de données du fait de sa non-exhaustivité. Néanmoins, la recherche de ces haplotypes sur la base de données en ligne EMPOP, dédiée aux séquences mitochondriales, s'est également révélée infructueuse. Il est donc possible que l'absence de ces haplotypes dans les bases de données actuelles soit due à une disparition de ces lignées au cours du temps ou tout du moins à une sous-représentation. La répartition des haplotypes au sein des populations eurasiennes actuelles est présentée dans la Figure 15. De façon intéressante, l'haplotype 037G 311C et 319A, partagé par deux individus de la culture Afanasievo et 1 individu de la culture Bol'shemyskaya (2, 11, TU31), qui n'était pas retrouvé chez les populations anciennes et que nous n'avons pas pu rattacher à un haplogroupe particulier, est retrouvé chez un Finlandais.

Figure 15 : Fréquence des haplotypes mitochondriaux au sein des populations eurasiennes actuelles

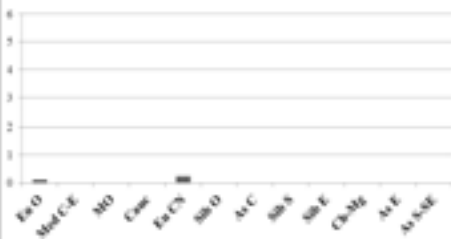
D'après les haplotypes disponibles dans notre base de données. Eu O : Europe de l'Ouest ; Med C-E : Méditerranée du Centre et de l'Est ; MO : Moyen-Orient ; Cauc : Caucase ; EU CN : Europe Centrale et du Nord ; Sib O : Sibérie Occidentale ; As C : Asie Centrale ; Sib S : Sibérie du Sud ; Sib E : Sibérie de l'Est ; Ch-Mg : Chine et Mongolie ; As E : Asie de l'Est ; As S-SE : Asie du Sud et du Sud-Est.

H

362C

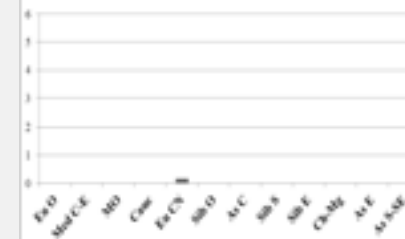


183C 189C 356C



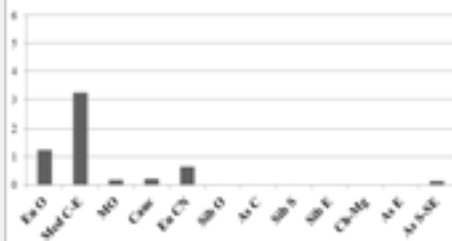
?

037G 311C 319A

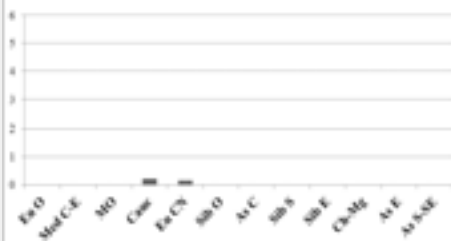


U

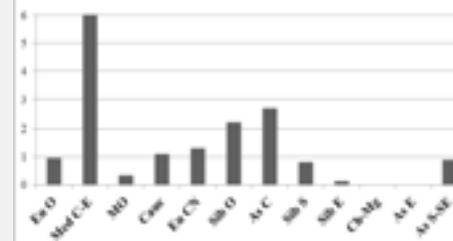
256T 270T



192T 256T 270T 304C

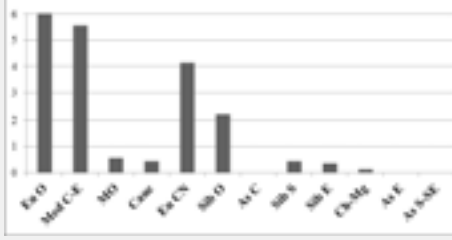


356C



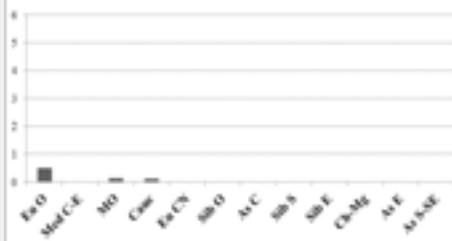
J

069T 126C

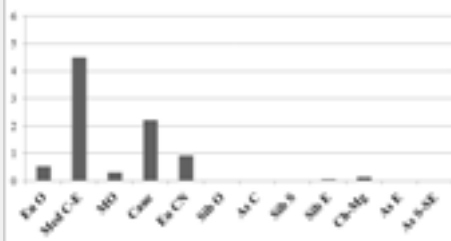


T

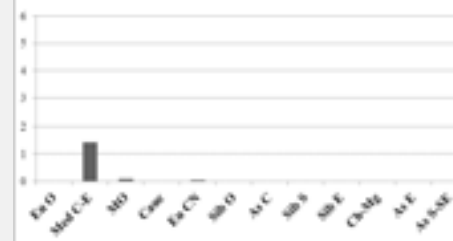
126C 292T 294T



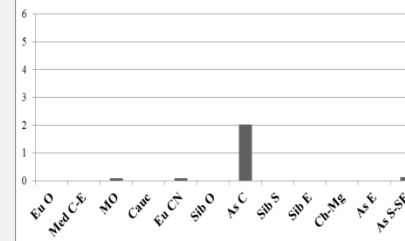
126C 294T 296T



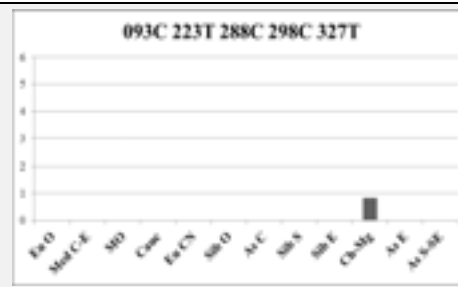
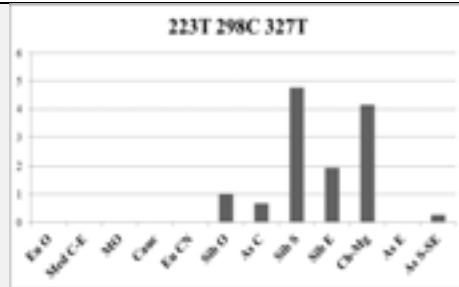
126C 294T 296T 324C



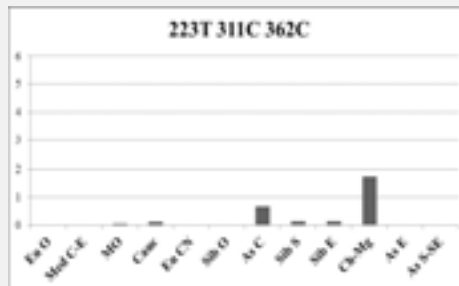
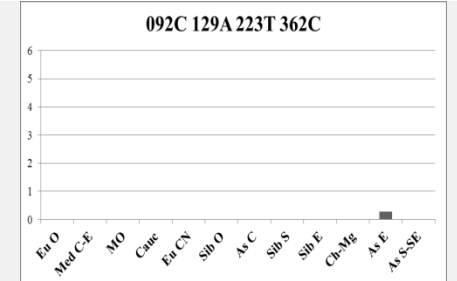
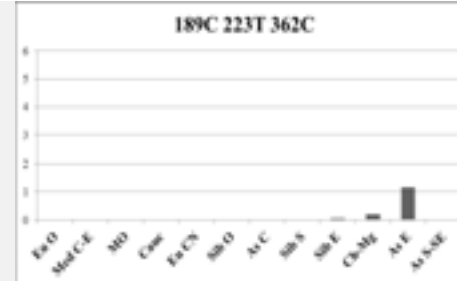
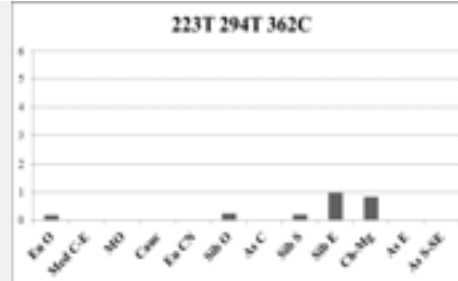
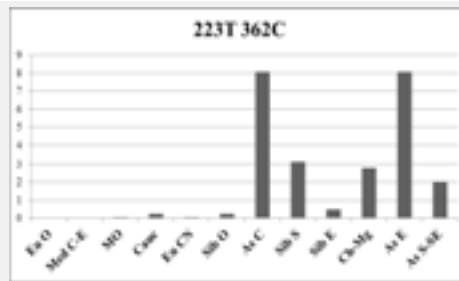
093C 126C 163G 186T 189C 294T



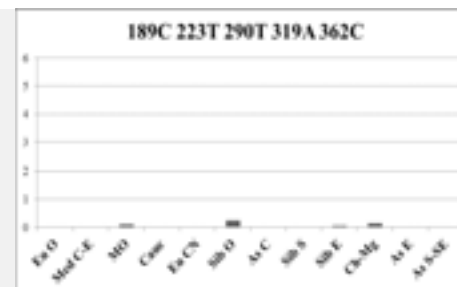
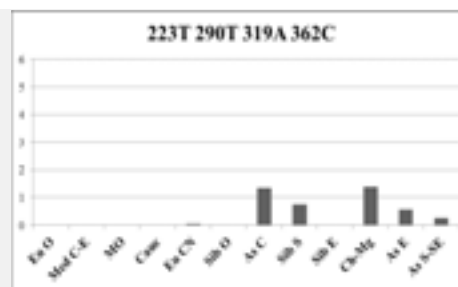
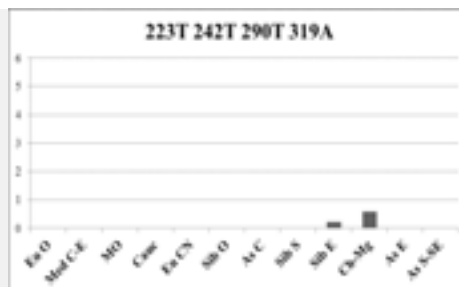
C



D



A



III.1.4 Les données du chromosome Y

III.1.4.1 Haplotypes et haplogroupes Y

L'analyse des STR localisés sur la région non recombinante du chromosome Y a permis de déterminer les haplotypes Y pour 26 des 28 individus masculins étudiés. Parmi ces haplotypes, 18 sont complets (Tableau 10). Quatre d'entre eux sont partagés par 2 individus : TA11 et TA12 ; Kh21 et Kh22 ; TA4 et TA10 ; TA14 et TU31. A partir de ces haplotypes, une prédiction de l'haplogroupe a été réalisée à l'aide de l'outil *haplogroup predictor* disponible en ligne (<http://www.hprg.com/hapest5/>). Les résultats ainsi que les probabilités d'appartenance à l'haplogroupe sont indiqués dans la dernière colonne du tableau.

Cette première analyse nous a ensuite permis de typer les échantillons avec l'une ou l'autre des réactions multiplexes développées. Les résultats du typage des SNP sont présentés dans le Tableau 11.

Echantillons	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATAH4	DYS437	DYS438	DYS448	haplogroup predictor	
Kaf	Kh20	16	13	(24)	29	16	(14)	11/14	12	10	(13)	(23)	(13)	12	15	12	19	R1b 100%
	Kh21	15	13	24	29	16	14	11/14	12	11	13	23	13	13	15	12	19	R1b 100%
	Kh22	15	13	24	29	16	14	11/14	12	11	13	23	13	13	15	12	19	R1b 100%
Ok	Kh7	16	13	(24)	(29)	17	(14)	11/14	12	10	(12)	(23)	(13)	13	15	12	19	R1b 100%
	Kh8	17	13	23	-	15	(13)	13/(18)	13	(11)	12	-	-	11	15	12	16	Q 98,1%
	Kh12	14	14	24	31	17	16	11/13	14	10	11	22	14	12	15	9.2	21	N 95,2%
	Kh13	14	13	24	31	18	15	11/13	14	11	12	21	14	12	14	10	21	N 99,6%
	Kh15	14	13	(24)	31	18	15	11/13	14	11	12	(21)	-	12	(14)	10	21	N 88,1%
	Kh17	15	(13)	-	-	-	(13)	-	(13)	(10)	(13)	-	-	(11)	-	-	-	Q 60%
Bol	11	(16)	(13)	-	(29)	(16)	-	(11/14)	(12)	-	(12)	-	-	(13)	-	-	-	R1b 95,7%
El	6	17	14	(24)	-	15	(13)	14/17	13	(10)	12	22	-	12	15	-	(19)	Q 99,4%
	7	16	(13)	-	-	(19)	-	(13/17)	13	10	(12)	(22)	-	(11)	-	-	-	Q 87,3%
	9	(15)	13	-	-	(16)	-	(12)	13	(9)	-	-	-	(9)	(14)	-	(20)	C3 81,1%
ChemN	AT_626	15	13	23	(30)	17	16	13/13	14	10	12	(21)	11	11	14	(10)	20	C3 99,6%
Sag	TA4	16	14	24	32	16	16	11/14	13	11	10	23	11	12	14	11	20	R1a 100%
	TA8	15	13	24	29	16	13	15/18	13	10	14	22	14	11	13	11	19	Q 100%
	TA10	16	14	24	32	16	16	11/14	13	11	10	23	11	12	14	11	20	R1a 100%
	TA11	15	13	24	29	16	13	15/(17)	13	10	13	22	14	11	13	11	(19)	Q 100%
	TA12	15	13	24	29	16	13	15/17	13	10	13	22	14	11	13	11	(19)	Q 100%
	TA14	16	13	25	31	15	16	11/14	13	11	10	23	11	13	14	11	20	R1a 100%
	TU17	15	12	23	27	17	16	11/16	14	10	11	21	11	11	14	10	21	C3 97,4%
	TU34	16	13	25	31	15	16	11/14	13	11	10	23	11	13	14	11	20	R1a 100%
Maf	TU31	17	12	22	28	17	13	14/17	14	11	13	23	16	10	14	11	19	Q 100%
Chem	13	15	13	22	(29)	16	13	11/17	13	10	13	22	15	10	14	11	19	Q 100%
MK	AT_614	15	14	24	31	18	16	11/13	14	10	11	22	14	12	15	9.2	21	C3 55,4%
	AT_861	15	14	23	(30)	18	-	11/12	13	10	-	(21)	-	11	14	(10)	20	C3 77,1%

Tableau 10 : Haplotypes Y consensus

Les haplotypes ont été déterminés pour l'ensemble des échantillons définis comme masculins après analyse du locus de l'amélogénine.
Les allèles entre parenthèses n'ont été observés qu'une seule fois.

Echantillons		Multiplexe 1										Multiplexe 2												Haplogroupes	
		NO-	O-	N-	Q-	Q1a3-	Q1a3a-	Q1a3a1-	Q1a1-	C-	D-	R-	R1-	R1a-	R1a1-	R1a1a-	M458	M434	Z280	Z93	R1b-	R1b1a-	R1b1a1-		R1b1a2-
		M214	M175	M231	M242	M346	L54	M3	M120	M130	M174	M207	M173	M420	M516	M17					M343	P297	M73		M269
Kaf	Kh20	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	-	T	A	G	A	G	C	G	A	C	GT	C	R1b1a2-M269
	Kh21	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	-	T	A	G	A	G	C	G	A	C	GT	C	R1b1a2-M269
	Kh22	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	-	T	A	G	A	G	C	G	A	C	(GT)	-	R1b1a-P297
Ok	Kh7	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	-	T	A	G	A	G	C	-	A	C	GT	C	R1b1a2-M269
	Kh8	T	TTCTC	-	T	G	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3-M346
	Kh12	C	TTCTC	-	C	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	NO (non O)
	Kh13	C	TTCTC	-	C	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	NO (non O)
	Kh15	C	TTCTC	-	-	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	NO (non O)
	Kh17	T	-	-	-	G	A	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3a-L54
Bol	11	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	(G)	-	T	(A)	-	A	G	-	G	A	C	(GT)	-	R1b1a-P297
El	6	T	TTCTC	-	(T)	G	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3-M346
	7	T	TTCTC	-	(T)	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	(Q_M242)
	9	T	TTCTC	-	-	C	G	G	-	-	T	nd	nd	nd	nd	nd	A	G	nd	nd	nd	nd	nd	nd	?
ChemN	AT_626	T	TTCTC	-	C	C	G	G	-	T	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	C-M130
Sag	TA4	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	C	A	G	DEL	A	G	C	A	C	G	GT	T	R1a1a1b2-Z93
	TA8	T	TTCTC	-	T	G	A	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3a-L54
	TA10	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	-	A	G	DEL	A	G	C	A	C	G	GT	T	R1a1a1b2-Z93
	TA11	T	TTCTC	-	T	G	A	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3a-L54
	TA12	T	TTCTC	-	T	G	A	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q1a3a-L54
	TA14	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	C	A	G	DEL	A	G	C	A	C	G	GT	T	R1a1a1b2-Z93
	TU17	T	TTCTC	-	C	C	G	G	T	T	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	C-M130
	TU34	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	G	C	A	G	DEL	A	G	C	A	C	G	GT	T	R1a1a1b2-Z93
Maf	TU31	T	TTCTC	-	T	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Q-M242
Chem	13	T	TTCTC	-	-	C	G	-	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	?
MK	AT_614	C	TTCTC	-	C	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	NO (non O)
	AT_861	C	TTCTC	-	C	C	G	G	T	C	T	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	NO (non O)

Tableau 11 : Résultats du typage des SNP de la région non recombinante du chromosome Y
Les allèles ayant permis de déterminer l'haplogroupe sont indiqués en gras.

S'il a été possible de typer tous les SNP présents dans la réaction multiplexe 2, cela n'a pas été le cas avec la réaction multiplexe 1. En effet, pour les marqueurs M217 (C3), M231 (N), Tat (N1c1) et M128 (N1c2a), des résultats ont été obtenus à partir d'ADN moderne mais, pas à partir d'ADN ancien. Les allèles étaient soit totalement absents, soit d'intensité trop faible pour être différenciés du bruit de fond.

En ce qui concerne le marqueur P43 (N1c2b), nous avons décidé de ne pas prendre en compte les résultats du typage. En effet, ce marqueur (sélectionné au regard de la littérature) est localisé sur le gène SMCY du chromosome Y qui a un gène homologue SMCX sur le chromosome X. Les amorces PCR et SBE peuvent donc tout à fait s'hybrider aux séquences du gène SMCX et SMCY durant la réaction d'amplification. Lorsque l'ADN étudié n'est pas dégradé (ADN contemporain), la séquence du chromosome Y est toujours amplifiée, en revanche, lorsqu'il s'agit d'ADN ancien, généralement dégradé, il est possible que l'amplification de la séquence du chromosome Y échoue. On ne sait donc pas si le résultat obtenu est bien l'allèle du marqueur P43 ou celle portée par le chromosome X.

Ainsi, au vu de l'efficacité de la réaction multiplexe 1, il est impossible d'affirmer l'appartenance d'un échantillon à l'haplogroupe N ou à l'un de ses sous-haplogroupes.

L'application de ces réactions multiplexes sur les individus masculins a néanmoins permis la détermination de l'haplogroupe Y pour 24 d'entre eux. Bien que pour certains échantillons il n'ait pas été possible d'amplifier tous les marqueurs, nous pouvons observer la distribution suivante :

- 4 individus sont affiliés à l'haplogroupe R1a1a1b2-Z93.
- 5 individus sont affiliés à l'haplogroupe R1b1a-P297 et 3 d'entre eux ont pu être rattachés au sous haplogroupe R1b1a2-M269.
- 2 individus sont affiliés à l'haplogroupe C-M130.
- 8 individus sont affiliés à l'haplogroupe Q-M242 ou à l'un de ses sous-haplogroupes (Q1a3-M346 et Q1a3a-L54).
- 5 individus appartiennent à l'ensemble NO(non O), souvent noté NO(xO). Effectivement, ils portent la mutation pour le marqueur NO-M214 et la présence de l'allèle ancestral pour le marqueur M175 exclut leur appartenance à l'haplogroupe O. Cependant, l'absence d'amplification pour le marqueur M231 ne permet pas de les affilier à l'haplogroupe N bien que les résultats obtenus à partir de l'analyse des STR du chromosome Y soient en faveur d'une prédiction pour l'haplogroupe N pour les 3

qui avaient un score de prédiction élevé (>88%). Pour deux autres individus (AT_614 et AT_861), à partir des données STR, c'est l'haplogroupe C3 qui avait été prédit mais ce résultat est nuancé par la faible valeur de prédiction obtenue.

Pour les individus 13 et 9, le typage des SNP s'est révélé non informatif, car seuls des allèles ancestraux ont été observés sur les marqueurs pour lesquels l'amplification a été possible. Néanmoins les résultats partiels obtenus ne vont pas à l'encontre des prédictions réalisées à partir des STR, à savoir l'appartenance à l'haplogroupe Q pour l'individu 13 et à l'haplogroupe C pour l'individu 9.

III.1.4.2 Analyse des données du chromosome Y

III.1.4.2.1 Haplotypes partagés avec les populations anciennes

Par comparaison avec les données mitochondriales, le nombre de données Y publiées dans la littérature est beaucoup plus faible. Les recherches ont été effectuées sur la base des haplogroupes et non des haplotypes car, dans la plupart des études, seuls les SNP définissant les haplogroupes ont été analysés. De plus, les analyses ne sont pas toutes réalisées avec le même niveau de profondeur dans la phylogénie.

Tous les haplogroupes présents chez les spécimens étudiés ont déjà été retrouvés chez des individus anciens. L'haplogroupe R1b a été identifié chez deux individus néolithiques (culture campaniforme) en Allemagne. L'haplogroupe R1a montre une distribution plus large, il est présent à la fois chez un individu néolithique de la culture cordée en Allemagne, chez plusieurs Scytho-Sibériens de l'âge du Bronze et de l'âge du Fer, dans le bassin du Tarim à l'âge du Bronze ainsi que chez un individu Xiongnu à l'est de la Mongolie. L'haplogroupe Q a été identifié chez des individus du nord de la Chine il y a plus de 2000 ans. L'haplogroupe N était présent au nord-est de la Chine à l'âge du Bronze. Enfin, l'haplogroupe C a été identifié chez un individu de la culture d'Andronovo dans le sud de la Sibérie.

III.1.4.2.2 Haplotypes partagés avec les populations modernes

Seuls les échantillons ayant un haplotype minimal complet (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393) ont été comparés à la base de données contemporaines (Tableau 12). La recherche a été effectuée sur la base de ces 7 loci, car, pour certaines populations de comparaison, seules ces données sont disponibles. Parmi les 12

haplotypes recherchés, des correspondances chez les populations actuelles ont été trouvées pour 7 haplotypes, notamment pour les haplotypes rattachés aux haplogroupes R1a et R1b. Les haplotypes R1b sont principalement retrouvés dans des populations ouest-eurasiennes. Les haplotypes R1a sont eux retrouvés à la fois dans des populations ouest-eurasiennes, chez quelques individus sibériens ou encore dans le sud de l'Asie. L'haplotype Q des individus TA11 et TA12 est retrouvé uniquement chez deux individus asiatiques. L'haplotype C (individu TU17) est présent uniquement chez un individu est-asiatique et enfin l'haplotype porté par les individus Kh12 et AT_614 est retrouvé uniquement chez un Finlandais. Les haplotypes R1a1a1b2–Z93 ont également été recherchés dans les données publiées par Underhill et al. (2010), et Pamjav et al. (2012), articles dans lesquels les sous-haplogroupes de R1a1a ont été décrits, afin d'estimer l'affinité de ces haplotypes avec les haplotypes affiliés aux sous-haplogroupes R1a1a-M458, Z280, et Z93. Nos haplotypes sont partagés avec des individus R1a1a(xM458) d'après Underhill et al. (2010) et de façon quasi-exclusive avec des individus R1a1a1b2-Z93 d'après Pamjav et al. (2012).

Echantillons	haplotype minimal (19, 389I, 389II, 390, 391, 392, 393, 385a, 385b)	Individus modernes
TA4 / TA10	16 14 32 24 11 11 13 11 14	5 ind : 2 Polonais, 1 Allemand, 1 Slovaque, 1 Mon-Khmer
TA11 / TA12	13 13 29 24 10 14 13 15 17	2 ind : 1 Tibétain, 1 Even
TA8	13 13 29 24 10 14 13 15 18	No match
TA14 / TU34	16 13 31 25 11 11 13 11 14	33 ind : 5 Polonais, 4 Russes (ouest), 3 Allemands, 2 Slovaques, 2 Lettons, 2 Afghans, 2 Tuviniens, 2 Indiens, 1 Altaïen-Kizhi, 1 Croate, 1 Tchèque, 1 Russe (Est), 1 Evenk, 1 Iranien, 1 Malaisien, 1 Suédois, 1 Turque, 1 Ouïghour, 1 Yakoute.
TU17	16 12 27 23 10 11 14 11 16	1 ind : 1 Japonnais
TU31	13 12 28 22 11 16 14 14 17	non retrouvé
Kh20 / Kh7	14 13 29 24 10 13 12 11 14	17 ind : 2 espagnols, 2 Portugais, 2 Polonais, 2 Allemands, 1 Irlandais, 1 Belge, 1 Sicilien, 1 Grecque, 1 Slovaque, 1 Iranien, 1 Tchèque, 1 Coréen, 1 Afghane
Kh21 / Kh22	14 13 29 24 11 13 12 11 14	28 ind : 8 Espagnols, 1 Autrichien, 1 Belge, 1 Croate, 2 Allemands, 2 Grecques, 2 Bosniaques, 2 Polonais, 1 Macedonien, 1 Italien, 1 Iranien, 5 Tchèques, 1 Ouïghour
Kh12 / AT_614	16 14 31 24 10 14 14 11 13	1 ind : 1 Finlandais
Kh13 / Kh15	15 13 31 24 11 14 14 11 13	non retrouvé
13	13 13 29 22 10 15 13 11 17	non retrouvé
AT_626	16 13 30 23 10 11 14 13 13	non retrouvé

Tableau 12 : Haplotypes Y partagés entre les individus anciens étudiés et les populations modernes

Le nom des STR composant l'haplotype minimal sont indiqués entre parenthèses avec la mention DYS exclue.

III.1.4.3 Analyses phylogénétiques

A partir des données STR du chromosome Y, une analyse phylogénétique à l'aide du logiciel Network a été réalisée, afin de déterminer l'affinité des haplotypes anciens avec les haplotypes présents dans les populations actuelles.

Afin de voir où se placent les haplotypes anciens R1b parmi les haplotypes R1b actuels, nous avons réalisé un *Median Joining Network* (Figure 16). Pour cela les haplotypes minimaux des 4 individus anciens R1b1a2-M269 ainsi que ceux publiés par Myres et al., ont été utilisés. Le réseau obtenu montre que les individus anciens (en noir) se placent majoritairement avec des individus modernes affiliés à l'haplogroupe R1b1a2a-L23 (en orange). Le marqueur L23 est présent majoritairement au sud-est de l'Europe, dans le Caucase, en Turquie et dans la région de l'Oural avec une fréquence souvent supérieure à 10% (Myres et al., 2011).

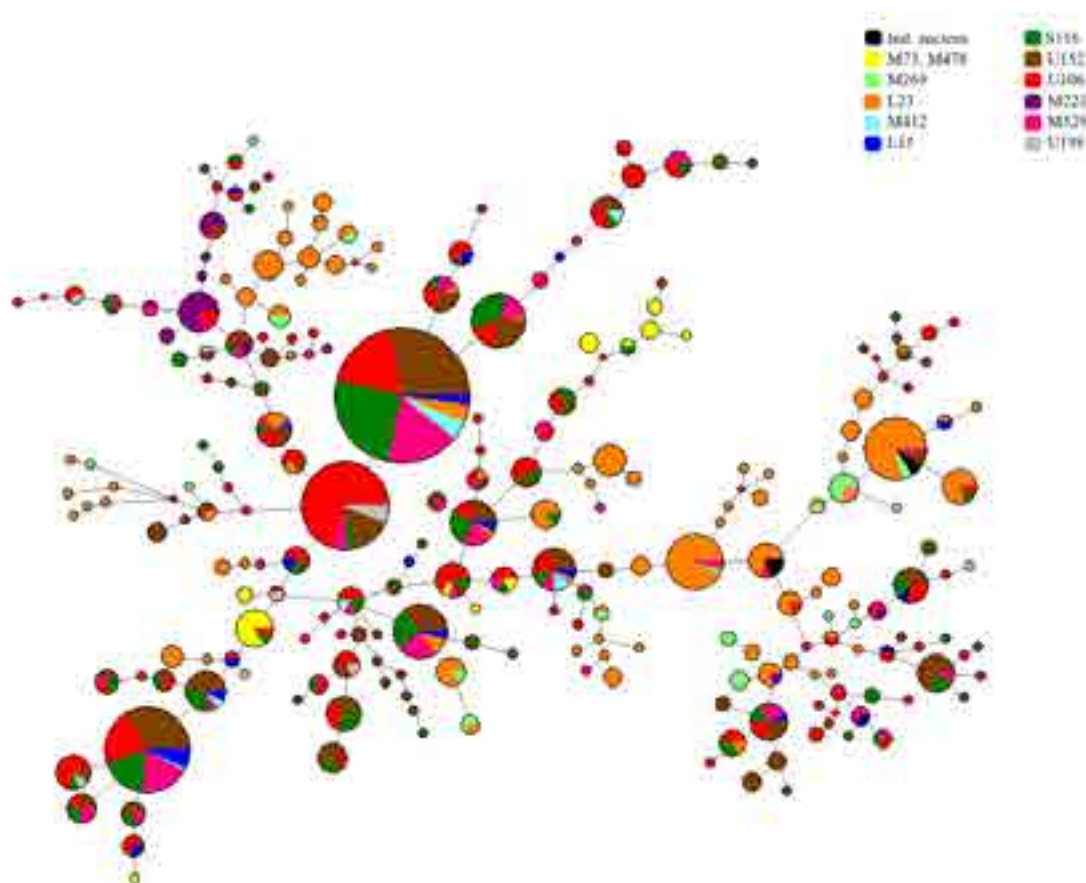


Figure 16 : *Median Joining Network* réalisé à partir des haplotypes R1b

Réseau phylogénétique réalisé à partir des haplotypes minimaux R1b observés chez nos individus anciens et ceux publiés par Myres et al.. Afin d'éviter un nombre trop important de réticulation l'option « star contraction » a été utilisée.

III.1.5 Résultats du typage des SNP autosomaux

III.1.5.1 Résultats obtenus à partir des échantillons modernes

Détermination des caractères phénotypiques

Afin de déterminer la couleur des cheveux et des yeux des individus anciens étudiés, ainsi que leur origine biogéographique, nous avons adapté le test HIrisPlex (Walsh et al., 2013) pour une application sur spectromètre de masse et y avons ajouté des marqueurs de l'origine biogéographique (SNP AIM). Dans un premier temps, il était impératif d'évaluer ce test sur des échantillons d'ADN provenant d'individus modernes pour lesquels les informations relatives au phénotype et à l'origine biogéographique étaient connues.

Dix-sept échantillons modernes ont été analysés avec le consentement des donneurs. Vingt-deux SNP ont été analysés par PCR multiplexe suivie d'une détection par spectrométrie de masse MALDI-TOF, et six SNP, localisés sur le gène MC1R, ont été analysés par séquençage direct du fait de leur proximité sur le gène. Il n'a pas été possible de typer par spectrométrie de masse MALDI-TOF le SNP rs1805009. Ce dernier, localisé sur le gène MC1R, a pourtant un impact particulièrement élevé sur le phénotype roux (Walsh et al., 2013). C'est pourquoi il a été analysé par séquençage direct sur les personnes présentant un phénotype décrit comme roux. Pour les personnes avec un autre phénotype, nous avons réalisé les prédictions des caractères pigmentaires en indiquant dans l'outil prédictif le génotype homozygote G/G (allèle ancestral) pour le SNP rs1805009.

A partir des génotypes obtenus pour chaque individu, nous avons réalisé une prédiction de la couleur des cheveux et des yeux. Pour cela, nous avons utilisé l'outil informatique associé au test HIrisPlex disponible en matériel supplémentaire de la publication. Les génotypes obtenus, ainsi que les informations relatives aux phénotypes (observés et prédits), sont présentés en Annexe 6. Les résultats des prédictions sont cohérents pour la couleur des yeux pour tous les individus sauf un qui est décrit comme bleu/gris mais qui est prédit comme brun. Pour cet individu, aucune valeur élevée de prédiction n'est observée pour une catégorie en particulier, mais toutes les catégories (bleu, brun, intermédiaire) ont des valeurs supérieures ou égales à 0.2. De plus, en observant bien la couleur de l'iris de ce sujet, des pigments jaunes sont distingués, ce qui pourrait expliquer la valeur élevée obtenue pour la catégorie «brun». Il est

intéressant de noter que les 3 individus d'origine non-européenne (2 asiatiques et 1 africain) ont une valeur de prédiction de 0 pour la couleur «bleu».

Concernant la couleur des cheveux, les prédictions sont correctes pour tous les individus aux cheveux blonds, bruns ou noirs. Pour les individus aux cheveux roux, les résultats obtenus sont plus discutables, puisque 1 individu a été prédit comme roux mais 2 individus ont été prédits blonds. L'individu prédit comme roux est hétérozygote pour deux loci fortement associés à la couleur rousse (rs1805007 et rs1805008), contrairement aux deux autres individus qui ne sont hétérozygotes que pour un locus (rs1805008). L'état hétérozygote pour ce seul marqueur MC1R n'a pas été déterminé comme suffisamment puissant pour prédire le phénotype roux à lui tout seul, puisque ce phénotype est décrit comme associé soit à un état homozygote récessif soit à une hétérozygotie combinée (Branicki et al., 2007; Walsh et al., 2013). En effet, le marqueur rs1805008 à l'état hétérozygote seul est également retrouvé chez un individu aux cheveux bruns dans notre échantillon. Ce génotype n'est donc pas considéré comme suffisamment discriminant par le logiciel de prédiction pour établir à lui seul le phénotype roux.

III.1.5.1.1 Détermination de l'origine biogéographique

Afin de déterminer l'origine biogéographique des individus, 5 SNP AIM (rs1545397 ; rs2814778 ; rs1426654 ; rs16891982 ; rs3827760) ont été sélectionnés au vu de leurs différences de fréquences alléliques au sein de populations africaines, asiatiques ou européennes. Afin de s'assurer de l'efficacité des SNP sélectionnés pour estimer l'origine biogéographique, les génotypes de 3 populations de référence (européenne, africaine et asiatique) ont été recueillis. Les génotypes ont été analysés sans aucune indication de la population d'origine avec le logiciel STRUCTURE. Tous les individus ont bien été regroupés au sein de 3 clusters correspondant aux 3 origines biogéographiques différentes. Nous avons ensuite testé les génotypes des individus modernes typés au laboratoire. L'origine biogéographique estimée par le logiciel STRUCTURE pour chacun d'entre eux correspondait bien à leur origine biogéographique (Annexe 7).

Dans l'objectif d'appliquer ces tests sur des échantillons anciens, donc généralement à des échantillons contenant une faible quantité d'ADN, la réaction multiplexe a été appliquée sur des échantillons modernes contenant des quantités décroissantes d'ADN. Une gamme de dilution a été réalisée. Les résultats obtenus montrent que jusqu'à 100 pg d'ADN, il y a peu de

perte d'information, seule la perte d'un état hétérozygote ou la perte d'un génotype est observée pour 1 à 4 marqueurs sur les 22 typés (Annexe 8). Ces résultats sont cohérents avec les résultats obtenus par Mendisco et al. (2011) qui montraient que pour l'ADN nucléaire, à partir de 100 pg les haplotypes obtenus n'étaient pas tous complets.

III.1.5.2 Résultats obtenus sur les échantillons d'ADN ancien

III.1.5.2.1 Détermination des caractères phénotypiques

Les 22 SNP contenus dans la réaction multiplexe ont été analysés à partir des échantillons anciens pour lesquels des données génétiques avaient déjà été obtenues. Le séquençage de la région du gène MC1R contenant les 6 SNP a été réalisé uniquement sur les échantillons Sag car ils présentaient une bonne préservation de l'ADN lui-même disponible en quantité relativement importante. Les génotypes et les prédictions obtenues pour tous les individus anciens sont présentés dans le tableau en Annexe 9. Les résultats montrent qu'il n'a pas été possible d'obtenir des génotypes complets pour tous les individus. En effet, à l'issue du typage par spectrométrie de masse MALDI-TOF 23 individus présentaient un génotype complet (en omettant le rs1805009). Pour ces individus, la prédiction de la couleur des yeux et des cheveux a pu être réalisée à partir de l'outil informatique utilisé pour les échantillons modernes. Neuf individus anciens présentaient un génotype partiel, il n'a donc pas été possible de déterminer la couleur des cheveux mais nous avons pu prédire la couleur des yeux basée sur l'analyse de seulement 6 SNP à l'aide de l'outil de prédiction informatique associé à la publication relative au test Iris Plex (Walsh et al., 2011). Enfin, pour 5 individus, les génotypes très partiels obtenus ne permettaient pas d'utiliser l'un ou l'autre des outils de prédiction. Néanmoins nous avons pu estimer la couleur des yeux à partir uniquement du SNP rs12913832 localisé sur le gène HERC2 qui a un fort pouvoir de prédiction sur la couleur des yeux (Eiberg et al., 2008; Liu et al., 2009; Sturm et al., 2008; Walsh et al., 2011). En effet, le génotype G/G serait associé au phénotype « bleu » et les génotypes G/A et A/A au phénotype « brun ».

Echantillons	Cheveux		Yeux	
	N	phénotype	N	phénotype
Kaf	2	brun foncé - noir	2	brun
	2	brun clair - brun foncé	1	intermédiaire
			1	bleu
Ok	4	brun foncé - noir	9	brun
	1	brun clair - brun foncé	1	bleu
Bol	1	-	1	brun
Aaf	0	-	1	brun
El	0	-	3	brun
ChemN	0	-	2	brun
Sag	4	brun foncé - noir	8	brun
	5	brun clair - brun foncé	1	intermédiaire
			1	bleu
Maf	1	brun foncé - noir	1	brun
Chem	3	brun foncé - noir	3	brun
MK	1	brun foncé - noir	1	brun

Tableau 13 : Bilan par groupe de la prédiction des caractères pigmentaires pour les individus anciens

III.1.5.2.2 Détermination de l'origine biogéographique

L'estimation de l'origine biogéographique a également été réalisée pour les individus anciens à partir des 5 SNP AIM. Les génotypes obtenus ont été analysés avec le logiciel STRUCTURE (Annexe 7). Les résultats obtenus sont présentés sur les graphiques en triangle de la Figure 17.

Au vu des données obtenues, il apparaît que 17 échantillons ont une composante européenne majoritaire alors que 4 individus semblent plus proches des populations asiatiques actuelles. Six individus semblent d'origine mixte. Si on réalise une analyse en fonction des groupes culturels on observe des différences. Les individus appartenant à la culture Afanasievo (Kaf, Aaf et Maf) semblent tous proches des populations européennes actuelles. Pour le groupe Sagsai, certains individus semblent avoir une origine mixte alors que d'autres sont plus proches des populations européennes actuelles. Enfin, les résultats obtenus pour les individus de la culture Okunevo suggèrent des origines biogéographiques plus variées.

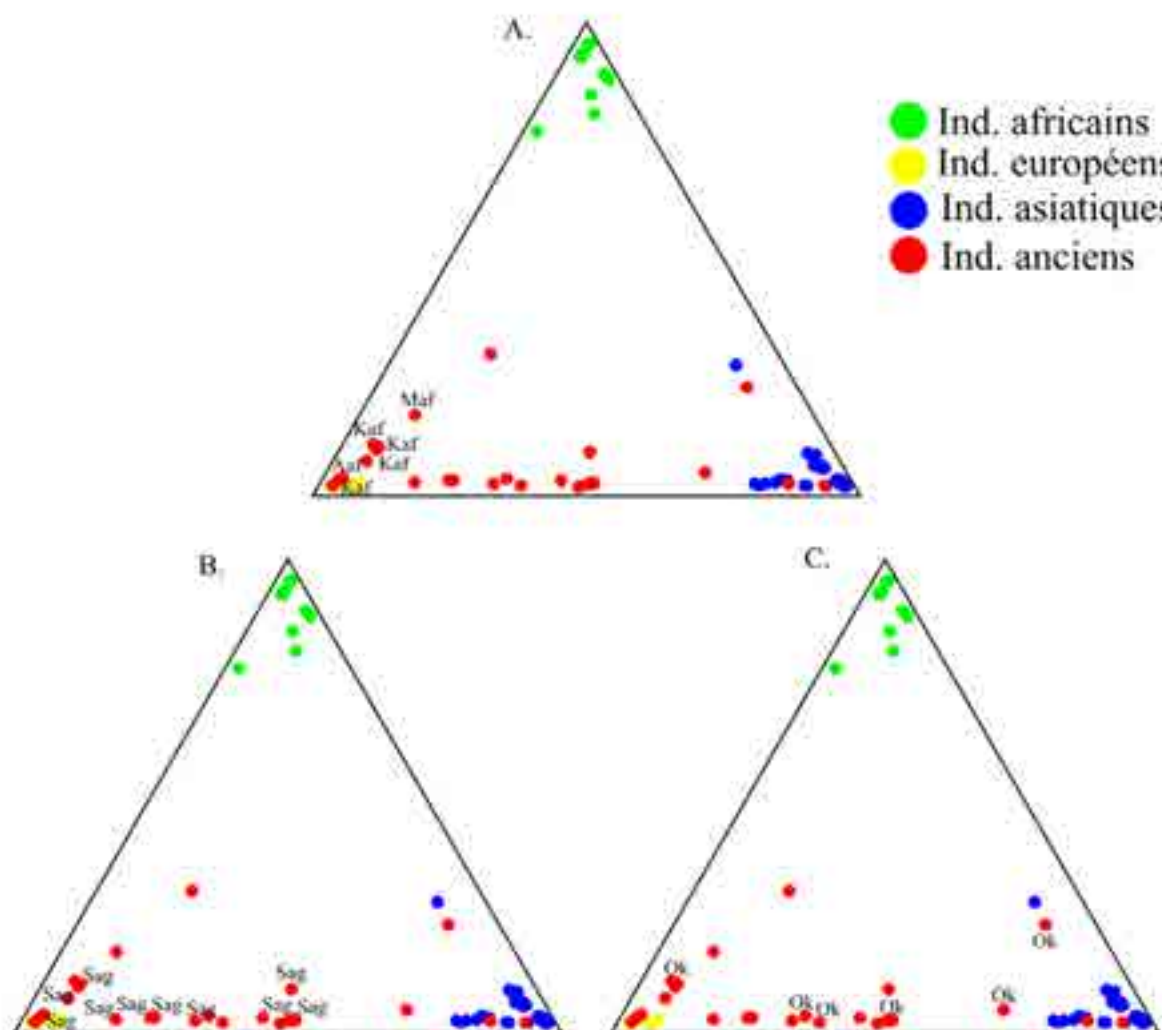


Figure 17 : Graphiques triangulaires générés à partir du logiciel STRUCTURE représentant l’affinité génétique des individus anciens avec les 3 populations de référence

A : Individus de la culture Afanasievo ; B : Individu de la culture Sagsai ; C : Individus de la culture Okunevo.

III.1.6 Synthèse des résultats obtenus pour chaque échantillon

Au final, lorsque la qualité de l’ADN le permettait, nous avons pu déterminer pour chaque échantillon le profil génétique, le sexe, les haplogroupes maternels et paternels, les caractères pigmentaires (couleur des cheveux et couleur des yeux) ainsi que la probable origine biogéographique. Un résumé des résultats obtenus pour chaque échantillon est présenté dans le Tableau 14.

Echantillons		Sexe moléculaire	Hg mt	Hg Y	Cheveux	Yeux	AIM		
							Europ	Afr	As
Kaf	Kh19	XX	C	R1b1a2-M269 R1b1a2-M269 R1b1a-P297	brun foncé - noir	brun	0.961	0.024	0.015
	Kh20	XY	C		brun clair - brun foncé	Int.	0.991	0.005	0.004
	Kh21	XY	H		brun clair - brun foncé	bleu	0.968	0.020	0.011
	Kh22	XY	U		brun foncé - noir	brun	0.857	0.091	0.051
Ok	Kh1	XX	C	R1b1a2-M269 Q1a3-M346	-	brun	-	-	-
	Kh2	XX	C		brun foncé - noir	brun	0.260	0.036	0.705
	Kh3	XX	H		-	(brun)	-	-	-
	Kh7	XY	J		-	bleu	0.626	0.008	0.366
	Kh8	XY	H	Q1a3-M346	brun foncé - noir	brun	0.649	0.021	0.329
	Kh9	XX	A	-	-	-	-	-	
	Kh12	XY	U	NO (non O)	brun foncé - noir	brun	0.085	0.225	0.690
	Kh13	XY	H	NO (non O)	brun clair - brun foncé	brun	0.486	0.012	0.503
	Kh14	XX	T	-	brun foncé - noir	brun	0.967	0.020	0.013
	Kh15	X(Y)	D	NO (non O)	-	brun	-	-	-
	Kh16	XX	C	-	-	-	-	-	
	Kh17	X(Y)	A	Q1a3a-L54	-	(brun)	-	-	-
	Kh18	(XY)	A	-	-	-	-	-	-
Bol	11	XY	?	R1b1a-P297	-	(brun)	-	-	-
Aaf	2	XX	?		-	(brun)	0.990	0.005	0.005
	3	(XY)	U		-	-	-	-	
El	6	XY	C	Q1a3-M346	-	brun	-	-	-
	7	XY	C	(Q_M242)	-	(brun)	-	-	-
	8	XX	D	-	-	-	-	-	
	9	XY	C	?	-	brun	0.496	0.011	0.493
ChemN	AT_626	XY	D	C-M130	-	brun	-	-	-
	AT_635	XX	A	-	-	(brun)	0.045	0.007	0.948
Sag	TA4	XY	T	R1a1a1b2-Z93	brun foncé - noir	brun	0.519	0.004	0.477
	TA8	XY	U	Q1a3a-L54	-	int.	0.899	0.060	0.041
	TA9	XX	D5a2	R1a1a1b2-Z93 Q1a3a-L54 Q1a3a-L54 R1a1a1b2-Z93	brun foncé - noir	brun	0.486	0.014	0.499
	TA10	XY	C4a1		brun clair - brun foncé	brun	0.751	0.018	0.231
	TA11	XY	D		brun foncé - noir	brun	0.458	0.083	0.459
	TA12	XY	?		brun foncé - noir	brun	0.867	0.086	0.047
	TA14	XY	D	R1a1a1b2-Z93	brun clair - brun foncé	bleu	0.761	0.016	0.222
	TU15	XX	H	C-M130	brun clair - brun foncé	brun	0.679	0.011	0.310
	TU17	XY	A		brun clair - brun foncé	brun	0.829	0.013	0.158
	TU18	(XX)	D		-	-	-	-	
TU34	XY	H	R1a1a1b2-Z93		brun clair - brun foncé	brun	0.991	0.005	0.004
Maf	TU31	XY	?	Q-M242	brun foncé - noir	brun	0.752	0.164	0.084
Chem	13	XY	T	?	brun foncé - noir	brun	0.540	0.298	0.162
	14	-	H		-	-	-	-	
	15	XX	U		brun foncé - noir	brun	0.865	0.097	0.037
	16	XX	C		brun foncé - noir	brun	0.546	0.019	0.435
MK	AT_614	XY	H	NO (non O)	brun foncé - noir	brun	0.111	0.011	0.878
	AT_861	XY	D	NO (non O)	-	-	-	-	-
	AT_862	-	C	-	-	-	-	-	
ChemS	AT 592_3	XX	T		-	-	-	-	-

Tableau 14 : Tableau récapitulatif des données obtenues pour chaque échantillon ancien

Hg : haplogroupe ; mt : mitochondrial ; Europ : Européenne ; Afr : Africaine ; As : asiatique.

III.1.7 Etude des liens de parentés

L'étude des STR autosomaux n'a révélé aucun lien de parenté de type parent/enfant dans les différents groupes étudiés puisqu'aucune paire d'individus ne partage un allèle en commun pour chaque locus. Cependant, les profils n'étant pas toujours complets, il est possible que des liens de proche parenté n'aient pas été détectés.

Afin de rechercher des liens de parentés plus éloignés, nous avons regroupé les individus présentant les mêmes haplotypes mitochondriaux ou Y au sein des différents groupes, puis analysé les résultats plus en détail. Ces derniers sont présentés dans le tableau 15.

Echantillons			Sexe	Haplotypes mt	Haplotypes Y	STRa partagés
Kaf	Kh19/Kh20	Itkol'II, kourgane 24, tombe 1 / Itkol'II, kourgane 23, tombe 2	F / H	Identiques	-	12/15
	Kh21/Kh22	Itkol'II, kourgane 27, tombe 1 / Itkol'II, kourgane 27, tombe 1	H / H	Différents	Identiques (17/17 STR)	5/15
Ok	Kh1/Kh2	Krasny Kamen, kourgane 1, tombe 1 / Krasny Kamen, kourgane 1, tombe 3	F / F	Identiques	-	11/15
	Kh13/Kh15	Itkol'II, kourgane 14, tombe 2 / Itkol'II, kourgane 14, tombe 5	H / H	Différents	Identiques (15/15 STR)	9/15
El	6/9	Teleytskiy Vzvoz-I, tombe 7 / Teleytskiy Vzvoz-I, tombe 41	H / H	Identiques	Différents	3/11
Sag	TA4/TA10	Tsagaan Asga, tombe 4 / Tsagaan Asga, tombe 10	H / H	Différents	Identiques (17/17 STR)	11/15
	TA11/TA12	Tsagaan Asga, tombe 11 / Tsagaan Asga, tombe 12	H / H	Différents	Identiques (17/17 STR)	11/15
	TA14/TU34	Tsagaan Asga, tombe 14 / Takhilgat Uzuur-5, tombe 34	H / H	Différents	Identiques (17/17 STR)	9/15

Tableau 15 : Echantillons présentant les mêmes lignées maternelles ou paternelles

STRa : STR autosomaux, mt : mitochondrial.

Les résultats indiquent que les paires d'individus sont, soit reliées par la même lignée paternelle, soit par la même lignée maternelle. Ensuite on peut distinguer plusieurs cas. Les couples (Kh21/Kh22 et 6/9) ne partageant que très peu d'allèles en commun (3/11 ou 5/15), sont probablement liés par une parenté éloignée et impossible à déterminer sans populations de référence. Concernant les paires d'hommes (Kh21/Kh22; Kh13/Kh15; TA4/TA10;

TA11/TA12; TA14/TU34), aucune ne partagent à la fois le même haplotype Y et le même haplotype mitochondrial, ils ne sont donc pas de la même fratrie. Enfin, concernant la paire Kh19/Kh20, les résultats indiquent que l'individu Kh19 est une femme qui, d'après les données anthropologiques aurait entre 40 et 50 ans alors que l'individu Kh20 serait un adolescent. Les données STR semblent exclure un lien de type parent-enfant. Néanmoins, aucune exclusion n'est basée sur deux génotypes hétérozygotes différents mais uniquement sur des homozygoties. Il est donc possible que des génotypes hétérozygotes soient apparus homozygotes du fait de la dégradation de l'ADN et aient faussé les analyses. D'autres liens de type tante/neveu ou grand-mère maternelle/petit-fils sont également envisageables. Enfin, pour les individus féminins Kh1 et Kh2, inhumés dans le même kourgane, et qui partagent le même haplotype mitochondrial, les données archéologiques indiquent qu'elles seraient âgées respectivement de 20-25 et 35-40 ans. De la même façon que pour les individus Kh19/Kh20, les STR semblent exclure un lien de type mère/fille mais l'état de dégradation des échantillons, notamment Kh1, pourrait avoir entraîné des pertes d'allèles. Au vu de leurs âges respectifs, il est également possible d'envisager un lien de parenté de type sœurs. Bien que la stratégie d'analyse employée soit complète, c'est-à-dire à la fois l'établissement des profils génétiques mais aussi l'analyse des lignées paternelles et maternelles des individus, la déduction de liens de parenté autres que parent-enfant semble difficile à partir d'une simple observation des résultats. Il serait donc nécessaire de tester ces hypothèses par des calculs statistiques afin de déterminer une probabilité de parenté. Cela peut se faire avec des logiciels tels DNA-VIEWTM (Brenner, 1997), développé pour la criminalistique, qui est capable de calculer des liens de parenté jusqu'à oncle/neveu. Cependant ce type de logiciel utilise des algorithmes ou calculs basés sur les fréquences alléliques dans les populations actuelles or, il est possible que les fréquences étaient différentes dans les populations anciennes étudiées, ce qui pourrait biaiser les analyses. En outre, il est difficile d'établir quelle(s) population(s) contemporaine(s) serai(en)t la(les) plus proche(s) de la population ancienne du point de vue génétique et pourrait servir de population(s) de référence. Pour cela, il faudrait avoir un effectif bien plus important, afin de voir s'il existe des différences entre les populations du passé et actuelles. D'autres logiciels comme ML-RELATE (Kalinowski et al., 2006) pourraient, si nous disposions d'un groupe plus important, déterminer les différents liens de parenté probables entre les individus, au sein du groupe. Néanmoins, ce qu'il ressort de ces analyses est que peu de liens de très proche parenté semblent exister au sein des kourganes étudiés.

III.2 Authenticité des résultats

Afin de prouver l'authenticité des résultats, de nombreuses précautions ont été prises lors des analyses et une stratégie réfléchie et cohérente a été mise en place. Bien que tous les critères d'authenticité décrits dans la littérature (cf. I.2.3) n'aient pu être respectés, les points suivants semblent attester de la validité des résultats obtenus :

Des témoins négatifs ont été utilisés pour chaque extraction et chaque amplification, et aucune amplification n'a été observée dans ces témoins sans ADN.

Tous les profils génétiques obtenus à partir des échantillons anciens sont uniques et sont différents de ceux des personnes ayant été en contact avec les prélèvements.

Les analyses ont été répétées sur différents extraits d'ADN et parfois différents prélèvements à plusieurs mois d'intervalle et les résultats obtenus étaient reproductibles.

Lors de l'analyse des STR autosomaux ou du chromosome Y, une diminution de l'intensité du signal (voire une perte) a été observée avec l'augmentation de la taille des fragments amplifiés (figure 17) pour l'ensemble des échantillons étudiés. Ce comportement est caractéristique d'un ADN dégradé.

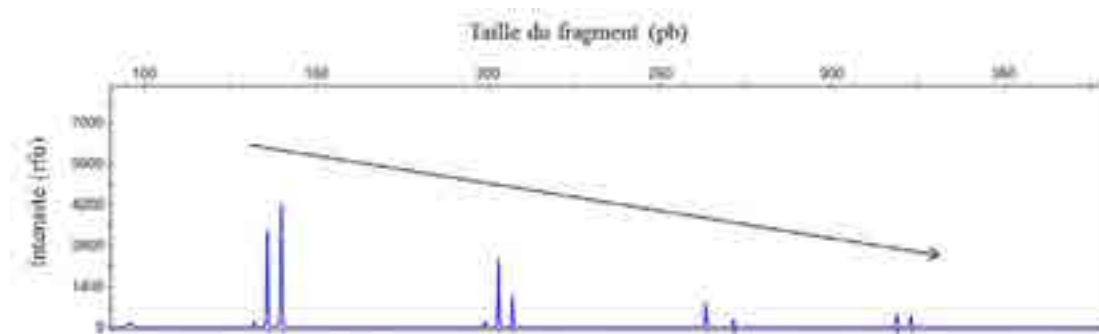


Figure 18 : Corrélation inverse entre la taille des fragments amplifiés et l'intensité du signal

Résultats issus de l'analyse des STR autosomaux de l'individu TA10.

En plus de l'analyse de la région HV1 de l'ADN mitochondrial et des STR du chromosome Y, des positions diagnostiques de la région codante de l'ADN mitochondrial et du chromosome Y ont été typées afin de déterminer les haplogroupes. Les résultats obtenus entre les deux approches étaient cohérents, permettant ainsi de valider l'appartenance aux haplogroupes.

Enfin, lorsque nous disposons d'informations sur les sujets étudiés, comme c'est le cas pour le site de Takhilgat Uzuur-5, nous avons pu observer une cohérence entre le taux de réussite des analyses et l'état de préservation des échantillons décrit dans le rapport de fouille, mais aussi en fonction l'âge du sujet. En effet, il est souvent plus difficile d'obtenir des résultats, particulièrement à partir du génome nucléaire, pour les individus immatures.

IV DISCUSSION

Les steppes eurasiennes connectent l'Europe, l'Asie Centrale, la Chine, l'Asie du Sud et le Moyen-Orient. Au cours du temps, de nombreuses populations, issues d'Europe ou d'Asie, s'y sont succédées. Ces multiples mouvements migratoires expliquent certainement le pool génique présent dans la population actuelle, constitué à la fois de lignées ouest- et est-eurasiennes. Pour comprendre et retracer ces mouvements migratoires, l'accès direct au génome des populations anciennes semble être le moyen le plus fiable pour étudier la diversité génétique présente à un moment donné en s'affranchissant des événements postérieurs qui ont pu considérablement la modifier.

Ce travail s'est plus particulièrement intéressé à la dynamique de peuplement du sud de la Sibérie et de l'Altaï à l'âge du Bronze, période pendant laquelle les steppes eurasiennes auraient connu d'après les données archéologiques de nombreux mouvements migratoires d'Ouest en Est.

Ainsi, l'objectif général de ce travail a été, à partir d'une ensemble composé de 69 individus, de caractériser génétiquement ces populations de l'âge du Bronze, d'étudier les relations génétiques qui existent entre elles, mais aussi, dans un contexte plus large, d'essayer de déterminer leur origine biogéographique afin d'apporter de nouveaux éléments pour la compréhension des événements migratoires complexes survenus dans cette région.

Afin de répondre à cette problématique, un de nos objectifs a été de mettre en place une approche originale, permettant d'étudier au mieux l'origine biogéographique des sujets anciens. Pour cela, en plus de l'analyse des lignées maternelles couramment analysées dans les études de paléogénétiques, nous avons étudié les lignées paternelles des individus masculins, ainsi que des marqueurs de caractères phénotypiques (couleur des yeux et des cheveux) puisque l'apparence physique peut également être informative de l'origine biogéographique, notamment dans cette région tampon à la rencontre des mondes asiatiques et européens.

Ainsi, dans cette quatrième partie, nous discuterons tout d'abord de l'évolution du pool génique au sein de la région pendant l'âge du Bronze. Puis, la confrontation de ces données aux données archéologiques, anthropologiques et génétiques nous permettra de discuter des hypothèses de peuplement. Enfin, nous discuterons de la stratégie expérimentale employée et

des perspectives méthodologiques aujourd'hui envisageables pour de telles études, au vu du développement technologique actuel et de l'évolution de la discipline.

IV.1 Apport de ce travail à la compréhension de la dynamique de peuplement du sud de la Sibérie et de l'Altaï à l'âge du Bronze

Afin d'étudier la dynamique de peuplement du sud de la Sibérie et de l'Altaï, nous disposons d'un groupe de 69 échantillons. Ces échantillons proviennent de sites et de cultures différentes allant du milieu 4^{ème} millénaire au début du premier millénaire avant J.-C. Cet étalement temporel nous permet d'avoir accès à des échantillons couvrant toute la période de l'âge du Bronze. Cependant, cette diversité fait également la faiblesse de cet échantillonnage, puisque le nombre de sujets étudiés par site et par culture est parfois très faible. En effet, pour certaines cultures (Elunino, Munkh-Khairkan, Baitag ou encore Chemurchek nord) nous avons eu accès à seulement 3 ou 4 individus, ne permettant pas, par la suite, de réaliser des analyses statistiques, car le biais imputé à ce faible échantillonnage est extrêmement fort.

IV.1.1 Evolution du pool génique

Afin d'étudier l'évolution du pool génique des populations dans la région du sud de la Sibérie et de l'Altaï, les prélèvements seront répartis au sein de trois groupes pour adopter une organisation temporelle des échantillons. Le premier groupe, daté du début de l'âge du Bronze, est composé des individus porteurs des cultures de Bol'shemyskaya (IV^{ème} millénaire av. J.-C) et d'Afanasievo (env. 3000 av. J.-C). Le second groupe, daté du milieu de l'âge du Bronze, est composé des individus appartenant aux cultures d'Okunevo (2300-1800 av. J.-C), d'Elunino (2300-1700 av. J.-C) et de Chemurchek (2500-1800 av. J.-C). Enfin, le troisième groupe, daté de la fin de l'âge du Bronze, est composé des individus appartenant aux cultures de Munkh Khairkan (1700-1400 av. J.-C) et de Sagsai (1400-900 av. J.-C). Pour chaque groupe, les résultats sont schématisés sur la Figure 19 page 116.

IV.1.1.1 Début de l'âge du Bronze

Pour cette période, nous avons rassemblé 7 individus : la totalité des individus attachés à la culture d'Afanasievo, soit 5 individus dans le sud sibérien (2 dans l'Altaï et 4 en Khakassie), et 1 individu dans l'Altaï mongol. Nous traiterons également avec ces échantillons l'individu 11, présenté comme appartenant à la culture Bol'shemyskaya, datée du énéolithique.

IV.1.1.1.1 Données mitochondriales

Les données mitochondriales obtenues pour cet ensemble ou période sont réparties de la sorte : 2 individus sont affiliés à l'haplogroupe C, 2 individus sont affiliés à l'haplogroupe U (U4 et U5) et 3 individus, provenant de site différents, partagent le même haplotype que nous n'avons pu affilier à un haplogroupe particulier.

Une des données intéressantes est la présence de l'haplogroupe est-eurasien C dans le sud de la Sibérie au début de l'âge du Bronze. Les individus sont porteurs de l'haplotype fondateur. D'après les données de génétique moderne, l'haplogroupe C serait apparu dans l'Est de l'Asie et aurait diffusé depuis la région péri-Baïkal au cours de nombreuses migrations avant et après le dernier maximum glaciaire (Derenko et al., 2010). Les études paléogénétiques avaient déjà mis en évidence la présence de cet haplotype dans la culture contemporaine d'Ust-Tartas dans l'ouest de la Sibérie (Molodin et al., 2012), dans une population du nord-est de l'Europe il y a 3500 ans (Der Sarkissian et al., 2013), ainsi qu'en Ukraine chez des individus néolithiques (Nikitin et al., 2012), suggérant ainsi une diffusion de l'haplogroupe C d'Est en Ouest avant l'âge du Bronze.

Deux individus sont porteurs de l'haplogroupe U (probablement U4 et U5). L'haplogroupe U est présent à des fréquences élevées chez les chasseurs-cueilleurs d'Europe (Bramanti et al., 2009; Der Sarkissian et al., 2013; Malmstrom et al., 2009). L'haplotype U5a, porté par l'individu 3, a été retrouvé chez des individus de l'âge du Bronze à l'ouest de la Sibérie (Molodin et al., 2012). Cet haplotype est aujourd'hui principalement présent dans les populations du Caucase et celles du centre et du nord de l'Europe.

De façon intéressante, l'haplogroupe H est déjà présent chez un individu Afanasievo. L'haplogroupe H est porté par environ 40% de la population ouest-européenne actuelle (Richards et al., 2000). L'haplotype détecté, non retrouvé dans les populations anciennes, présente une large distribution aujourd'hui, principalement dans les populations européennes.

Comme 71 % de la diversité de l'haplogroupe H se trouve en dehors de la région de contrôle de l'ADNmt, l'analyse de la région HV1 seule ne nous permet pas d'être informatif des sous-haplogroupes (Brotherton et al., 2013), il est donc difficile d'aller plus loin dans l'analyse. Néanmoins, ces deux haplogroupes, U et H, suggèrent la présence d'une composante maternelle ouest-eurasienne.

Un second élément particulièrement intéressant est le partage d'un même haplotype par 3 individus exhumés de sites différents : 2 dans le sud sibérien (1 Afanasievo et 1 Bol'shemyskaya) et 1 dans l'Altai mongol. La recherche d'haplotypes partagés n'a pas montré de correspondance parmi les autres populations anciennes et seulement un finlandais actuel partage cette séquence. Celle-ci semble donc assez rare dans les populations actuelles. Néanmoins, malgré le faible effectif étudié, cette distribution au sein d'ensemble différents tend à montrer un pool génique commun à ces individus qui, en plus de partager des caractères culturels, semblent donc partager des caractères génétiques au niveau des lignées maternelles. Ce résultat suggère ainsi une origine génétique commune ou un flux de gènes entre ces populations.

IV.1.1.1.2 Données du chromosome Y

Les haplotypes paternels retrouvés se répartissent au sein de deux haplogroupes. Les individus sud-sibériens appartiennent à l'haplogroupe R1b1a-P297 et parmi eux, 2 ont pu être affiliés au sous-haplogroupe R1b1a2-M269. L'individu de l'Altai mongol appartient quant à lui à l'haplogroupe Q-M242.

La présence de l'haplogroupe R1b1a2-M269 dans le sud de l'Altai tout au début de l'âge du Bronze constitue un des résultats marquants de ce travail.

L'haplogroupe R est surtout fréquent en Europe puisque plus de 50% des hommes en sont porteurs. Mais cet haplogroupe est également présent en Asie Centrale, en Asie du Sud et en Asie de l'Ouest (Chiaroni et al., 2009). La majorité des européens appartient à l'haplogroupe R1, divisé en 2 sous-haplogroupes : R1a et R1b. Ces sous-haplogroupes sont majoritairement retrouvés au niveau de l'Europe de l'Est et de l'Europe de l'Ouest, respectivement, avec une zone de différenciation, en Allemagne. R1b est scindé en deux branches : R1b1a2-M269, qui représente la branche européenne, et R1b1a1-M73 la branche asiatique. Nos échantillons n'appartiennent pas à la branche asiatique M73 puisqu'ils possèdent l'allèle ancestral pour ce marqueur. Deux individus, en revanche, ont pu être associés au sous- haplogroupe R1b1a2-

M269. Le marqueur M269 est présent majoritairement en Europe de l'Ouest suivant un gradient de fréquence Est-Ouest avec un maximum dans le nord-ouest de l'Europe. L'origine de ce marqueur n'est pas clairement définie (Busby et al., 2012). Certains auteurs ont tout d'abord argumenté en faveur d'une origine paléolithique (Rosser et al., 2000; Semino et al., 2000), puis d'autres ont proposé une introduction en Europe avec les fermiers néolithiques (Balaesque et al., 2010; Sjodin and Francois, 2011). Néanmoins, bien que peu de données anciennes sur le chromosome Y aient été publiées en Europe jusqu'à maintenant, l'haplogroupe R1b a été retrouvé en Allemagne chez deux individus de la culture campaniforme, daté d'environ 2500 ans avant J.-C (Lee et al., 2012). Les données actuellement disponibles (Haak et al., 2010; Haak et al., 2008; Lacan et al., 2011a; Lacan et al., 2011b), semblent indiquer que les individus plus anciens ne sont pas porteurs de cet haplogroupe, laissant penser à une entrée de cet haplogroupe en Europe seulement à la fin du néolithique. D'autres données européennes seront bien évidemment nécessaires pour discuter cette hypothèse.

Au sein du sous-haplogroupe R1b1a2-M269, l'étude publiée par Myres et al. (2011) a montré que le marqueur M412 séparait les lignées M269 de l'ouest et du centre de l'Europe de celles rencontrées en Europe de l'Est, dans la région de l'Oural, au Proche-Orient ainsi que dans le Caucase et au Pakistan. Afin d'évaluer l'affinité de nos échantillons anciens avec les individus appartenant à l'haplogroupe R1b-M269 et ses sous-haplogroupes, nous avons construit un réseau phylogénétique. Sur ce réseau, nos individus se placent majoritairement avec des individus appartenant au sous-haplogroupe L23(xM412)², c'est à dire avec des populations du Caucase, d'Europe de l'Est, de la région de l'Oural et du Proche-Orient. Ce résultat suggère un lien avec ces populations de l'est de l'Europe et met en évidence pour la première fois la présence de l'haplogroupe R1b1a2 dans le sud de la Sibérie au début de l'âge du Bronze, chez des individus porteurs des cultures d'Afanasievo et de Bol'shemyskaya. Ainsi, ces individus sont les porteurs les plus anciens de l'haplogroupe R1b1a2 connus à ce jour.

Dans l'Altaï mongol, l'individu Afanasievo est porteur de l'haplogroupe Q-M242. Cet haplogroupe est distribué largement dans le nord de l'Eurasie et trouvé à de fortes fréquences dans le sud de la Sibérie, contrairement à l'Europe, l'Asie de l'Est et le Moyen-Orient où il

² Rappel : la nomenclature suivante « nom de l'haplogroupe ou nom du marqueur » (x nom d'un marqueur), par exemple L23(xM412), signifie que l'individu est porteur de l'allèle muté pour le marqueur L23 mais possède l'allèle ancestral pour le marqueur M412. Cela peut se trouver également sous la forme L23 (non M412).

très peu présent (Karafet et al., 2002). Il serait apparu en Asie puis se serait répandu et différencié avant de passer la Béringie et d'entrer en Amérique. En effet, le sous- haplogroupe Q1a3a1-M3 est retrouvé de façon quasi exclusive chez les amérindiens (Zegura et al., 2004). Cet haplogroupe serait donc une signature génétique paléosibérienne.

IV.1.1.1.3 Données des SNP autosomaux

L'analyse des SNP autosomaux a été possible principalement pour les individus Afanasievo de Khakassie, l'ADN des autres individus semblait trop dégradé pour obtenir l'ensemble des informations nécessaires à l'exploitation des données. Ainsi, l'analyse des caractères pigmentaires (couleur des cheveux et couleur des yeux) montre des phénotypes variés parmi ces individus. Deux individus avaient probablement les yeux bruns et les cheveux foncés alors que deux autres individus avaient probablement les yeux clairs et les cheveux clairs.

L'analyse de 5 SNP AIM a permis d'estimer l'origine biogéographique des individus anciens. Les résultats obtenus montrent clairement une composante européenne très majoritaire dans les génomes des individus Afanasievo.

IV.1.1.1.4 Bilan

Ainsi, d'après l'ensemble des données génétiques, que ce soit les données mitochondriales, du chromosome Y, des marqueurs phénotypiques ou des AIM, il semblerait que ces individus du début de l'âge du Bronze, pour la plupart affiliés à la culture d'Afanasievo, soient proches des populations européennes actuelles, notamment pour les individus du sud de la Sibérie. L'individu mongol porte, quant à lui, une signature génétique paternelle plutôt paléosibérienne.

Les résultats obtenus montrent également un partage du pool génique entre ces individus. En effet, une lignée mitochondriale rare est partagée par 3 individus, dont l'un de l'Altaï mongol, et l'haplogroupe R1b1a2 est présent aussi bien dans l'Altaï russe qu'en Khakassie. Ces résultats suggèrent donc, en plus d'une homogénéité culturelle, une origine commune ou des flux géniques entre ces régions.

IV.1.1.2 Bronze moyen

A la suite de la culture d'Afanasievo, vers le milieu du III^{ème} millénaire avant J.C apparaît, dans le bassin du Minoussinsk, la culture d'Okunevo (13 individus), et au pied de l'Altaï, la culture d'Elunino (4 individus). Dans l'Altaï mongol est retrouvée, sur l'ensemble du massif, la culture Chemurchek (7 individus).

IV.1.1.2.1 Données mitochondriales

Les 13 haplotypes obtenus pour la culture d'Okunevo se répartissent de la façon suivante : 54% appartiennent à des haplogroupes est-eurasiens (C, D et A) et 46% à des haplogroupes ouest-eurasiens (U, H, J et T). Les 3 haplotypes détectés pour les individus de la culture d'Elunino sont à 100% est-eurasiens (C et D). Les individus de la culture Chemurchek, répartis sur l'ensemble de l'Altaï, ont des haplotypes uniquement est-eurasiens au nord et en majorité ouest-eurasiens au sud.

Certains haplogroupes déjà présents dans le groupe précédent sont retrouvés dans ce groupe plus récent, il s'agit des haplogroupes U, H et C. Par contre, de nouveaux haplogroupes ont fait leur apparition : les haplogroupes ouest-eurasiens J et T, ainsi que les haplogroupes est-eurasiens C et D. Les haplogroupes J et T seraient issus du Proche-Orient après les dernières glaciations (Pala et al., 2012). L'haplotype fondateur J présente une large distribution parmi les individus anciens déjà analysés. Il a été retrouvé chez des individus néolithiques d'Europe Centrale (Bramanti et al., 2009; Haak et al., 2010; Haak et al., 2005), du Sud-Ouest (Lacan et al., 2011a; Sampietro et al., 2007), et chez des individus de l'ouest-sibérien datés de la transition entre l'âge du Bronze et l'âge du Fer (Molodin et al., 2012), ainsi que dans l'Altaï mongol à l'âge du Fer (Gonzalez-Ruiz et al., 2012). Parmi les 3 haplotypes T, un haplotype est partagé avec un individu néolithique d'Allemagne (Haak et al., 2005) et un individu du Kazakhstan de l'âge du Fer (Lalueza-Fox et al., 2004).

Nous pouvons ainsi observer l'apport de nouveaux haplogroupes et une diversification des haplotypes dans ces populations plus récentes. La lignée présente chez trois individus du début de l'âge du Bronze n'est plus retrouvée dans ce groupe. Seul l'haplotype H 16362C est partagé entre les deux groupes du Bronze ancien et moyen.

IV.1.1.2.2 Données du chromosome Y

Concernant les données du chromosome Y, une modification très importante est observée. Dans le sud de la Sibérie, seul l'haplogroupe R1b1a2-M269 était présent chez les individus du début de l'âge du Bronze. Or, celui-ci n'est plus retrouvé que chez un individu masculin de la culture d'Okunevo. Le reste du pool génique est constitué des haplogroupes Q et NO(xO).

L'haplogroupe Q était déjà présent dans l'Altai mongol au début de l'âge du Bronze. Si l'individu Afanasievo appartenait au macro-haplogroupe Q*, les individus sont cette fois-ci affiliés aux sous-haplogroupes Q1a3-M346 et Q1a3a-L54. Ces résultats marquent l'apparition d'haplogroupes paléosibériens dans le sud de la Sibérie.

Comme nous l'avons expliqué dans la section « résultats », il n'a pas été possible de déterminer finement l'haplogroupe des individus appartenant à l'ensemble NO(xO). L'haplogroupe N a une large distribution dans le nord de l'Eurasie et présente les fréquences les plus hautes aux latitudes les plus élevées (Derenko et al., 2007).

Ainsi, bien qu'une légère continuité semble exister entre les individus Afanasievo et Okunevo dans le bassin du Minoussinsk, les données suggèrent un changement dans le pool génique des individus masculins.

Dans le massif de l'Altai, l'haplogroupe Y n'a pu être déterminé que pour un individu. Celui-ci portait l'haplogroupe asiatique C-M130 dont la distribution actuelle s'étend de l'Asie du Sud, à l'Asie du Nord-Est ainsi qu'en Océanie et en Amérique.

IV.1.1.2.3 Données des SNP autosomaux

L'étude des caractères phénotypiques montre une majorité de phénotypes yeux bruns et cheveux foncés. Un seul individu semble avoir les yeux bleus. Ainsi, il est possible qu'il y ait eu une modification des phénotypes présents dans le sud de la Sibérie. Ces résultats seraient en accord avec des données anthropologiques qui suggèrent la présence de caractères mongoloïdes chez les individus d'Okunevo.

De même que pour les données phénotypiques, l'étude des SNP AIM suggère des contributions européennes et asiatiques variables entre les individus. En effet, certains individus semblent plus proches des populations européennes avec néanmoins, des valeurs inférieures à celles obtenues pour les individus Afanasievo. D'autres individus semblent avoir

autant de contribution européenne et asiatique, et 3 individus ont une composante asiatique majoritaire au sein de leur génome.

IV.1.1.2.4 Bilan

Au vu de l'ensemble des données génétiques obtenues il semble qu'il y ait eu une modification à la fois du pool génique mitochondrial et du chromosome Y dans la région ciblée à cette période, par l'apport de nouveaux haplogroupes (est- et ouest-eurasiens) d'une part, et par l'augmentation de la proportion d'haplogroupes est-eurasiens d'autre part. Cette modification se retrouve dans les résultats obtenus à partir de l'analyse des marqueurs nucléaires car nous observons une augmentation des phénotypes yeux bruns et cheveux foncés, assez caractéristiques des populations d'Asie. L'étude des AIM supporte cette hypothèse puisque la composante asiatique semble fortement augmentée, et même majoritaire chez plusieurs individus.

IV.1.1.3 Bronze final

Enfin, le troisième groupe temporel est constitué d'individus appartenant aux cultures de Sagsai (11 individus) et de Munkh-Khairkan (3 individus), toutes deux situées dans l'Altai mongol. Nous ne disposons pas d'échantillons provenant du sud de la Sibérie pour cette période.

IV.1.1.3.1 Données mitochondriales

Les individus Sagsai mis au jour dans le nord de l'Altai mongol, pour une partie sur le versant Russe (TU) et l'autre partie sur le versant Chinois (TA), présentent un mélange d'haplotypes ouest-eurasiens et est-eurasiens. La plupart des haplogroupes étaient déjà présents dans le groupe précédent. On peut noter l'apparition de l'haplogroupe T1a, qui est très présent dans les populations actuelles d'Asie Centrale et qui est retrouvé chez un individu de l'âge du Bronze (Lalueza-Fox et al., 2004) et le partage d'une séquence mitochondriale avec un individu du bassin du Tarim. Concernant les individus de la culture Munkh-Khairkan, l'individu situé dans l'ouest de la Mongolie porte l'haplogroupe H alors que les deux individus situés dans la partie plus centrale portent des haplogroupes asiatiques (C et D).

IV.1.1.3.2 Données du chromosome Y

De nouveau, une modification du pool génique est observée à cette période. Celui-ci est composé majoritairement des haplogroupes Q1a3a-L54, R1a1a1b2-Z93 et C-M130 pour les individus Sagsai, et de l'haplogroupe NO (xO) pour les individus Munkh-Khairkan.

L'apparition de l'haplogroupe R1a1a1b2-Z93 est un élément particulièrement intéressant. L'haplogroupe R1a1a a une distribution très large au sein de l'Eurasie, s'étendant de l'Europe de l'Est au sud de l'Asie en passant par le sud-sibérien. Son origine mal connue, est aujourd'hui très discutée. Certains auteurs argumentent en faveur d'une origine est-européenne après le dernier maximum glaciaire (Passarino et al., 2001; Semino et al., 2000), alors que d'autres, du fait de la fréquence très élevée et de la forte diversité des haplotypes présents en Asie du Sud, proposent un foyer d'origine dans cette région (Mirabal et al., 2009; Sharma et al., 2009). Afin d'améliorer la compréhension de la diffusion de cet haplogroupe, plusieurs études se sont attachées à trouver des marqueurs présentant des fréquences discriminantes parmi les différentes populations, définissant ainsi des sous-haplogroupes informatifs de l'origine biogéographique. Ainsi, le marqueur M458 a été décrit par Underhill et al. (2010) comme ayant une distribution restreinte aux populations d'Europe Centrale et à quelques populations d'Europe de l'Est, avec un pic de fréquence en Pologne, mais ne s'étendant pas au-delà de l'Oural. Son aire de répartition semble concorder avec l'aire de répartition de la culture Cordée. Antérieurement à cette étude, un individu ancien associé à la culture Cordée avait déjà été affilié à l'haplogroupe R1a1a (Haak et al., 2008). Ce marqueur est donc présenté comme informatif d'une branche européenne de l'haplogroupe R1a1. R1a1a*(x M458) a, à l'inverse, en Europe une fréquence maximale en Biélorussie et dans le sud-ouest de la Russie. La fréquence d'apparition de ces deux marqueurs semble également pouvoir être corrélée aux différentes langues indo-européennes. En effet, R1a1a-M458 a un pic de fréquence chez les indo-européens de langues slaves et finno-ougriennes alors que R1a1a*(xM458) a un pic de fréquence chez les individus de langues indo-iraniennes et dravidiennes (Underhill et al., 2010). Selon Underhill, la diversité et la fréquence des haplotypes M458 suggèrent une origine durant le début de l'holocène suivi de son expansion associée à des événements culturels préhistoriques tels que la diffusion de la culture Cordée. L'absence de M458 en dehors de l'Europe irait ainsi à l'encontre d'un flux génique depuis l'Europe de l'Est vers l'Asie, au moins depuis environ 5000 ans. En 2012, Pamjav et al. ont identifié deux nouveaux marqueurs, Z280 et Z93, ayant des aires de répartition restreintes à l'Asie Centrale et l'Asie du Sud pour le marqueur Z93, et à l'Europe de l'Est pour le

marqueur Z280. Ces zones géographiques correspondent à l'aire de répartition de R1a1a*(xM458) chez Underhill. Les aires de répartition du marqueur Z280 et Z93 se chevauchent au niveau de l'Asie Centrale. L'origine de la différenciation de R1a1a pourrait donc se trouver quelque part entre l'Asie du Sud-Est et l'Europe de l'Est. Les steppes eurasiennes (Ukraine, sud de la Russie, Kazakhstan et Caucase) et le Moyen-Orient sont présentés comme des candidats potentiels (Pamjav et al., 2012). Dans notre étude les individus anciens porteurs de l'haplogroupe R1a1a sont tous porteurs du sous-haplogroupe R1a1a1b2-Z93 correspondant, selon la littérature, à la branche centre et sud-asiatique. L'haplogroupe R1a1a-M17 (les sous-haplogroupes n'étant pas encore définis) avait déjà été retrouvé dans le sud de la Sibérie, chez des individus de l'âge du Bronze (Andronovo (1800-1400 avant J.-C) et Karasouk (1400-800 av. J.-C)) et de l'âge du Fer (Keyser et al., 2009), dans le bassin du Tarim il y a 3900 ans (Li et al., 2010), et chez un individu Xiongnu à l'est de la Mongolie (Kim et al., 2010). Ces résultats suggèrent donc un pool génique commun à ces individus.

IV.1.1.3.3 Données autosomales

Les phénotypes observés parmi ces individus sont très divers ; 2 individus ont les yeux bleus ou intermédiaires et 5 individus sur 8 ont les cheveux bruns.

L'analyse des AIM révèle que ces individus ont, pour la plupart (7/11), une composante européenne majoritaire.

IV.1.1.3.4 Bilan

Si au niveau des lignées maternelles aucun changement majeur dans le pool génique n'est observé, l'élément principal est l'apparition de l'haplogroupe R1a1a1b2-Z93, déjà présent chez des individus sud-sibériens de l'âge du Bronze et dans le bassin du Tarim.

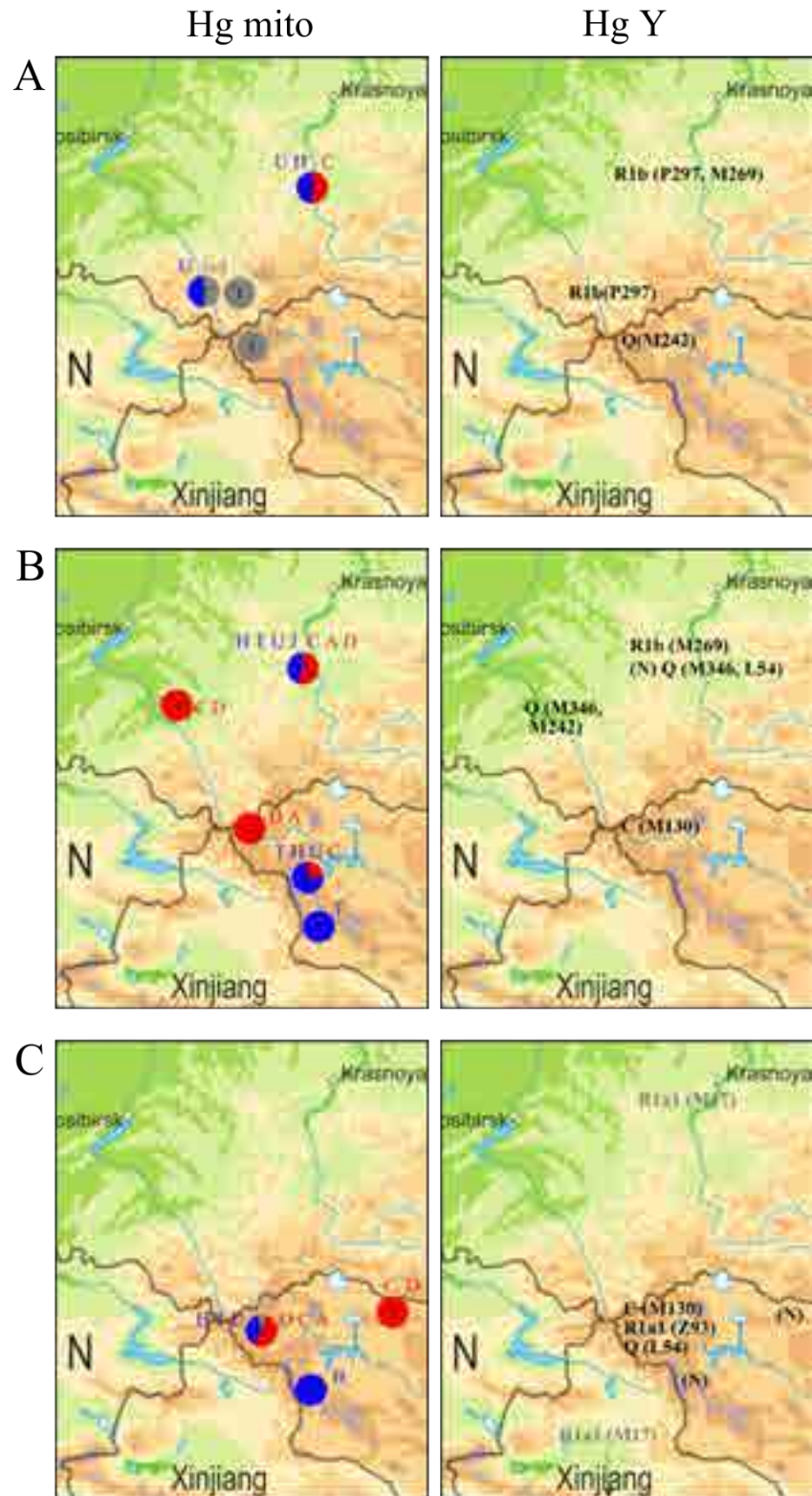


Figure 19 : Répartition géographique et temporelle des haplogroupes mitochondriaux et Y retrouvés dans les échantillons étudiés

A : début de l'âge du Bronze (cultures d'Afanasievo et de Bol'shemyskaya) B : Bronze moyen (cultures d'Okunevo, d'Elunino et de Chemurchek), C : fin de l'âge du Bronze (cultures de Sagsai et de Munkh Khairkan). Les cercles représentent les proportions d'haplogroupes ouest-eurasiens (en bleu), est-eurasiens (en rouge) et autres (en gris) pour chaque groupe. L'effectif total est indiqué au centre du cercle. Les haplogroupes indiqués en gris correspondent à la distribution de l'haplogroupe R1a1a décrit dans la littérature (Keyser et al., 2009; Li et al., 2010)

IV.1.2 Dynamique de peuplement de la région, premières hypothèses

Dans un premier temps, l'étude et la compréhension de l'évolution du pool génique au sein même de la région nous ont permis de comprendre le processus micro-évolutif. Il est maintenant nécessaire de le discuter au vu des études paléogénétiques précédentes mais aussi par rapport aux données et hypothèses archéologiques proposées, afin de mieux appréhender les différentes hypothèses de peuplement de cette région.

Dans ce travail, nous avons étudié l'une des cultures les plus intéressantes de l'âge du Bronze, la culture d'Afanasievo. L'étude de cette culture serait un élément clé pour la compréhension des premières migrations indo-européennes. Selon des données archéologiques et anthropologiques, son origine a été associée aux cultures de l'ouest des steppes eurasiennes, la culture de Yamna ou la culture antérieure de Repin (Anthony, 2007). Les données moléculaires obtenues pour les individus Afanasievo, dans leur ensemble, suggèrent une affinité avec les populations européennes actuelles. Ces données génétiques semblent donc en accord avec les données archéologiques et anthropologiques, qui proposent une origine au niveau des steppes d'Europe de l'Est pour ces individus. Néanmoins, une étude récente a montré que l'aire de répartition des populations ouest-eurasiennes s'étendait jusqu'à l'ouest du lac Baïkal durant le Paléolithique, et que cette occupation avait perduré pendant le dernier maximum glaciaire (l'individu étudié était porteur de l'haplogroupe mitochondrial U et de l'haplogroupe Y R) (Raghavan et al., 2014). De la même façon, une étude publiée en 2006 a montré la présence de l'haplogroupe ouest-urasien U5 dans la région Baïkal au Néolithique (Mooder et al., 2006). Une origine locale des individus Afanasievo, à partir de peuples caucasoïdes déjà présents, pourrait donc être une hypothèse également envisageable. Seul l'accès à des sujets sud-sibériens / altaïques antérieurs à la culture d'Afanasievo pourra nous permettre de répondre de façon plus fiable à cette question. De la même façon, aucune donnée génétique du début de l'âge du Bronze, et même un peu avant (culture de Repin, de Yamna), n'est actuellement disponible pour les populations des steppes eurasiennes, alors que de nombreuses hypothèses s'appuient sur ces populations. L'accès à ces échantillons et leur étude est donc une étape importante pour mieux appréhender ces mouvements migratoires.

Malgré qu'un biais imputable au faible échantillonnage soit indéniable, le pool génique observé dans les cultures du début de l'âge du Bronze semble, d'après les données en notre

possession, avoir été modifié dans les cultures suivantes d'Okunevo et d'Elunino, et n'avoir que peu persisté puisque l'haplogroupe R1b1a2-M269, très présent au début de l'âge du Bronze, n'est plus présent que chez un individu dans la culture d'Okunevo, et que la séquence mitochondriale partagée par trois individus n'est plus retrouvée. En revanche, on voit clairement apparaître une composante asiatique, notamment au niveau du chromosome Y. Cette composante pourrait venir d'un retour de populations paléosibériennes du nord-est de la Sibérie (van Geel et al., 2004). Néanmoins, l'origine de cette composante asiatique est, au vu des données actuelles, difficile à déterminer. Ces résultats génétiques semblent cependant en accord avec les données anthropologiques qui présentaient la population d'Okunevo comme une population plus « mongoloïde » que celle d'Afanasievo. Cependant, il semble que des flux géniques aient persisté en provenance de l'Ouest, puisque l'on note l'apparition de nouveaux haplogroupes mitochondriaux ouest-eurasiens.

Au milieu du II^{ème} millénaire, ce pool génique pourrait avoir de nouveau évolué avec l'apparition de l'haplogroupe R1a1a1b2 dans l'Altaï mongol. Cet haplogroupe, comme décrit précédemment, avait déjà été retrouvé chez plusieurs individus de l'âge du Bronze dans le sud de la Sibérie ainsi que dans le bassin du Tarim. De plus, une lignée paternelle (17 STR) est partagée entre deux individus inhumés sur deux sites différents dans l'Altaï, et une lignée mitochondriale non retrouvée dans la population actuelle est partagée entre un individu de l'Altaï et un du bassin du Tarim. Les résultats obtenus dans ce travail semblent donc montrer un partage de caractères génétiques entre ces trois ensembles géographiques. Ces résultats suggèrent donc une origine génétique commune, ou un flux de gènes entre ces régions, qui aurait modifié considérablement le pool génique présent dans la région. Du fait de sa répartition géographique, certains auteurs ont associé l'haplogroupe R1a1a à l'expansion des peuples indo-européens (Semino et al., 2000; Wells et al., 2001). L'analyse du marqueur M458, réalisée après le travail de thèse de Caroline Bouakaze chez les sujets sud-sibériens de la culture d'Andronovo, les plus anciens de notre échantillon, a révélé la présence de l'allèle ancestral pour ce marqueur. Ces individus seraient donc R1a1a(xM458), haplogroupe présent chez les populations actuelles de langues indo-iraniennes (Underhill et al., 2010). Ainsi, ces résultats semblent en accord avec le lien proposé entre la culture d'Andronovo et la diffusion de la branche indo-iranienne des langues indo-européennes (Lamberg-Karlovsky, 2002). La diffusion de l'haplogroupe R1a1a(xM458) pourrait ainsi être liée à la diffusion de la culture d'Andronovo puisque les individus R1a1a1b2 étudiés dans l'Altaï sont postérieurs à cette culture, et que des hypothèses ont été émises entre la culture d'Andronovo et le peuplement

du bassin du Tarim (où R1a1a a également été retrouvé mais pour qui les sous-haplogroupes ne sont pas connus). Cette hypothèse est bien évidemment à considérer avec prudence. Seul l'augmentation du nombre de données sur le chromosome Y des populations de cette région nous permettra de la discuter.

Dans l'article publié par Keyser et al. en 2009, le partage de caractéristiques phénotypiques (apparence euroïde) et de caractères génétiques (partage de lignées mitochondriales) entre des individus de la culture d'Andronovo et les momies du bassin du Tarim avaient permis d'argumenter en faveur de l'hypothèse d'un peuplement du bassin du Tarim *via* les steppes du Nord. Des données du chromosome Y sont venues supporter cette hypothèse quelque temps après en montrant la présence exclusive de lignées Y R1a1a chez 7 individus du site de Xiaohe dans le Bassin du Tarim (Li et al., 2010). Dans l'hypothèse des steppes, deux vagues de migrations se seraient succédées, une première par des individus porteurs de la culture d'Afanasievo puis une seconde par les porteurs de la culture d'Andronovo. Si tous les résultats sont en accord avec un lien entre Andronovo et le bassin du Tarim, a contrario, le lien entre la culture d'Afanasievo et le bassin du Tarim est plus discutable. En effet, si les marqueurs génétiques analysés suggèrent une origine ouest-eurasienne pour ces individus et si certains sont porteurs d'un phénotype plutôt euroïde, les lignées paternelles sont, elles, différentes (R1b vs R1a). Ces données sont néanmoins à nuancer puisque seuls quelques individus Afanasievo ont été analysés dans ce travail et les données des lignées paternelles du bassin du Tarim ne proviennent que d'un seul site.

Enfin, si le massif de l'Altaï culmine parfois à plus de 4000 mètres d'altitude, les données obtenues dans ce travail montrent un partage de lignées (maternelles ou paternelles) entre les individus du sud de la Sibérie et de l'Altaï mongol, ou tout du moins une homogénéité dans les haplogroupes présents. Il semble donc que le relief montagneux n'ait pas été un frein aux mouvements de populations dans cette région et que des flux géniques ont eu lieu, au moins dès le début de l'âge du Bronze. Les différences de pools géniques semblent, d'après ce travail, beaucoup plus liées à la composante culturelle et temporelle qu'à la composante géographique.

IV.2 Discussion sur la stratégie expérimentale employée et perspectives

Les résultats présentés dans ce travail nous ont permis d'apporter de nouveaux éléments quant à la compréhension de la dynamique de peuplement de la région ciblée. Néanmoins, travailler avec de rares et précieux échantillons d'ADN ancien dégradés, nécessite d'adopter une stratégie expérimentale réfléchie et cohérente vis à vis la quantité de matériel disponible, aux données recherchées et aux méthodes mises en place pour y parvenir.

La stratégie proposée dans cette étude est une stratégie « multi systèmes » qui consiste à analyser à la fois des marqueurs uniparentaux, localisés sur l'ADN mitochondrial et le chromosome Y, ainsi que des marqueurs biparentaux (STR et SNP autosomaux). Cette stratégie est particulièrement intéressante pour plusieurs raisons. D'une part, elle permet à contrario de nombreuses études d'ADN ancien, d'étudier à la fois l'histoire maternelle et paternelle des populations anciennes mais également d'obtenir des informations sur leur apparence physique. D'autre part, elle s'avère particulièrement intéressante pour étudier les liens de proches parentés (Keyser-Tracqui et al., 2003a). Néanmoins, afin d'être appliquée, cette stratégie requiert des critères particuliers. En effet, l'analyse des marqueurs nucléaires nécessite de travailler à partir d'un ADN relativement bien préservé. Il sera toujours plus difficile d'obtenir les données nucléaires que mitochondriales puisque l'ADN mitochondrial est présent en quantité plus importante dans la cellule. Dans le présent travail, pour deux individus, nous avons pu obtenir uniquement des données mitochondriales. De la même façon ce phénomène s'observe avec les individus immatures pour lesquels il est plus difficile d'obtenir des données nucléaires. Le second point faible de cette stratégie est la grande consommation d'ADN. Nous devons disposer d'une quantité suffisante d'ADN pour réaliser l'ensemble des analyses mais également pour répéter ces analyses afin de respecter les critères d'analyses de l'ADN ancien et ainsi assurer l'authenticité des résultats.

Ainsi, dans l'objectif d'obtenir toujours plus d'informations à partir d'une faible quantité de matériel il est nécessaire, en parallèle des études paléogénétiques à proprement parler, d'avoir un développement méthodologique constant. Ce développement s'applique à toutes les étapes de l'analyse.

L'amélioration des techniques d'extraction d'ADN pour lesquelles nous avons déjà vu, ces dernières années, une nette progression puisque de 1 à 2 grammes de poudre nécessaires à l'extraction d'ADN il y a quelques années (Ricaud et al., 2004b), nous travaillons actuellement à partir de 10 fois moins soit 150 à 200 mg de poudre. Néanmoins la compréhension du mécanisme d'adsorption/désorption de l'ADN sur l'hydroxyapatite reste encore à approfondir afin d'améliorer les méthodes d'extraction et de libération de l'ADN. Le laboratoire travaille dans ce sens puisqu'il s'agit du sujet de thèse d'une étudiante de notre équipe, Anne Grunenwald. Nous testons également en parallèle un protocole d'extraction comportant une étape de décalcification par l'EDTA plus longue, afin d'évaluer l'impact sur la quantité d'ADN extrait. En effet, une déminéralisation totale de l'os semble, selon certains auteurs, augmenter la quantité et la qualité d'ADN récupéré (Amory et al., 2012; Loreille et al., 2007).

L'amélioration doit également se faire au niveau des techniques utilisées. Dans ce travail nous avons principalement utilisé, pour l'analyse des SNP, des réactions d'amplification multiplexes suivies d'une détection par spectrométrie de masse MALDI-TOF et plus particulièrement par le système commercialisé par la société SEQUENOM. Ce système a la propriété d'avoir été spécifiquement développé pour l'analyse d'ADN et non l'analyse protéique, évitant ainsi le long processus de mise au point pour analyser des acides nucléiques sur des systèmes dédiés aux analyses protéiques comme présenté dans le travail de thèse de Elizabeth Petkovski (Petkovski, 2006). Des précédents travaux ont montré l'intérêt de cette méthode pour l'analyse d'ADN ancien (Mendisco et al., 2011). Dans ce travail nous l'avons utilisée pour la détermination des haplogroupes mitochondriaux et Y, mais aussi pour la détermination des caractères phénotypiques et l'analyse des SNP AIM. Contrairement à l'ADN mitochondrial et au chromosome Y, pour lesquels il n'y a pas de souci d'interprétation puisque le résultat obtenu est binaire (un allèle ou rien), les marqueurs localisés sur les autosomes, peuvent être hétérozygotes. Si cela ne pose pas de problème avec un ADN de bonne qualité, cela est plus problématique à partir d'un ADN dégradé et présent en faible quantité. En effet, comme observé lors de la gamme de dilution, pour des petites concentrations en ADN on voit apparaître des phénomènes de perte d'allèles entraînant la transformation d'un génotype hétérozygote en génotype homozygote. Si les analyses ne sont pas répétées cela peut entraîner un biais dans les résultats. Dans le cas du test HIris Plex, selon le poids du SNP impliqué dans la prédiction du phénotype, le résultat obtenu sera donc plus ou moins fiable. Les analyses ont été répétées afin de limiter ce phénomène mais pour autant

le typage de SNP reste plus délicat à appliquer sur des marqueurs diploïdes que sur des systèmes haploïdes. La mesure du taux d'hétérozygotie est également un problème observé lors des analyses avec les séquenceurs de nouvelles générations du fait des faibles taux de couverture généralement obtenus (Veeramah and Hammer, 2014).

Afin d'obtenir l'haplotype puis l'haplogroupe mitochondrial des individus anciens, nous avons séquencé la région HV1 de l'ADN mitochondrial. Pour réaliser ces analyses nous travaillons à chaque fois à partir de 20 µl d'ADN et celles-ci doivent être répétées afin d'assurer l'authenticité des séquences obtenues. Nous nous basons ensuite sur l'haplotype HV1 ainsi que sur les quelques positions codantes typées pour en déduire l'haplogroupe et comparer avec les données disponibles dans la littérature, afin de déterminer l'origine biogéographique des lignées maternelles. Or, aujourd'hui, les études des génomes mitochondriaux totaux des populations modernes ont permis d'affiner très fortement la phylogénie et de décrire ainsi de nombreux sous-haplogroupes souvent associés à des événements migratoires particuliers, et avec des répartitions géographiques très informatives (Derenko et al., 2010; Pala et al., 2012). Ce même phénomène apparaît pour l'étude du chromosome Y (Myres et al., 2011), pour lequel, aujourd'hui, nous disposons de séquences complètes ou de typage d'un très grand nombre de SNP. Ainsi, lors de nos analyses nous sommes désormais parfois limités dans l'interprétation du fait de la faible résolution phylogénétique obtenue comparée à celle présentée dans la littérature depuis quelques mois. Les méthodes de séquençage et de typage utilisées dans ce travail semblent donc peu adaptées à cette évolution puisque nous sommes limités par la quantité d'ADN disponible, par le nombre de SNP (pas plus de 30) à analyser par réactions multiplexes, et par le nombre de multiplexes à réaliser sur chaque échantillon. Une des alternatives aujourd'hui envisageable est l'utilisation des séquenceurs de nouvelles générations (NGS). En effet, les progrès réalisés ces dix dernières années permettent d'avoir des outils capables de séquencer un génome entier en quelques heures. Si le séquençage du génome entier est envisageable pour des spécimens particuliers ou quelques individus, comme cela est réalisé actuellement, il reste difficilement applicable à des études de populations anciennes par toutes les équipes de recherche. En effet, cela impliquerait un coût actuellement encore trop élevé et nécessiterait des capacités d'analyse très importantes. Néanmoins, des solutions alternatives, plus ciblées sur les données recherchées, telles que l'analyse des génomes mitochondriaux totaux et d'un nombre important de SNP semblent envisageables par l'utilisation des séquenceurs de paillasse de 2^{ème} génération, tels que le GS Junior de Roche, le MiSeq de Illumina ou l'Ion Torrent PGM

de Life technologies. Ces séquenceurs ont la capacité d'analyser des petits génomes ou des ensembles de gènes en quelques heures sur plusieurs échantillons en parallèle. Même si toutes les méthodes sont en théorie applicables sur des échantillons d'ADN dégradé, la méthode Ion Torrent PGM de Life technologies semble particulièrement intéressante. Une partie du développement des kits utilisables sur ce séquenceur est réalisée en vue d'une utilisation dans le domaine de la criminalistique, correspondant donc à nos problématiques, aussi bien en termes d'analyses à réaliser, qu'en termes de « qualité » du substrat de départ. C'est pourquoi des études sont actuellement réalisées par d'autres équipes afin de séquencer des génomes mitochondriaux à partir de petits fragments d'ADN et d'un ADN de mauvaise qualité (Parson et al., 2013). De plus, un kit, particulièrement intéressant pour notre équipe, est en cours de commercialisation. Il s'agit du kit « EVC SNP panel » qui devrait permettre l'analyse de 245 SNP dont les 24 SNP du test Hiris Plex et de SNP AIM permettant de déterminer l'origine biogéographique (Jurado, 2013). Ce kit n'étant pas encore commercialisé, il est possible que des modifications y soient apportées. Ces panels étant destinés à une utilisation de routine en criminalistique, des logiciels d'analyse ont également été développés permettant ainsi d'avoir une optimisation du temps d'analyse, bien qu'en France, contrairement à d'autres pays tels que les Pays-Bas et les pays anglo-saxons, la législation ne permette pas la détermination de caractères phénotypiques dans un cadre judiciaire. En effet, dans le domaine judiciaire, seule l'identification d'une personne par ses empreintes génétiques est envisageable, l'examen des caractéristiques génétiques étant réservé au domaine médical et à la recherche scientifique (art 16-10 et 16-11 du Code Civil). C'est pourquoi nous pouvons les étudier dans le cadre de projets de recherche et donc en même temps, tester et participer au développement de ces nouvelles technologies.

L'utilisation de toutes ces méthodes semble donc prometteuse pour les prochaines études d'ADN ancien. Malgré le progrès technologique, la qualité des échantillons de départ reste cependant le point le plus important. Il sera donc nécessaire, dans un premier temps, d'évaluer les capacités réelles d'analyses de ces nouvelles technologies, à partir de ce type de substrats. Néanmoins, les premiers résultats obtenus pour le domaine criminalistique à partir de l'Ion torrent semblent encourageants car, bien que le système soit proposé pour travailler à partir de 1 ng d'ADN, une étude récente, après avoir apporté une légère modification du protocole en augmentant le nombre de cycles d'amplification, a prouvé son efficacité pour des quantités d'ADN allant jusqu'à 100 pg lors de l'analyse des 136 SNP inclus dans le kit HID développé pour l'identification humaine (Seo et al., 2013).

Enfin, la collaboration avec les archéologues est essentielle afin d'assurer une qualité optimale des échantillons prélevés, c'est à dire limiter les contaminations et la dégradation de l'ADN. Cela passe par la mise en place de bonnes pratiques de prélèvement et de stockage des échantillons et, lorsque ce que cela est possible, l'utilisation de matériel fraîchement exhumé (Pruvost et al., 2007).

V CONCLUSION

Si dans un premier temps, les problématiques de peuplement étaient uniquement abordées à partir des données archéologiques, anthropologiques et linguistiques, ces dernières années, les données génétiques sont venues apporter de nouveaux éléments. D'abord remises en question, les études paléogénétiques sont finalement devenues essentielles pour la compréhension des événements migratoires. C'est par la confrontation de l'ensemble de ces données complémentaires que des hypothèses solides peuvent désormais être établies.

Dans ce travail nous nous sommes intéressés à la dynamique du peuplement du sud de la Sibérie et de l'Altaï pendant l'âge du Bronze. Pour cela, 69 individus provenant de sites et de cultures différentes ont été étudiés. Afin d'estimer l'origine biogéographique de ces sujets, nous avons adopté une stratégie, consistant, en plus de l'étude de l'ADN mitochondrial couramment étudié dans les études d'ADN ancien, à étudier les lignées paternelles de ces individus, ainsi que des marqueurs autosomaux, informatifs de l'origine biogéographique (AIM) et des caractères pigmentaires tels que la couleur des yeux et des cheveux.

L'état de préservation de l'ADN était très variable d'un site à l'autre mais néanmoins, la relativement bonne conservation des échantillons nous a permis d'obtenir des données mitochondriales pour 62 % d'entre eux. Le taux de réussite des analyses des marqueurs autosomaux s'est lui révélé plus bas (environ 40%).

Grâce aux données génétiques obtenues nous avons pu mettre en évidence une probable évolution du pool génique au sein de la région étudiée. La confrontation de ces données aux autres données de paléogénétiques mais aussi aux données archéo- et anthropo-logiques nous a permis de discuter et d'émettre des hypothèses concernant les mouvements migratoires qu'a connus cette région.

Tout d'abord, il semble qu'en plus de partager des caractères culturels, les porteurs de la culture d'Afanasievo possédaient des caractères génétiques communs, laissant penser qu'ils proviennent d'un même groupe ancestral. Les données obtenues semblent également indiquer que ces individus étaient génétiquement proches des populations européennes actuelles. Ce résultat semble donc en accord avec l'hypothèse proposée par de nombreux archéo- et anthropologues, consistant à présenter la culture d'Afanasievo comme une diffusion vers l'est, jusque dans l'Altaï, des porteurs des cultures de l'ouest des steppes, la culture de Yamna ou de Repin. L'analyse du pool génique des populations ayant succédé à la culture d'Afanasievo n'a pas révélé de fortes continuités génétiques mais plutôt un apport de lignées est-eurasiennes, notamment au niveau des lignées paternelles. Des populations asiatiques

autochtones se seraient donc probablement mélangées avec ces peuples plutôt caucasoïdes. Enfin, durant le II^{ème} millénaire avant J.-C, le pool génique de ces populations semble avoir été de nouveau modifié avec la présence, sur une large zone s'étendant du sud de la Sibérie au bassin du Tarim en passant par l'Altai, de l'haplogroupe R1a1a. L'haplogroupe (R1a1a (xM458)) pourrait ainsi être associé à l'expansion de la culture d'Andronovo et aux locuteurs de langues indo-iraniennes.

Les résultats présentés dans ce travail, semblent, d'après les données du chromosome Y, écarter le lien entre les individus Afanasievo et le peuplement du bassin du Tarim. Néanmoins, le petit nombre de sujets étudiés dans les deux populations peut fortement biaiser l'analyse. Pour apporter de nouveaux éléments quant au processus de peuplement du bassin du Tarim, des analyses supplémentaires et une analyse plus poussée des lignées R1a1a dans cette région, semblent vraiment nécessaires.

Le faible échantillonnage observé par groupe ainsi que le processus de recrutement funéraire peut bien évidemment avoir biaisé une partie des analyses. C'est pourquoi, afin de discuter ces hypothèses, il est indispensable d'envisager d'autres analyses paléogénétiques. L'accès à des spécimens anciens du début de l'âge du Bronze ayant peuplé les steppes de l'Ouest semble indispensable pour la compréhension de ces phénomènes migratoires puisque de nombreuses hypothèses restent à valider ou invalider. De la même façon, l'accès à des sujets antérieurs à la culture d'Afanasievo serait très intéressant, afin d'estimer l'arrivée de la composante européenne dans cette région. Néanmoins, il semble que peu de vestiges archéologiques de la période néolithique soient présents dans le sud de la Sibérie. Concernant les études génétiques elles-mêmes, bien que les informations soient plus difficiles à obtenir qu'à partir de l'ADN mitochondrial, il est nécessaire de réaliser des études sur le chromosome Y. Peu de données sont aujourd'hui disponibles or, nous avons pu voir que les lignées paternelles sont souvent très informatives des mouvements migratoires. En effet, les femmes ayant un taux de migration plus élevé que les hommes (Stoneking, 1998), on observera plus facilement une forte homogénéité entre les populations. Ainsi, les lignées maternelles peuvent, dans certains cas, se révéler moins informatives pour retracer les mouvements migratoires.

Néanmoins, ces mouvements migratoires se sont poursuivis, puisque d'après les études portées sur les sujets de l'âge du Fer, ces modifications du pool génique ont continué. Durant le I^{er} millénaire avant J.-C, la composante mitochondriale d'origine asiatique semble avoir

augmenté aussi bien dans le sud de la Sibérie qu'au Kazakhstan (Keyser et al., 2009; Lalueza-Fox et al., 2004). Les données mitochondriales obtenues sur les populations Scythes montrent un pool génique composé à la fois d'haplotypes est- et ouest-eurasiens, et de même que les observations issues de ce travail, une similarité génétique en plus d'une homogénéité culturelle entre les différents groupes Scythes étudiés dans la région (Gonzalez-Ruiz et al., 2012). Aucune donnée sur le chromosome Y n'est disponible or, cela pourrait certainement procurer des données intéressantes pour discuter de l'émergence et de l'origine de ces peuples Scythes et pour ainsi reconstruire l'histoire complexe des populations des steppes eurasiennes.

A l'issue de ce travail et avant d'envisager les prochaines études paléogénétiques, se pose la également la question de la méthodologie employée. Comme décrit précédemment, la discipline connaît depuis ces dernières années une mutation technologique importante. Actuellement, de nombreux travaux de paléogénomique rapportent le séquençage de spécimens rares, très anciens en repoussant toujours un peu plus loin les limites d'analyse de l'ADN ancien. Sans aller jusque-là, il est néanmoins essentiel de bien réfléchir aux outils actuellement disponibles pour réaliser les prochaines études de populations. Dans ce contexte, la pluridisciplinarité de notre équipe nous permet de bénéficier de l'accès aux technologies les plus récentes en matière d'identification génétique des individus, et ainsi, pourquoi pas, d'envisager leur utilisation dans nos prochaines études de paléogénétique voire de paléogénomique.

VI BIBLIOGRAPHIE

A

Abu-Amero, K.K., Larruga, J.M., Cabrera, V.M., and Gonzalez, A.M. (2008). Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evol Biol* 8, 45.

Achakzai, N.M., Rahman, Z., Shahzad, M.S., Daud, S., Zar, M.S., Israr, M., Husnain, T., Willuweit, S., and Roewer, L. (2012). Y-chromosomal STR analysis in the Pashtun population of Southern Afghanistan. *Forensic Sci Int Genet* 6, e103-105.

Adams, S.M., Bosch, E., Balaesque, P.L., Ballereau, S.J., Lee, A.C., Arroyo, E., Lopez-Parra, A.M., Aler, M., Grifo, M.S., Brion, M., *et al.* (2008). The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *American journal of human genetics* 83, 725-736.

Adcock, G.J., Dennis, E.S., Easteal, S., Huttley, G.A., Jermin, L.S., Peacock, W.J., and Thorne, A. (2001). Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proceedings of the National Academy of Sciences of the United States of America* 98, 537-542.

Alakoc, Y.D., Gokcumen, O., Tug, A., Gultekin, T., Gulec, E., and Schurr, T.G. (2010). Y-chromosome and autosomal STR diversity in four proximate settlements in Central Anatolia. *Forensic Sci Int Genet* 4, e135-137.

Alam, S., Ali, M.E., Ferdous, A., Hossain, T., Hasan, M.M., and Akhteruzzaman, S. (2010). Haplotype diversity of 17 Y-chromosomal STR loci in the Bangladeshi population. *Forensic Sci Int Genet* 4, e59-60.

Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T., Willerslev, E., *et al.* (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings Biological sciences / The Royal Society* 279, 4724-4733.

Allwood, J.S., and Harbison, S. (2013). SNP model development for the prediction of eye colour in New Zealand. *Forensic science international Genetics* 7, 444-452.

Alshamali, F., Brandstatter, A., Zimmermann, B., and Parson, W. (2008). Mitochondrial DNA control region variation in Dubai, United Arab Emirates. *Forensic Sci Int Genet* 2, e9-10.

Amory, S., Huel, R., Bilic, A., Loreille, O., and Parsons, T.J. (2012). Automatable full demineralization DNA extraction procedure from degraded skeletal remains. *Forensic science international Genetics* 6, 398-406.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457-465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics* 23, 147.

Anthony, D.W. (2007). *The Horse, The Wheel and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton.

B

Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., *et al.* (2010). A predominantly neolithic origin for European paternal lineages. *PLoS biology* 8, e1000285.

Ballard, D.J., Phillips, C., Thacker, C.R., and Court, D.S. (2006). Y chromosome STR haplotype data for an Irish population. *Forensic Sci Int* 161, 64-68.

Ballard, D.J., Phillips, C., Thacker, C.R., Robson, C., Revoir, A.P., and Syndercombe Court, D. (2005). Y chromosome STR haplotypes in three UK populations. *Forensic Sci Int* 152, 289-305.

Bandelt, H.J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution* 16, 37-48.

Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogvali, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *American journal of human genetics* 90, 675-684.

Bennett, C.C., and Kaestle, F.A. (2010). Investigation of ancient DNA from Western Siberia and the Sargat culture. *Hum Biol* 82, 143-156.

Berger, B., Lindinger, A., Niederstatter, H., Grubwieser, P., and Parson, W. (2005). Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *Int J Legal Med* 119, 241-246.

Biagini, P., Theves, C., Balaresque, P., Geraut, A., Cannet, C., Keyser, C., Nikolaeva, D., Gerard, P., Duchesne, S., Orlando, L., *et al.* (2012). Variola virus in a 300-year-old Siberian mummy. *The New England journal of medicine* 367, 2057-2059.

Bianucci, R., Rahalison, L., Massa, E.R., Peluso, A., Ferroglio, E., and Signoli, M. (2008). Technical note: a rapid diagnostic test detects plague in ancient human remains: an example of the interaction between archeological and biological approaches (southeastern France, 16th-18th centuries). *American journal of physical anthropology* 136, 361-367.

Blyakharchuk, T.A., and Chernova, N.A. (2013). Vegetation and climate in the Western Sayan Mts according to pollen data from Lugovoe Mire as background for prehistoric cultural change in southern Middle Siberia. *Quaternary Science Reviews* 75, 22-42.

Bodner, M., Zimmermann, B., Rock, A., Kloss-Brandstatter, A., Horst, D., Horst, B., Sengchanh, S., Sanguansermisri, T., Horst, J., Kramer, T., *et al.* (2011). Southeast Asian diversity: first insights into the complex mtDNA structure of Laos. *BMC Evol Biol* 11, 49.

Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., *et al.* (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506-510.

- Bosch, E., Calafell, F., Gonzalez-Neira, A., Flaiz, C., Mateu, E., Scheil, H.G., Huckenbeck, W., Efremovska, L., Mikerezi, I., Xirotiris, N., *et al.* (2006). Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70, 459-487.
- Bouakaze, C., Keyser, C., Crubezy, E., Montagnon, D., and Ludes, B. (2009). Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *International journal of legal medicine* 123, 315-325.
- Bouwman, A.S., Chilvers, E.R., Brown, K.A., and Brown, T.A. (2006). Brief communication: identification of the authentic ancient DNA sequence in a human bone contaminated with modern DNA. *American journal of physical anthropology* 131, 428-431.
- Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., *et al.* (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-140.
- Brandstatter, A., Klein, R., Duftner, N., Wiegand, P., and Parson, W. (2006). Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int J Legal Med* 120, 310-314.
- Branicki, W., Brudnik, U., Kupiec, T., Wolanska-Nowak, P., and Wojas-Pelc, A. (2007). Determination of phenotype associated SNPs in the MC1R gene. *Journal of forensic sciences* 52, 349-354.
- Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pospiech, E., Walsh, S., Kupiec, T., Wojas-Pelc, A., and Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Human genetics* 129, 443-454.
- Brenner, C.H. (1997). Symbolic kinship program. *Genetics* 145, 535-542.
- Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., *et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* 104, 14616-14621.
- Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Sarkissian, C.D., Ganslmeier, R., Friederich, S., *et al.* (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature communications* 4, 1764.
- Burger, J., Kirchner, M., Bramanti, B., Haak, W., and Thomas, M.G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences of the United States of America* 104, 3736-3741.
- Busby, G.B., Brisighelli, F., Sanchez-Diz, P., Ramos-Luis, E., Martinez-Cadenas, C., Thomas, M.G., Bradley, D.G., Gusmao, L., Winney, B., Bodmer, W., *et al.* (2012). The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings Biological sciences / The Royal Society* 279, 884-892.

C

- Cadenas, A.M., Regueiro, M., Gayden, T., Singh, N., Zhivotovsky, L.A., Underhill, P.A., and Herrera, R.J. (2007). Male amelogenin dropouts: phylogenetic context, origins and implications. *Forensic science international* 166, 155-163.
- Cai, X., Qin, Z., Wen, B., Xu, S., Wang, Y., Lu, Y., Wei, L., Wang, C., Li, S., Huang, X., *et al.* (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 6, e24282.
- Cakir, A.H., Celebioglu, A., and Yardimci, E. (2004). Y-STR haplotypes in Central Anatolia region of Turkey. *Forensic Sci Int* 144, 59-64.
- Calafell, F., Underhill, P., Tolun, A., Angelicheva, D., and Kalaydjieva, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60, 35-49.
- Campos, P.F., Craig, O.E., Turner-Walker, G., Peacock, E., Willerslev, E., and Gilbert, M.T. (2012). DNA in ancient bone - where is it located and how should we extract it? *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* 194, 7-16.
- Cano, R.J., Poinar, H.N., Pieniazek, N.J., Acra, A., and Poinar, G.O., Jr. (1993). Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* 363, 536-538.
- Chang, Y.M., Perumal, R., Keat, P.Y., and Kuehn, D.L. (2007). Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci Int* 167, 70-76.
- Charlier, P., Huynh-Charlier, I., Poupon, J., Keyser, C., Lancelot, E., Favier, D., Vignal, J.N., Sorel, P., Chaillot, P.F., Boano, R., *et al.* (2010). Multidisciplinary medical identification of a French king's head (Henri IV). *Bmj* 341, c6805.
- Chen, F., Wang, S.Y., Zhang, R.Z., Hu, Y.H., Gao, G.F., Liu, Y.H., and Kong, Q.P. (2008). Analysis of mitochondrial DNA polymorphisms in Guangdong Han Chinese. *Forensic Sci Int Genet* 2, 150-153.
- Chiaroni, J., Underhill, P.A., and Cavalli-Sforza, L.L. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America* 106, 20174-20179.
- Chikisheva, T.A., Gubina, M.A., Kulikov, I.V., Karafet, T.M., Voevoda, M.I., and Romaschenko, A.G. (2007). A paleogenetic study of the prehistoric populations of the Altai. *Archeology, Ethnology & Anthropology of Eurasia* 4, 130-142.
- Cieslak, M., Pruvost, M., Benecke, N., Hofreiter, M., Morales, A., Reissmann, M., and Ludwig, A. (2010). Origin and history of mitochondrial DNA lineages in domestic horses. *PloS one* 5, e15311.

Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G.L., Lillie, A.S., Roseman, C.C., Lin, A.A., Prince, K., *et al.* (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114, 127-148.

Clisson, I., Keyser, C., Francfort, H.P., Crubezy, E., Samashev, Z., and Ludes, B. (2002). Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC). *International journal of legal medicine* 116, 304-308.

Coble, M.D., Loreille, O.M., Wadhams, M.J., Edson, S.M., Maynard, K., Meyer, C.E., Niederstatter, H., Berger, C., Berger, B., Falsetti, A.B., *et al.* (2009). Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PloS one* 4, e4838.

Comas, D., Calafell, F., Bendukidze, N., Fananas, L., and Bertranpetit, J. (2000). Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol* 112, 5-16.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., and Bertranpetit, J. (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13, 1067-1077.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., *et al.* (1998). Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 63, 1824-1838.

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., and Ward, R. (2001). Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409, 704-707.

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

Cui, Y., Li, C., Gao, S., Xie, C., and Zhou, H. (2010). Early Eurasian migration traces in the Tarim Basin revealed by mtDNA polymorphisms. *Am J Phys Anthropol* 142, 558-564.

D

De Maesschalck, K., Vanhoutte, E., Knaepen, K., Vanderheyden, N., Cassiman, J.J., and Decorte, R. (2005). Y-chromosomal STR haplotypes in a Belgian population sample and identification of a micro-variant with a flanking site mutation at DYS19. *Forensic Sci Int* 152, 89-94.

Deguilloux, M.F., Pemonge, M.H., Mendisco, F., Thibon, D., Cartron, I., and Castex, D. (2014). Ancient DNA and kinship analysis of human remains deposited in Merovingian necropolis sarcophagi (Jau Dignac et Loirac, France, 7th–8th century AD). *Journal of Archaeological Science* 41, 399-405.

- Deguilloux, M.F., Soler, L., Pemonge, M.H., Scarre, C., Joussaume, R., and Laporte, L. (2011). News from the west: ancient DNA from a French megalithic burial chamber. *American journal of physical anthropology* 144, 108-118.
- Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., Zaporozhchenko, V., Gronenborn, D., Moiseyev, V., Kolpakov, E., *et al.* (2013). Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS genetics* 9, e1003296.
- Derbeneva, O.A., Starikovskaia, E.B., Volod'ko, N.V., Wallace, D.C., and Sukernik, R.I. (2002a). [Mitochondrial DNA variation in Kets and Nganasans and the early peoples of Northern Eurasia]. *Genetika* 38, 1554-1560.
- Derbeneva, O.A., Starikovskaya, E.B., Wallace, D.C., and Sukernik, R.I. (2002b). Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70, 1009-1014.
- Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2007). Y-chromosome haplogroup N dispersals from south Siberia to Europe. *Journal of human genetics* 52, 763-770.
- Derenko, M., Malyarchuk, B., Denisova, G.A., Wozniak, M., Dambueva, I., Dorzhu, C., Luzina, F., Miscicka-Sliwka, D., and Zakharov, I. (2006). Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Human genetics* 118, 591-604.
- Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., Dambueva, I., and Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PloS one* 5, e15214.
- Derenko, M.V., Grzybowski, T., Malyarchuk, B.A., Dambueva, I.K., Denisova, G.A., Czarny, J., Dorzhu, C.M., Kakpakov, V.T., Miscicka-Sliwka, D., Wozniak, M., *et al.* (2003). Diversity of mitochondrial DNA lineages in South Siberia. *Annals of human genetics* 67, 391-411.
- DeSalle, R., Gatesy, J., Wheeler, W., and Grimaldi, D. (1992). DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257, 1933-1936.
- Di Benedetto, G., Erguven, A., Stenico, M., Castri, L., Bertorelle, G., Togan, I., and Barbujani, G. (2001). DNA diversity and population admixture in Anatolia. *Am J Phys Anthropol* 115, 144-156.
- Di Gaetano, C., Cerutti, N., Crobu, F., Robino, C., Inturri, S., Gino, S., Guarrera, S., Underhill, P.A., King, R.J., Romano, V., *et al.* (2009). Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur J Hum Genet* 17, 91-99.
- Dimo-Simonin, N., Grange, F., Taroni, F., Brandt-Casadevall, C., and Mangin, P. (2000). Forensic evaluation of mtDNA in a population from south west Switzerland. *Int J Legal Med* 113, 89-97.

Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., and Raoult, D. (1998). Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proceedings of the National Academy of Sciences of the United States of America* 95, 12637-12640.

Drancourt, M., Signoli, M., Dang, L.V., Bizot, B., Roux, V., Tzortzis, S., and Raoult, D. (2007). *Yersinia pestis* Orientalis in remains of ancient plague patients. *Emerging infectious diseases* 13, 332-333.

Draus-Barini, J., Walsh, S., Pospiech, E., Kupiec, T., Glab, H., Branicki, W., and Kayser, M. (2013). Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative genetics* 4, 3.

Dubut, V., Chollet, L., Murail, P., Cartault, F., Beraud-Colomb, E., Serre, M., and Mogentale-Profizi, N. (2004). mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet* 12, 293-300.

Dulik, M.C., Osipova, L.P., and Schurr, T.G. (2011). Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PloS one* 6, e17548.

Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T.G. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *American journal of human genetics* 90, 229-246.

Dupuy, B.M., Stenersen, M., Lu, T.T., and Olaisen, B. (2006). Geographical heterogeneity of Y-chromosomal lineages in Norway. *Forensic Sci Int* 164, 10-19.

E

Edwards, C.J., Bollongino, R., Scheu, A., Chamberlain, A., Tresset, A., Vigne, J.D., Baird, J.F., Larson, G., Ho, S.Y., Heupink, T.H., *et al.* (2007). Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings Biological sciences / The Royal Society* 274, 1377-1385.

Egyed, B., Furedi, S., and Padar, Z. (2006). Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers. *Forensic Sci Int* 164, 257-265.

Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K.W., and Hansen, L. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Human genetics* 123, 177-187.

F

Fedorova, S.A., Bermisheva, M.A., VILLEMS, R., Maksimova, N.R., and Khusnutdinova, E.K. (2003). [Analysis of mitochondrial DNA haplotypes in yakut population]. *Mol Biol (Mosk)* 37, 643-653.

Fondevila, M., Phillips, C., Santos, C., Freire Aradas, A., Vallone, P.M., Butler, J.M., Lareu, M.V., and Carracedo, A. (2013). Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic science international Genetics* 7, 63-74.

Frank, W.E., Ralph, H.C., and Tahir, M.A. (2008). Y chromosome STR haplotypes and allele frequencies in a southern Indian male population. *J Forensic Sci* 53, 248-251.

G

Gabriel, S., Ziaugra, L., and Tabbaa, D. (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al] Chapter 2, Unit 2* 12.

Gamba, C., Fernandez, E., Tirado, M., Deguilloux, M.F., Pemonge, M.H., Utrilla, P., Edo, M., Molist, M., Rasteiro, R., Chikhi, L., *et al.* (2012). Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Molecular ecology* 21, 45-56.

Gao, S., Cui, Y., Yang, Y., Duan, R., Abuduresule, I., Mair, V.H., Zhu, H., and Zhou, H. (2008). Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China. *Science in China Series C, Life sciences / Chinese Academy of Sciences* 51, 205-213.

Garcia-Bour, J., Perez-Perez, A., Alvarez, S., Fernandez, E., Lopez-Parra, A.M., Arroyo-Pardo, E., and Turbon, D. (2004). Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization. *American journal of physical anthropology* 123, 361-370.

Gayden, T., Bukhari, A., Chennakrishnaiah, S., Stojkovic, O., and Herrera, R.J. (2011). Y-chromosomal microsatellite diversity in three culturally defined regions of historical Tibet. *Forensic Sci Int Genet* 6, 437-446.

Ghiani, M.E., Mameli, A., Piras, G., Berti, A., Calo, C.M., and Vona, G. (2009). Population data for Y-chromosome haplotypes defined by AmpFI STR Yfiler PCR amplification kit in North Sardinia (Italy). *Collegium antropologicum* 33, 643-651.

Gilbert, M.T., Bandelt, H.J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. *Trends in ecology & evolution* 20, 541-544.

Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P.F., *et al.* (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787-1789.

Gilbert, M.T., Menez, L., Janaway, R.C., Tobin, D.J., Cooper, A., and Wilson, A.S. (2006). Resistance of degraded hair shafts to contaminant DNA. *Forensic science international* 156, 208-212.

Gilbert, M.T., Wilson, A.S., Bunce, M., Hansen, A.J., Willerslev, E., Shapiro, B., Higham, T.F., Richards, M.P., O'Connell, T.C., Tobin, D.J., *et al.* (2004). Ancient mitochondrial DNA from hair. *Current biology : CB* 14, R463-464.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature genetics* 6, 130-135.

Ginolhac, A., Vilstrup, J., Stenderup, J., Rasmussen, M., Stiller, M., Shapiro, B., Zazula, G., Froese, D., Steinmann, K.E., Thompson, J.F., *et al.* (2012). Improving the performance of true single molecule sequencing for ancient DNA. *BMC genomics* 13, 177.

Goebel, T. (1999). Pleistocene Human Colonization of Siberia and Peopling of the Americas: An Ecological Approach. *Evolutionary Anthropology*, 208-227.

Gokcumen, O., Dulik, M.C., Pai, A.A., Zhadanov, S.I., Rubinstein, S., Osipova, L.P., Andreenkov, O.V., Tabikhanova, L.E., Gubina, M.A., Labuda, D., *et al.* (2008). Genetic variation in the enigmatic Altaian Kazakhs of South-Central Russia: insights into Turkic population history. *Am J Phys Anthropol* 136, 278-293.

Gonzalez-Ruiz, M., Santos, C., Jordana, X., Simon, M., Lalueza-Fox, C., Gigli, E., Aluja, M.P., and Malgosa, A. (2012). Tracing the origin of the east-west population admixture in the Altai region (Central Asia). *PloS one* 7, e48904.

Gonzalez, A.M., Brehm, A., Perez, J.A., Maca-Meyer, N., Flores, C., and Cabrera, V.M. (2003). Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* 120, 391-404.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., *et al.* (2010). A draft sequence of the Neandertal genome. *Science* 328, 710-722.

Grzybowski, T., Malyarchuk, B.A., Derenko, M.V., Perkova, M.A., Bednarek, J., and Wozniak, M. (2007). Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet* 1, 141-147.

Gubina, M.A., Damba, L.D., Babenko, L., Romashchenko, V.N., and Voevoda, A.G. (2013). [Haplotype diversity in mtDNA and Y-chromosome in populations of Altai-Sayan region]. *Genetika* 49, 376-391.

H

Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S., Brandt, G., Schwarz, C., Nicklisch, N., *et al.* (2010). Ancient DNA from

European early neolithic farmers reveals their near eastern affinities. *PLoS biology* 8, e1000536.

Haak, W., Brandt, G., de Jong, H.N., Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C., Pike, A.W., Meller, H., and Alt, K.W. (2008). Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proceedings of the National Academy of Sciences of the United States of America* 105, 18226-18231.

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tanzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., *et al.* (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310, 1016-1018.

Haas, C., Wangenstein, T., Giezendanner, N., Kratzer, A., and Bar, W. (2006). Y-chromosome STR haplotypes in a population sample from Switzerland (Zurich area). *Forensic Sci Int* 158, 213-218.

Haber, M., Platt, D.E., Ashrafian Bonab, M., Youhanna, S.C., Soria-Hernanz, D.F., Martinez-Cruz, B., Douaihy, B., Ghassibe-Sabbagh, M., Rafatpanah, H., Ghanbari, M., *et al.* (2012). Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events. *PLoS One* 7, e34288.

Haensch, S., Bianucci, R., Signoli, M., Rajerison, M., Schultz, M., Kacki, S., Vermunt, M., Weston, D.A., Hurst, D., Achtman, M., *et al.* (2010). Distinct clones of *Yersinia pestis* caused the black death. *PLoS pathogens* 6, e1001134.

Haile, J., Froese, D.G., Macphee, R.D., Roberts, R.G., Arnold, L.J., Reyes, A.V., Rasmussen, M., Nielsen, R., Brook, B.W., Robinson, S., *et al.* (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* 106, 22352-22357.

Halder, I., Shriver, M., Thomas, M., Fernandez, J.R., and Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Human mutation* 29, 648-658.

Haliti, N., Carapina, M., Masic, M., Strinovic, D., Klaric, I.M., and Kubat, M. (2009). Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croatians. *Forensic Sci Int Genet* 3, e137-138.

Hallenberg, C., Nielsen, K., Simonsen, B., Sanchez, J., and Morling, N. (2005). Y-chromosome STR haplotypes in Danes. *Forensic Sci Int* 155, 205-210.

Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., *et al.* (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS genetics* 4, e1000074.

Hansen, A.J., Mitchell, D.L., Wiuf, C., Paniker, L., Brand, T.B., Binladen, J., Gilichinsky, D.A., Ronn, R., and Willerslev, E. (2006). Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173, 1175-1179.

Hashiyada, M., Umetsu, K., Yuasa, I., Tamura, A., Matsusue, A., Suzuki, K., Kashimura, S., and Funayama, M. (2008). Population genetics of 17 Y-chromosomal STR loci in Japanese. *Forensic Sci Int Genet* 2, e69-70.

Hedman, M., Brandstatter, A., Pimenoff, V., Sistonen, P., Palo, J.U., Parson, W., and Sajantila, A. (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci Int* 172, 171-178.

Hemphill, B.E., and Mallory, J.P. (2004). Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *American journal of physical anthropology* 124, 199-222.

Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., *et al.* (2008). Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PloS one* 3, e3426.

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

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., and Wilson, A.C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282-284.

Ho, S.Y., Heupink, T.H., Rambaut, A., and Shapiro, B. (2007). Bayesian estimation of sequence damage in ancient DNA. *Molecular biology and evolution* 24, 1416-1422.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. *Nature reviews Genetics* 2, 353-359.

Hohoff, C., Dewa, K., Sibbing, U., Hoppe, K., Forster, P., and Brinkmann, B. (2007). Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *Int J Legal Med* 121, 359-363.

Holmlund, G., Nilsson, H., Karlsson, A., and Lindblom, B. (2006). Y-chromosome STR haplotypes in Sweden. *Forensic Sci Int* 160, 66-79.

Huang, T.Y., Hsu, Y.T., Li, J.M., Chung, J.H., and Shun, C.T. (2008). Polymorphism of 17 Y-STR loci in Taiwan population. *Forensic Sci Int* 174, 249-254.

I

Imaizumi, K., Parsons, T.J., Yoshino, M., and Holland, M.M. (2002). A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med* 116, 68-73.

Irwin, J.A., Saunier, J.L., Strouss, K.M., Diegoli, T.M., Sturk, K.A., O'Callaghan, J.E., Paintner, C.D., Hohoff, C., Brinkmann, B., and Parsons, T.J. (2008). Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med* 122, 257-259.

J

Jenkins, D.L., Davis, L.G., Stafford, T.W., Jr., Campos, P.F., Hockett, B., Jones, G.T., Cummings, L.S., Yost, C., Connolly, T.J., Yohe, R.M., 2nd, *et al.* (2012). Clovis age Western Stemmed projectile points and human coprolites at the Paisley Caves. *Science* 337, 223-228.

Jin, H.J., Kwak, K.D., Hong, S.B., Shin, D.J., Han, M.S., Tyler-Smith, C., and Kim, W. (2006). Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: an expanded Korean population database. *Forensic Sci Int* 158, 125-130.

Jurado, T. (2013). Human Identification Applications for the Ion PGMTM System. Oral communication, 2013 Future trends in forensic DNA technology, User meetings, Life Technologies.

K

Kalinowski, S., Wagner, A., and Taper, M. (2006). ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* 6, 576-579.

Karafet, T.M., Osipova, L.P., Gubina, M.A., Posukh, O.L., Zegura, S.L., and Hammer, M.F. (2002). High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Human biology* 74, 761-789.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L.F., Jespersen, J., Bakker, E., *et al.* (2001). An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *American journal of human genetics* 68, 990-1018.

Kayser, M., Liu, F., Janssens, A.C., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M.M., van Ijcken, W.F., *et al.* (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *American journal of human genetics* 82, 411-423.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., *et al.* (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature communications* 3, 698.

Keyser-Tracqui, C., Crubezy, E., and Ludes, B. (2003a). 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., Ricaut, F.X., Blandin, P., and Ludes, B. (2003b). French allele frequencies and haplotypes of nine Y-specific STRs. *J Forensic Sci* 48, 242-244.

Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V.G., Montagnon, D., Reis, T., and Ludes, B. (2009). Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Human genetics* 126, 395-410.

- Khar'kov, V.N., Khamina, K.V., Medvedeva, O.F., Shtygasheva, O.V., and Stepanov, V.A. (2011). [Genetic diversity of Khakassian gene pool: subethnic differensiation and the structure of Y-chromosome haplogroups,]. *Molekuliarnaia biologii* 45, 446-458.
- Khar'kov, V.N., Medvedeva, O.F., Luzina, F.A., Kolbasko, A.V., Gafarov, N.I., Puzyrev, V.P., and Stepanov, V.A. (2009). [Comparative characteristics of the gene pool of Teleuts inferred from Y-chromosomal marker data]. *Genetika* 45, 1132-1142.
- Kim, K., Brenner, C.H., Mair, V.H., Lee, K.H., Kim, J.H., Gelegdorj, E., Batbold, N., Song, Y.C., Yun, H.W., Chang, E.J., *et al.* (2010). A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *American journal of physical anthropology* 142, 429-440.
- Kim, S.H., Kim, N.Y., Hong, S.B., Cho, N.S., Kim, J.J., Han, M.S., and Kim, W. (2008). Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic Sci Int Genet* 2, e9-10.
- King, R.J., Di Cristofaro, J., Kouvatsi, A., Triantaphyllidis, C., Scheidel, W., Myres, N.M., Lin, A.A., Eissautier, A., Mitchell, M., Binder, D., *et al.* (2011). The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. *BMC evolutionary biology* 11, 69.
- King, R.J., Ozcan, S.S., Carter, T., Kalfoglu, E., Atasoy, S., Triantaphyllidis, C., Kouvatsi, A., Lin, A.A., Chow, C.E., Zhivotovsky, L.A., *et al.* (2008). Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. *Ann Hum Genet* 72, 205-214.
- Klaric, I.M., Lauc, L.B., Pericic, M., Janicijevic, B., Terzic, R., Colak, I., Kvesic, A., and Rudan, P. (2005). Evaluation of Y-STR variation in Bosnian and Herzegovinian population. *Forensic Sci Int* 154, 252-256.
- Kolman, C.J., Sambuughin, N., and Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142, 1321-1334.
- Kong, Q.P., Yao, Y.G., Liu, M., Shen, S.P., Chen, C., Zhu, C.L., Palanichamy, M.G., and Zhang, Y.P. (2003). Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 113, 391-405.
- Kosoy, R., Nassir, R., Tian, C., White, P.A., Butler, L.M., Silva, G., Kittles, R., Alarcon-Riquelme, M.E., Gregersen, P.K., Belmont, J.W., *et al.* (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human mutation* 30, 69-78.
- Kovalev, A.A., and Erdenebaatar, D. (2009). Discovery of new cultures of the Bronze Age in Mongolia according to the data obtained by the international central asian archaeological expedition. *current archaeological research in Mongolia*.
- Kovatsi, L., Saunier, J.L., and Irwin, J.A. (2009). Population genetics of Y-chromosome STRs in a population of Northern Greeks. *Forensic Sci Int Genet* 4, e21-22.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.J., Hanni, C., Fortea, J., de la Rasilla, M., *et al.* (2007a). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current biology : CB* 17, 1908-1912.

Krause, J., Orlando, L., Serre, D., Viola, B., Prufer, K., Richards, M.P., Hublin, J.J., Hanni, C., Derevianko, A.P., and Pääbo, S. (2007b). Neanderthals in central Asia and Siberia. *Nature* 449, 902-904.

Kwak, K.D., Suren, G., Tundewrentsen, S., Jin, H.J., Hong, S.B., Han, M.S., and Kim, W. (2006). Y-chromosome STR haplotype profiling in the Mongolian population. *Legal medicine* 8, 58-61.

L

Lacan, M., Keyser, C., Ricaut, F.X., Brucato, N., Duranthon, F., Guilaine, J., Crubezy, E., and Ludes, B. (2011a). Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9788-9791.

Lacan, M., Keyser, C., Ricaut, F.X., Brucato, N., Tarrus, J., Bosch, A., Guilaine, J., Crubezy, E., and Ludes, B. (2011b). Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proceedings of the National Academy of Sciences of the United States of America* 108, 18255-18259.

Lacau, H., Bukhari, A., Gayden, T., La Salvia, J., Regueiro, M., Stojkovic, O., and Herrera, R.J. (2011). Y-STR profiling in two Afghanistan populations. *Leg Med (Tokyo)* 13, 103-108.
Lalueza-Fox, C., Gigli, E., Bini, C., Calafell, F., Luiselli, D., Pelotti, S., and Pettener, D. (2011). Genetic analysis of the presumptive blood from Louis XVI, King of France. *Forensic science international Genetics* 5, 459-463.

Lalueza-Fox, C., Rompler, H., Caramelli, D., Staubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., *et al.* (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318, 1453-1455.

Lalueza-Fox, C., Sampietro, M.L., Gilbert, M.T., Castri, L., Facchini, F., Pettener, D., and Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proceedings Biological sciences / The Royal Society* 271, 941-947.

Lalueza, C., Perez-Perez, A., Prats, E., Cornudella, L., and Turbon, D. (1997). Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. *Human molecular genetics* 6, 41-46.

Lamberg-Karlovsky, C.C. (2002). Archaeology and Language. *Current Anthropology* 43, 63-88.

Lamers, R., Hayter, S., and Matheson, C.D. (2009). Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. *Journal of molecular evolution* 68, 40-55.

- Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.L., and Lahermo, P. (2008). Migration waves to the Baltic Sea region. *Ann Hum Genet* 72, 337-348.
- Larson, G., Liu, R., Zhao, X., Yuan, J., Fuller, D., Barton, L., Dobney, K., Fan, Q., Gu, Z., Liu, X.H., *et al.* (2010). Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* 107, 7686-7691.
- Lebedynsky, I. (2011). *Les Scythes. éditions errance 2ème édition.*
- Lee, E.J., Makarewicz, C., Renneberg, R., Harder, M., Krause-Kyora, B., Muller, S., Ostritz, S., Fehren-Schmitz, L., Schreiber, S., Muller, J., *et al.* (2012). Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany. *American journal of physical anthropology* 148, 571-579.
- Lessig, R., Edelmann, J., Kleemann, W.J., and Kozhemyako, V. (2006). Population data of Y-chromosomal STRs in Russian males of the Primorye region population. *Forensic Sci Int* 159, 71-76.
- Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., Mair, V.H., Xu, Z., Zhang, Q., Abuduresule, I., *et al.* (2010). Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC biology* 8, 15.
- Li, H., Zhao, X., Zhao, Y., Li, C., Si, D., Zhou, H., and Cui, Y. (2011). Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J Hum Genet* 56, 815-822.
- Li, W., Hu, L., Chen, S., and Qiu, N. (2004). 10 Y-STRs haplotypes in Chinese. *Forensic Sci Int* 139, 85-88.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.
- Liu, C., Wang, S.Y., Zhao, M., Xu, Z.Y., Hu, Y.H., Chen, F., Zhang, R.Z., Gao, G.F., Yu, Y.S., and Kong, Q.P. (2011). Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. *Forensic Sci Int Genet* 5, e4-10.
- Liu, F., van Duijn, K., Vingerling, J.R., Hofman, A., Uitterlinden, A.G., Janssens, A.C., and Kayser, M. (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current biology : CB* 19, R192-193.
- Llamas, B., Holland, M.L., Chen, K., Cropley, J.E., Cooper, A., and Suter, C.M. (2012). High-resolution analysis of cytosine methylation in ancient DNA. *PloS one* 7, e30226.
- Loreille, O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P., and Hanni, C. (2001). Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Current biology : CB* 11, 200-203.

Loreille, O.M., Diegoli, T.M., Irwin, J.A., Coble, M.D., and Parsons, T.J. (2007). High efficiency DNA extraction from bone by total demineralization. *Forensic science international Genetics* 1, 191-195.

Lutz, S., Weisser, H.J., Heizmann, J., and Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111, 67-77.

M

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., and Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64, 232-249.

Mallory, J.P., and Mair, V.H. (2000). *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. London: Thames and Hudson.

Malmstrom, H., Gilbert, M.T., Thomas, M.G., Brandstrom, M., Stora, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Gotherstrom, A., *et al.* (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Current biology : CB* 19, 1758-1762.

Malmstrom, H., Linderholm, A., Liden, K., Stora, J., Molnar, P., Holmlund, G., Jakobsson, M., and Gotherstrom, A. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC evolutionary biology* 10, 89.

Malyarchuk, B., Derenko, M., Denisova, G., and Kravtsova, O. (2010). Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol Biol Evol* 27, 2220-2226.

Malyarchuk, B., Derenko, M., Denisova, G., Maksimov, A., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2011). Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *Journal of human genetics* 56, 583-588.

Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Drobnic, K., and Miscicka-Sliwka, D. (2003). Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67, 412-425.

Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Wozniak, M., and Miscicka-Sliwka, D. (2002). Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66, 261-283.

Malyarchuk, B.A., Perkova, M.A., Derenko, M.V., Vanecek, T., Lazur, J., and Gomolcak, P. (2008). Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet* 72, 228-240.

Malyarchuk, B.A., Vanecek, T., Perkova, M.A., Derenko, M.V., and Sip, M. (2006). Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol* 78, 681-696.

- Martinez, L., Underhill, P.A., Zhivotovsky, L.A., Gayden, T., Moschonas, N.K., Chow, C.E., Conti, S., Mamolini, E., Cavalli-Sforza, L.L., and Herrera, R.J. (2007). Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet* 15, 485-493.
- McEvoy, B., Richards, M., Forster, P., and Bradley, D.G. (2004). The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* 75, 693-702.
- Meinila, M., Finnila, S., and Majamaa, K. (2001). Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered* 52, 160-170.
- Mendisco, F., Keyser, C., Hollard, C., Seldes, V., Nielsen, A.E., Crubezy, E., and Ludes, B. (2011). Application of the iPLEX Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: benefits for ancient population studies. *Electrophoresis* 32, 386-393.
- Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., *et al.* (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26.
- Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.L., Martinez, I., Gracia, A., de Castro, J.M., Carbonell, E., *et al.* (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505, 403-406.
- Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., *et al.* (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222-226.
- Mirabal, S., Regueiro, M., Cadenas, A.M., Cavalli-Sforza, L.L., Underhill, P.A., Verbenko, D.A., Limborska, S.A., and Herrera, R.J. (2009). Y-chromosome distribution within the geolinguistic landscape of northwestern Russia. *European journal of human genetics : EJHG* 17, 1260-1273.
- Mirabal, S., Varljen, T., Gayden, T., Regueiro, M., Vujovic, S., Popovic, D., Djuric, M., Stojkovic, O., and Herrera, R.J. (2010). Human Y-chromosome short tandem repeats: a tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkan Peninsula. *American journal of physical anthropology* 142, 380-390.
- Molodin, V.I., Pilipenko, A.S., Romaschenko, A.G., Zhuravlev, A.A., Trapezov, R.O., Chikisheva, T.A., and Pozdnyakov, D.V. (2012). Human migrations in the southern region of the West Siberian Plain during the Bronze Age: Archaeological, palaeogenetic and anthropological data. *Population dynamics in prehistory and early history*, Ed DE GRUYTER, 93-111.
- Mooder, K.P., Schurr, T.G., Bamforth, F.J., Bazaliiski, V.I., and Savel'ev, N.A. (2006). Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. *American journal of physical anthropology* 129, 349-361.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology* 155, 335-350.

Muzzio, M., Ramallo, V., Motti, J.M., Santos, M.R., Lopez Camelo, J.S., and Bailliet, G. (2011). Software for Y-haplogroup predictions: a word of caution. *International journal of legal medicine* 125, 143-147.

Myres, N.M., Rootsi, S., Lin, A.A., Jarve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusbayev, B., *et al.* (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European journal of human genetics : EJHG* 19, 95-101.

N

Nair, S.P., Geetha, A., and Jagannath, C. (2011). Y-short tandem repeat haplotype and paternal lineage of the Ezhava population of Kerala, south India. *Croatian medical journal* 52, 344-350.

Nasidze, I., Quinque, D., Dupanloup, I., Cordaux, R., Kokshunova, L., and Stoneking, M. (2005). Genetic evidence for the Mongolian ancestry of Kalmyks. *American journal of physical anthropology* 128, 846-854.

Nasidze, I., Schadlich, H., and Stoneking, M. (2003). Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Sci Int* 137, 85-93.

Nassir, R., Kosoy, R., Tian, C., White, P.A., Butler, L.M., Silva, G., Kittles, R., Alarcon-Riquelme, M.E., Gregersen, P.K., Belmont, J.W., *et al.* (2009). An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC genetics* 10, 39.

Newton, J.R. (2011). Ancient mitochondrial DNA from pre-historical south-eastern Europe: the presence of east eurasian haplogroups provides evidence of interactions with south siberians across the central asian steppe belt. Master Theses, Grand Valley State University.

Nievergelt, C.M., Maihofer, A.X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K.K., and Kidd, J.R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative genetics* 4, 13.

Nikitin, A.G., Newton, J.R., and Potekhina, I.D. (2012). Mitochondrial haplogroup C in ancient mitochondrial DNA from Ukraine extends the presence of East Eurasian genetic lineages in Neolithic Central and Eastern Europe. *Journal of human genetics* 57, 610-612.

O

Olalde, I., Allentoft, M.E., Sanchez-Quinto, F., Santpere, G., Chiang, C.W., Degiorgio, M., Prado-Martinez, J., Rodriguez, J.A., Rasmussen, S., Quilez, J., *et al.* (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*.

Orlando, L., Ginolhac, A., Raghavan, M., Vilstrup, J., Rasmussen, M., Magnussen, K., Steinmann, K.E., Kapranov, P., Thompson, J.F., Zazula, G., *et al.* (2011). True single-molecule DNA sequencing of a pleistocene horse bone. *Genome research* 21, 1705-1719.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., *et al.* (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74-78.

Outram, A.K., Stear, N.A., Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., Thorpe, N., and Evershed, R.P. (2009). The earliest horse harnessing and milking. *Science* 323, 1332-1335.

Overballe-Petersen, S., Orlando, L., and Willerslev, E. (2012). Next-generation sequencing offers new insights into DNA degradation. *Trends in biotechnology* 30, 364-368.

Ovodov, N.D., Crockford, S.J., Kuzmin, Y.V., Higham, T.F., Hodgins, G.W., and van der Plicht, J. (2011). A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the Last Glacial Maximum. *PloS one* 6, e22821.

P

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314, 644-645.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* 86, 1939-1943.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual review of genetics* 38, 645-679.

Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., and Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120, 334-353.

Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., and Stoneking, M. (2007). Mating patterns amongst Siberian reindeer herders: inferences from mtDNA and Y-chromosomal analyses. *American journal of physical anthropology* 133, 1013-1027.

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Hooshier Kashani, B., *et al.* (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *American journal of human genetics* 90, 915-924.

Palmer, S.A., Smith, O., and Allaby, R.G. (2012). The blossoming of plant archaeogenetics. *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* 194, 146-156.

- Pamjav, H., Feher, T., Nemeth, E., and Padar, Z. (2012). Brief communication: new Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1. *American journal of physical anthropology* 149, 611-615.
- Parson, W., Huber, G., Niederstatter, H., Mayr-Eduardoff, M., Strobl, C., Röck, W., Bodner, M., and Consortium, E.-N. (2013). Next Generation mtGenome Sequencing of Forensic samples with the Ion Torrent PGM. Oral communication, 25th ISFG World Congress, Melbourne, Australia.
- Parson, W., Parsons, T.J., Scheithauer, R., and Holland, M.M. (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.
- Passarino, G., Cavalleri, G.L., Lin, A.A., Cavalli-Sforza, L.L., Borresen-Dale, A.L., and Underhill, P.A. (2002). Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *Eur J Hum Genet* 10, 521-529.
- Passarino, G., Semino, O., Magri, C., Al-Zahery, N., Benuzzi, G., Quintana-Murci, L., Andellnovic, S., Bullc-Jakus, F., Liu, A., Arslan, A., *et al.* (2001). The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. *Human immunology* 62, 922-932.
- Pedersen, J.S., Valen, E., Vargas Velazquez, A.M., Parker, B.J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D.J., Kelly, T.K., Vang, S., *et al.* (2013). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome research*.
- Peng, M.S., He, J.D., Liu, H.X., and Zhang, Y.P. (2011). Tracing the legacy of the early Hainan Islanders--a perspective from mitochondrial DNA. *BMC Evol Biol* 11, 46.
- Pepinski, W., Niemcunowicz-Janica, A., Ptaszynska-Sarosiek, I., Skawronska, M., Koc-Zorawska, E., Janica, J., and Soltyszewski, I. (2004). Population genetics of Y-chromosome STRs in a population of Podlasie, Northeastern Poland. *Forensic Sci Int* 144, 77-82.
- Pepinski, W., Skawronska, M., Niemcunowicz-Janica, A., Ptaszynska-Sarosiek, I., Koc-Zorawska, E., Janica, J., and Berent, J.A. (2005). Population genetics of Y-chromosome STRs in a population sample of the Lithuanian minority residing in the northeastern Poland. *Forensic Sci Int* 153, 264-268.
- Pereira, L., Cunha, C., and Amorim, A. (2004). Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med* 118, 132-136.
- Perez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martinez-Arias, R., Clarimon, J., Fiori, G., Luiselli, D., Facchini, F., *et al.* (1999). Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *American journal of human genetics* 65, 208-219.
- Petkovski, E. (2006). Polymorphismes ponctuels de séquence et identification génétique. Etude par spectrométrie de masse MALDI-TOF. Thèse de doctorat Université Louis Pasteur, Strasbourg.

- Petrejčikova, E., Sotak, M., Bernasovska, J., Bernasovsky, I., Rebala, K., Sovicova, A., Boronova, I., Bozikova, A., Gabrikova, D., Svickova, P., *et al.* (2011). Allele frequencies and population data for 11 Y-chromosome STRs in samples from Eastern Slovakia. *Forensic Sci Int Genet* 5, e53-62.
- Pfeiffer, H., Brinkmann, B., Huhne, J., Rolf, B., Morris, A.A., Steighner, R., Holland, M.M., and Forster, P. (1999). Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112, 291-298.
- Phillips-Krawczak, C., Devor, E., Zlojutro, M., Moffat-Wilson, K., and Crawford, M.H. (2006). MtDNA variation in the Altai-Kizhi population of southern Siberia: a synthesis of genetic variation. *Hum Biol* 78, 477-494.
- Phillips, C., Salas, A., Sanchez, J.J., Fondevila, M., Gomez-Tato, A., Alvarez-Dios, J., Calaza, M., de Cal, M.C., Ballard, D., Lareu, M.V., *et al.* (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic science international Genetics* 1, 273-280.
- Piercy, R., Sullivan, K.M., Benson, N., and Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106, 85-90.
- Pilipenko, A., Romaschenko, A., Molodin, V., Parzinger, H., and Kobzev, V. (2010). Mitochondrial DNA studies of the Pazyryk people (4th to 3rd centuries BC) from northwestern Mongolia. *Archaeol Anthropol Sci* 2, 231-236.
- Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., and Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur J Hum Genet* 16, 1254-1264.
- Pliss, L., Tambets, K., Loogvali, E.L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., and Villems, R. (2006). Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet* 70, 439-458.
- Ploski, R., Wozniak, M., Pawlowski, R., Monies, D.M., Branicki, W., Kupiec, T., Kloosterman, A., Dobosz, T., Bosch, E., Nowak, M., *et al.* (2002). Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Hum Genet* 110, 592-600.
- Poinar, H.N., Hofreiter, M., Spaulding, W.G., Martin, P.S., Stankiewicz, B.A., Bland, H., Evershed, R.P., Possnert, G., and Paabo, S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 281, 402-406.
- Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R.D., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., *et al.* (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311, 392-394.

Pontes, M.L., Caine, L., Abrantes, D., Lima, G., and Pinheiro, M.F. (2007). Allele frequencies and population data for 17 Y-STR loci (AmpFlSTR Y-filer) in a Northern Portuguese population sample. *Forensic Sci Int* 170, 62-67.

Powell, G.T., Yang, H., Tyler-Smith, C., and Xue, Y. (2007). The population history of the Xibe in northern China: a comparison of autosomal, mtDNA and Y-chromosomal analyses of migration and gene flow. *Forensic Sci Int Genet* 1, 115-119.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.

Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., *et al.* (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43-49.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., and Geigl, E.M. (2007). Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* 104, 739-744.

Puzyrev, V.P., Stepanov, V.A., Golubenko, M.V., Puzyrev, K.V., Maksimova, N.R., Khar'kov, V.N., Spiridonova, M.G., and Nogovitsyna, A.N. (2003). [MtDNA and Y-chromosome lineages in the Yakut population]. *Genetika* 39, 975-981.

Q

Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., and Mehdi, S.Q. (2002). Y-chromosomal DNA variation in Pakistan. *American journal of human genetics* 70, 1107-1124.

Qin, P., Li, Z., Jin, W., Lu, D., Lou, H., Shen, J., Jin, L., Shi, Y., and Xu, S. (2014). A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. *European journal of human genetics : EJHG* 22, 248-253.

Quincey, D., Carle, G., Alunni, V., and Quatrehomme, G. (2013). Difficulties of sex determination from forensic bone degraded DNA: A comparison of three methods. *Science & justice : journal of the Forensic Science Society* 53, 253-260.

Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., *et al.* (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74, 827-845.

R

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., *et al.* (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505, 87-91.

- Rajkumar, R., and Kashyap, V.K. (2003). Haplotype diversity in mitochondrial DNA hypervariable regions I and II in three communities of Southern India. *Forensic Sci Int* 136, 79-82.
- Rakha, A., Shin, K.J., Yoon, J.A., Kim, N.Y., Siddique, M.H., Yang, I.S., Yang, W.I., and Lee, H.Y. (2011). Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int J Legal Med* 125, 841-848.
- Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P., and Chakraborty, R. (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur J Hum Genet* 9, 695-700.
- Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., *et al.* (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334, 94-98.
- Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., *et al.* (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757-762.
- Rebala, K., and Szczerkowska, Z. (2005). Polish population study on Y chromosome haplotypes defined by 18 STR loci. *Int J Legal Med* 119, 303-305.
- Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., *et al.* (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053-1060.
- Ricaut, F.X., Fedoseeva, A., Keyser-Tracqui, C., Crubezy, E., and Ludes, B. (2005). Ancient DNA analysis of human neolithic remains found in northeastern Siberia. *Am J Phys Anthropol* 126, 458-462.
- Ricaut, F.X., Keyser-Tracqui, C., Bourgeois, J., Crubezy, E., and Ludes, B. (2004a). Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. *Human biology* 76, 109-125.
- Ricaut, F.X., Keyser-Tracqui, C., Cammaert, L., Crubezy, E., and Ludes, B. (2004b). Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *American journal of physical anthropology* 123, 351-360.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., *et al.* (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *American journal of human genetics* 67, 1251-1276.
- Robino, C., Inturri, S., Gino, S., Torre, C., Di Gaetano, C., Crobu, F., Romano, V., Matullo, G., and Piazza, A. (2006). Y-chromosomal STR haplotypes in Sicily. *Forensic Sci Int* 159, 235-240.
- Robino, C., Varacalli, S., Gino, S., Chatzikyriakidou, A., Kouvatsi, A., Triantaphyllidis, C., Di Gaetano, C., Crobu, F., Matullo, G., Piazza, A., *et al.* (2004). Y-chromosomal STR

haplotypes in a population sample from continental Greece, and the islands of Crete and Chios. *Forensic Sci Int* 145, 61-64.

Rodig, H., Roewer, L., Gross, A., Richter, T., de Knijff, P., Kayser, M., and Brabetz, W. (2008). Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Sci Int* 174, 182-188.

Rodriguez, V., Tomas, C., Sanchez, J.J., Castro, J.A., Ramon, M.M., Barbaro, A., Morling, N., and Picornell, A. (2009). Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci. *Int J Legal Med* 123, 137-141.

Roewer, L., Kruger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L., Rothamel, T., Kravchenko, S., Jobling, M.A., Stoneking, M., *et al.* (2007). Y-chromosomal STR haplotypes in Kalmyk population samples. *Forensic Sci Int* 173, 204-209.

Roewer, L., Willuweit, S., Kruger, C., Nagy, M., Rychkov, S., Morozowa, I., Naumova, O., Schneider, Y., Zhukova, O., Stoneking, M., *et al.* (2008). Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. *Int J Legal Med* 122, 219-223.

Roewer, L., Willuweit, S., Stoneking, M., and Nasidze, I. (2009). A Y-STR database of Iranian and Azerbaijanian minority populations. *Forensic Sci Int Genet* 4, e53-55.

Rogaev, E.I., Grigorenko, A.P., Moliaka, Y.K., Faskhutdinova, G., Goltsov, A., Lahti, A., Hildebrandt, C., Kittler, E.L., and Morozova, I. (2009). Genomic identification in the historical case of the Nicholas II royal family. *Proceedings of the National Academy of Sciences of the United States of America* 106, 5258-5263.

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., *et al.* (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American journal of human genetics* 67, 1526-1543.

Ruiz, Y., Phillips, C., Gomez-Tato, A., Alvarez-Dios, J., Casares de Cal, M., Cruz, R., Maronas, O., Sochtig, J., Fondevila, M., Rodriguez-Cid, M.J., *et al.* (2013). Further development of forensic eye color predictive tests. *Forensic science international Genetics* 7, 28-40.

S

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350-1354.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67, 718-726.

- Sampietro, M.L., Lao, O., Caramelli, D., Lari, M., Pou, R., Marti, M., Bertranpetit, J., and Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proceedings Biological sciences / The Royal Society* 274, 2161-2167.
- Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*.
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one* 7, e34131.
- Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *Journal of applied microbiology* 113, 1014-1026.
- Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B., and Wallace, D.C. (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.
- Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., *et al.* (2000). The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159.
- Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.E., Lin, A.A., Mitra, M., Sil, S.K., Ramesh, A., *et al.* (2006). Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *American journal of human genetics* 78, 202-221.
- Seo, S.B., King, J.L., Warshauer, D.H., Davis, C.P., Ge, J., and Budowle, B. (2013). Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *International journal of legal medicine* 127, 1079-1086.
- Shapiro, B., and Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* 343, 1236573.
- Sharma, S., Rai, E., Sharma, P., Jena, M., Singh, S., Darvishi, K., Bhat, A.K., Bhanwer, A.J., Tiwari, P.K., and Bamezai, R.N. (2009). The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. *Journal of human genetics* 54, 47-55.
- Shlush, L.I., Behar, D.M., Yudkovsky, G., Templeton, A., Hadid, Y., Basis, F., Hammer, M., Itzkovitz, S., and Skorecki, K. (2008). The Druze: a population genetic refugium of the Near East. *PLoS One* 3, e2105.
- Sjodin, P., and Francois, O. (2011). Wave-of-advance models of the diffusion of the Y chromosome haplogroup R1b1b2 in Europe. *PloS one* 6, e21592.

Soltyszewski, I., Pepinski, W., Spolnicka, M., Kartasinska, E., Konarzewska, M., and Janica, J. (2007). Y-chromosomal haplotypes for the AmpFI STR Yfiler PCR Amplification Kit in a population sample from Central Poland. *Forensic Sci Int* 168, 61-67.

Spichenok, O., Budimljia, Z.M., Mitchell, A.A., Jenny, A., Kovacevic, L., Marjanovic, D., Caragine, T., Prinz, M., and Wurmbach, E. (2011). Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic science international Genetics* 5, 472-478.

Stanciu, F., Cutar, V., Pirlea, S., Stoian, V., Stoian, I.M., Sevastre, O., and Popescu, O.R. (2010). Population data for Y-chromosome haplotypes defined by 17 STRs in South-East Romania. *Leg Med (Tokyo)* 12, 259-264.

Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., *et al.* (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69, 67-89.

Stoneking, M. (1998). Women on the move. *Nature genetics* 20, 219-220.

Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G., and Montgomery, G.W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *American journal of human genetics* 82, 424-431.

Sukernik, R.I., Volodko, N.V., Mazunin, I.O., Eltsov, N.P., Dryomov, S.V., and Starikovskaya, E.B. (2012). Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *Am J Phys Anthropol* 148, 123-138.

Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., *et al.* (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature genetics* 39, 1443-1452.

Sullivan, K.M., Mannucci, A., Kimpton, C.P., and Gill, P. (1993). A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques* 15, 636-638, 640-631.

Svyatko, S.V., Mallory, J.P., Murphy, E.M., Polyakov, A.V., Reimer, P.J., and Schulting, R.J. (2009). New radiocarbon dates and a review of the chronology of prehistoric populations from the Minusinsk basin, southern Siberia, Russia. *Radiocarbon* 51, 243-273.

T

Tang, K., Allman, S.L., and Chen, C.H. (1993). Matrix-assisted laser desorption ionization of oligonucleotides with various matrices. *Rapid communications in mass spectrometry : RCM* 7, 943-948.

Tarskaia, L.A., and Melton, P. (2006). [Comparative analysis of mitochondrial DNA of Yakuts and other Asian populations]. *Genetika* 42, 1703-1711.

Taubenberger, J.K., Reid, A.H., Lourens, R.M., Wang, R., Jin, G., and Fanning, T.G. (2005). Characterization of the 1918 influenza virus polymerase genes. *Nature* 437, 889-893.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpre, M.B., Sablin, M.V., Lopez-Giraldez, F., Domingo-Roura, X., *et al.* (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342, 871-874.

Theves, C., Balaesque, P., Evdokimova, L.E., Timofeev, I.V., Alekseev, A.N., Sevin, A., Crubezy, E., and Gibert, M. (2010). Population genetics of 17 Y-chromosomal STR loci in Yakutia. *Forensic Sci Int Genet* 4, e129-130.

Theves, C., Senescau, A., Vanin, S., Keyser, C., Ricaut, F.X., Alekseev, A.N., Dabernat, H., Ludes, B., Fabre, R., and Crubezy, E. (2011). Molecular identification of bacteria by total sequence screening: determining the cause of death in ancient human subjects. *PloS one* 6, e21733.

Thomas, M.G., Barnes, I., Weale, M.E., Jones, A.L., Forster, P., Bradman, N., and Pramstaller, P.P. (2008). New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *Eur J Hum Genet* 16, 124-134.

Tran, T.N., Forestier, C.L., Drancourt, M., Raoult, D., and Aboudharam, G. (2011a). Brief communication: co-detection of *Bartonella quintana* and *Yersinia pestis* in an 11th-15th burial site in Bondy, France. *American journal of physical anthropology* 145, 489-494.

Tran, T.N., Signoli, M., Fozzati, L., Aboudharam, G., Raoult, D., and Drancourt, M. (2011b). High throughput, multiplexed pathogen detection authenticates plague waves in medieval Venice, Italy. *PloS one* 6, e16735.

Turrina, S., Atzei, R., and De Leo, D. (2006). Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *Int J Legal Med* 120, 56-59.

U

Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.E., Semino, O., Battaglia, V., *et al.* (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *European journal of human genetics : EJHG* 18, 479-484.

V

Valverde, P., Healy, E., Jackson, I., Rees, J.L., and Thody, A.J. (1995). Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nature genetics* 11, 328-330.

Van Geel, B., Bokovenko, N.A., Burova, N.D., Chugunov, K.V., Dergachev, V.A., Dirksen, V.G., Kulkova, M., Nagler, A., Parzinger, H., Van der Plicht, J., *et al.* (2004). Climate change

and the expansion of the Scythian culture after 850 BC: a hypothesis. *Journal of Archaeological Science* 31, 1735-1742.

Veeramah, K.R., and Hammer, M.F. (2014). The impact of whole-genome sequencing on the reconstruction of human population history. *Nature reviews Genetics*.

Vernot, B., and Akey, J.M. (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science*.

Visser, M., Kayser, M., and Palstra, R.J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome research* 22, 446-455.

Volgyi, A., Zalan, A., Szvetnik, E., and Pamjav, H. (2009). Hungarian population data for 11 Y-STR and 49 Y-SNP markers. *Forensic Sci Int Genet* 3, e27-28.

Volodko, N.V., Starikovskaya, E.B., Mazunin, I.O., Eltsov, N.P., Naidenko, P.V., Wallace, D.C., and Sukernik, R.I. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am J Hum Genet* 82, 1084-1100.

W

Walsh, S., Liu, F., Ballantyne, K.N., van Oven, M., Lao, O., and Kayser, M. (2011). IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic science international Genetics* 5, 170-180.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., and Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic science international Genetics* 7, 98-115.

Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., *et al.* (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10244-10249.

Wen, B., Li, H., Gao, S., Mao, X., Gao, Y., Li, F., Zhang, F., He, Y., Dong, Y., Zhang, Y., *et al.* (2005). Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22, 725-734.

Wiechmann, I., Harbeck, M., and Grupe, G. (2010). *Yersinia pestis* DNA sequences in late medieval skeletal finds, Bavaria. *Emerging infectious diseases* 16, 1806-1807.

Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., and Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791-795.

Woodward, S.R., Weyand, N.J., and Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science* 266, 1229-1232.

Wozniak, M., Derenko, M., Malyarchuk, B., Dambueva, I., Grzybowski, T., and Miscicka-Sliwka, D. (2006). Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. *Forensic science international* 164, 271-275.

Wozniak, M., Grzybowski, T., Starzynski, J., and Marciniak, T. (2007). Continuity of Y chromosome haplotypes in the population of Southern Poland before and after the Second World War. *Forensic Sci Int Genet* 1, 134-140.

Y

Young, K.L., Sun, G., Deka, R., and Crawford, M.H. (2011). Paternal genetic history of the Basque population of Spain. *Human biology* 83, 455-475.

Z

Zalloua, P.A., Platt, D.E., El Sibai, M., Khalife, J., Makhoul, N., Haber, M., Xue, Y., Izaabel, H., Bosch, E., Adams, S.M., *et al.* (2008a). Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *American journal of human genetics* 83, 633-642.

Zalloua, P.A., Xue, Y., Khalife, J., Makhoul, N., Debiane, L., Platt, D.E., Royyuru, A.K., Herrera, R.J., Hernanz, D.F., Blue-Smith, J., *et al.* (2008b). Y-chromosomal diversity in Lebanon is structured by recent historical events. *American journal of human genetics* 82, 873-882.

Zastera, J., Roewer, L., Willuweit, S., Sekerka, P., Benesova, L., and Minarik, M. (2010). Assembly of a large Y-STR haplotype database for the Czech population and investigation of its substructure. *Forensic Sci Int Genet* 4, e75-78.

Zegura, S.L., Karafet, T.M., Zhivotovsky, L.A., and Hammer, M.F. (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Molecular biology and evolution* 21, 164-175.

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R.S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., *et al.* (2003). The genetic legacy of the Mongols. *American journal of human genetics* 72, 717-721.

Zhang, F., Xu, Z., Tan, J., Sun, Y., Xu, B., Li, S., Zhao, X., Zhou, H., Gong, G., Zhang, J., *et al.* (2010). Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am J Phys Anthropol* 142, 314-320.

Zhang, Q., Yan, J., Tang, H., Jiao, Z., and Liu, Y. (2006). Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. *Legal medicine* 8, 300-305.

Zhao, Y.B., Li, H.J., Cai, D.W., Li, C.X., Zhang, Q.C., Zhu, H., and Zhou, H. (2010). Ancient DNA from nomads in 2500-year-old archeological sites of Pengyang, China. *J Hum Genet* 55, 215-218.

Zhong, H., Shi, H., Qi, X.B., Duan, Z.Y., Tan, P.P., Jin, L., Su, B., and Ma, R.Z. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Molecular biology and evolution* 28, 717-727.

Zhu, B.F., Wang, Z.Y., Yang, C.H., Li, X.S., Zhu, J., Yang, G., Huang, P., and Liu, Y. (2005). Y-chromosomal STR haplotypes in Chinese Uigur ethnic group. *Int J Legal Med* 119, 306-309.

VII ANNEXES

Annexe 1 : Amorces utilisées pour l'analyse des SNP de l'ADN mitochondrial et du chromosome Y

Marqueurs	Amorces PCR		Taille amplicon	Amorce SBE
	P1	P2		
Z_9090	ACGTTGGATGGCGATTTCTAGGATAGTCAG	ACGTTGGATGCCACCTTAGCAATATCAACC	105	ACCATTAACCTTCCTC
U_12308	ACGTTGGATGAGCTATCCATTGGTCTTAGG	ACGTTGGATGGGGTGGTTATAGTAGTGTGC	103	GGAGTTGCACCAAAATT
C4a1_1715	ACGTTGGATGTTGCGCCAGGTTTCAATTTTC	ACGTTGGATGAACCCACTCCACCTTACTAC	107	ACCTTACTACCAGACAAC
D5a2_1438	ACGTTGGATGCCAGAAAACACGATAGCCC	ACGTTGGATGCCCTGTTCAACTAAGCACTC	108	CAACTAAGCACTCTACTCT
D_5178	ACGTTGGATGTTAAACTCCAGCACACGAC	ACGTTGGATGGGTGGATGGAATTAAGGGTG	101	AAGGGTGTAGTCATGTTA
HV_14766	ACGTTGGATGGGGAGGTCGATGAATGAGTG	ACGTTGGATGACAAGAACACCAATGACCCC	97	AATGACCCCAATACGCAAAA
D5a_752	ACGTTGGATGTGGCTAGGCTAAGCGTTTTC	ACGTTGGATGTCACCTCTAAATCACCAACG	99	ATCACCAAGTCAAAAAGGAA
H_7028	ACGTTGGATGGACATCGTACTACACGACAC	ACGTTGGATGATGATGGCAAAATACAGCTCC	100	tagAGGACATAGTGGAAAGTG
A_663	ACGTTGGATGGCTCACATCACCCCATAAAC	ACGTTGGATGGGATGCTTGATGTGTAATC	100	cCTAAGAGCTAATAGAAAGGC
D4j5_13086	ACGTTGGATGGGAAGCGGATGAGTAAGAAG	ACGTTGGATGACTCCCTCAGCCATGAGAAG	115	GTCTCAGCCCTACTCCACTCAAG
C_13263	ACGTTGGATGATCGTAGCCTTCTCCACTTC	ACGTTGGATGAGGAATGCTAGGTGTGGTTG	100	ATTGTAACATTATGAGTCCCTAG
LCT_C13910T	ACGTTGGATGCAACCTAAGGAGGAGAGTTC	ACGTTGGATGATGTACTAGTAGGCCTCTGC	102	GGCAATACAGATAAGATAATGTAG
J_13708	ACGTTGGATGCGAAAATAACCCACCCCTAC	ACGTTGGATGAAAATCCTGCGAATAGGCTTC	87	AATCCTGCGAATAGGCTTCCGGCTG
T_4917	ACGTTGGATGTTGAGAGATGAGGAGAAGG	ACGTTGGATGCATGACAAAACTAGCCCCC	101	TCATATACCAAATCTCTCCCTCACTA
M458	ACGTTGGATGGAAAGTGACTTACAGCTCAGG	ACGTTGGATGGCTTTGAAAGACATTCCTCC	95	ATTCTCTGGCTCTCT
R1a1a_M17	ACGTTGGATGTTGGCCACTTAACAAACCCC	ACGTTGGATGTCAGATTCTGTCTACTCACC	107	TTGCTGGTTGTTACGGG
R1a_M420	ACGTTGGATGCTTGCCCTCCAATATGGGAC	ACGTTGGATGGGCAAACCTTTTCATTGCTGG	113	TTTTCAATTGCTGGCTCCA
R1b1a_P297	ACGTTGGATGGGAAAGGAACATAGCCAAAAG	ACGTTGGATGGGTGCTGCACTCAATCTCT	120	TTTTTAGGGTCAGGAATCT
R1b1a2_M269	ACGTTGGATGTCCTTTGTGTGCTTCTGAG	ACGTTGGATGAGGGGAATGATCAGGGTTTG	106	ATGATCAGGGTTTGGTTAAT
M434	ACGTTGGATGAAAATTAGCTGGGCGTGGTG	ACGTTGGATGAGGTTCAGCGATTCTCCTG	98	CCTGCATCAGCCTCCCAGGTA
R1a1_M516	ACGTTGGATGGGTAATTTGGCCAAAGCACAC	ACGTTGGATGTCAGCAGCAATAAGCTGAGG	117	cAGTAGACTGTTTACACCTAG
Z280	ACGTTGGATGTCAAAGGTCTTTACTTGTGC	ACGTTGGATGCGTTAGATGGCAGCTTTTAG	105	GATGGCAGCTTTTAGTTAGATC
R1b_M343	ACGTTGGATGTCCTGTCTTTCCCAACAC	ACGTTGGATGTATGCAAAATGCAAGATGCC	111	atgccGATGCCCTCGTGTCCA
R_M207	ACGTTGGATGGGGCAAATGTAAAGTCAAGC	ACGTTGGATGTCACCTTCAACCTCTGTGG	102	TCAAAAAGGTATTGTTATTCTCTTT
R1_M173	ACGTTGGATGTTTCTTACAATTCAAGGGC	ACGTTGGATGCCTGAAAAACAAACACTGGC	101	TCTGAATATTAACAGATGACAAAG
R1b1a1_M73	ACGTTGGATGCCAAAACTTCTGAATTAC	ACGTTGGATGGGTCTTTGTGATTCTCTG	109	TTGTGATTCTCTGAACGCTAACCA
Z93	ACGTTGGATGCCCCAAATGAGTTAAAGAC	ACGTTGGATGATGCATGTCAGGGATTGAGG	118	ACAATTATTATTAAGGGTTTGAAGC
N_M231	ACGTTGGATGCAACATTTACTGTTTCTAC	ACGTTGGATGCGTGTGTGTTGACCAACAG	108	CGATCTTTCCCCAATT
N1c2b_P43	ACGTTGGATGCCTGTAGTCCCAGCTACTTG	ACGTTGGATGTCCTTGCTCACTGCAAGCTC	99	TCCGCCTCCCGGGTTCA
C3_M217	ACGTTGGATGAAGCTGCTGTGGCTTTTCATC	ACGTTGGATGAGGAGAATGAAAAAGTTGGG	103	AAAAAGTTGGGTGACAC
Q1a3_M346	ACGTTGGATGTCACTCTGCCTACCTGATAG	ACGTTGGATGAGATGGGAAAGGCAGCCAAG	98	GCCAAGAGGACAGTAAGA
O_M175	ACGTTGGATGCTACTGTACCTTTGTTTCTG	ACGTTGGATGTGAATCAGGCATGCTTTC	84	ACATGCCTTCTCACTTCTC
D_M174	ACGTTGGATGGTTTTGCAAAATGCACCCCTC	ACGTTGGATGTCGCTTCTCTGAAATACCTTC	82	AATACCTTCTGGAGTGCC
Q1a3a_L54	ACGTTGGATGCCTGTAAAGGTAAACATGGAG	ACGTTGGATGAGGTCTCATCTAGTCTTC	99	TGTTGTACTCTCATAGCTC
NO_M214	ACGTTGGATGAAATGCCACTTCACTCCAGC	ACGTTGGATGCTACTGGTTACTTTCGTTG	106	ACTTTCGTTGTTTTATTTTTC
Q1a1_M120	ACGTTGGATGCGCAATAAAGTATAATTTCCC	ACGTTGGATGAACACACTGCTAATGATCCG	94	CCGTTTTTTGATGTGGAAATA
N1c2a_M128	ACGTTGGATGGTTGCAGAAATATCTACCTC	ACGTTGGATGCAGTGTTTTATCAGTACTGC	100	CAGTACTGCATCCTTTTTTTTCA
Q1a3a1_M3	ACGTTGGATGCATTTTAGGTACCAGCTCTTC	ACGTTGGATGTGCCAGGCTTTCAAATAGG	100	CAATGGGTACCTCTGGGACTGA
Q_M242	ACGTTGGATGGGGGAAAAACCATGTACG	ACGTTGGATGAGTTTGTGCAAAAAGGTGAC	113	GTGCAAAAAGGTGACCAAGGTGCT
N1c1_Tat	ACGTTGGATGGGACTCTGAGTGTAGACTTG	ACGTTGGATGTACCCTTCTCTTTGCTGTG	94	GCTCTGAAATATTAAATTAACAAC
C_M130	ACGTTGGATGGCAACAGTAAGTCGAATGCC	ACGTTGGATGTCCTCTTCTATTGAGGG	94	ATTGCAAGGCAATAAACCTTGGATTTC

Annexe 2 : Amorces utilisées pour l'analyse des SNP autosomaux

Amorces sequenom des 22 marqueurs (SNP) sélectionnés pour la détermination des traits phénotypiques

Marqueurs	Amorces PCR (3'→5')		Taille amplicon (pb)	Amorces SBE
	P1	P2		
rs1805006	ACGTTGGATGTCCAGCAGGAGGATGACG	ACGTTGGATGCTGGTGAGCTTGGTGGAGA	178	CCTGGCCTTGTCCGA
rs16891982	ACGTTGGATGTCCAAGTTGTGCTAGACCAGA	ACGTTGGATGCGAAAGAGGAGTCGAGGTTG	148	TGGATGTTGGGGCTT
rs1800407	ACGTTGGATGAAGGCTGCCTCTGTTCTAC	ACGTTGGATGCGATGAGACAGAGCATGATGA	144	aGCCCAACCCGTCCT
rs4959270	ACGTTGGATGTGAGAAATCTACCCCCACGA	ACGTTGGATGGTGTCTTACCCCCTGTG	160	tCCCCCTGTGGAAGTG
rs12913832	ACGTTGGATGTCAACATCAGGGTAAAAATCATGT	ACGTTGGATGGGCCCTGATGATGATAGC	170	CGTGCAGAACTTGACA
rs180479	ACGTTGGATGTCCAGCAGGAGGATGACG	ACGTTGGATGCTGGTGAGCTTGGTGGAGA	178	GGTGAGCGGGAGCAAC
N29insA	ACGTTGGATGGCAGGGATCCCAGAGAAAGAC	ACGTTGGATGTCAGAGATGGACACCTCCAG	137	gACCGGGCTCCTGTCTG
rs1426654	ACGTTGGATGTTCAAGCCCTTGGATTGTCTC	ACGTTGGATGGCAGTAACTAATTCAGGAGC	101	aGCTGCCATGAAAGTTG
rs12203592	ACGTTGGATGGCTTCGTCATATGGCTAAACCT	ACGTTGGATGAGGGCAGCTGATCTCTTCAGGCTT	146	TGGTGGGTAAAAGAAGG
rs1042602	ACGTTGGATGCAACACCCATGTTTAAACGACA	ACGTTGGATGGCTTCATGGGC AAAATCAAT	144	ATGTCTCTCCAGATTTCA
rs3827760	ACGTTGGATGAATCTCATCCCTCTTCAGGC	ACGTTGGATGCTCAGCTCCACGTACAATCTC	107	caAACTCTGAGAAGGCTG
rs2402130	ACGTTGGATGACCTGTCTCACAGTGCTGCT	ACGTTGGATGTTACCTCGATGACGATGAT	170	TGATGATGATGGCAGCTA
rs2378249	ACGTTGGATGCGCATAACCCATCCCTCTAA	ACGTTGGATGCATTGCTTTTCAGCCACAC	156	aaacCCTCTCCTCAGCCCA
rs28777	ACGTTGGATGTCTTTGATGTCCCCTTCGAT	ACGTTGGATGTACTCGTGTGGGAGTTCCAT	170	aATGTGATCCTCACAGCAG
rs1805005	ACGTTGGATGTCCAGCAGGAGGATGACG	ACGTTGGATGCTGGTGAGCTTGGTGGAGA	178	gggtGAGAACGCGCTGGTG
rs1805009	ACGTTGGATGCAAGAACTTCAACCTCTTTCTCG	ACGTTGGATGCACCTCCTTGAGCGTCCTG	126	cagtCGTAGATGAGGGGT
rs1393350	ACGTTGGATGTTCTTTATCCCCCTGATGC	ACGTTGGATGGGGAAGGTGAATGATAACACG	144	GTAAAAGACCACACAGATTT
rs2814778	ACGTTGGATGATGGCAACGTTTGGTTACAGG	ACGTTGGATGATGGCCCTCATTAAGTCCTTG	116	CTCATTAAGTCCTTGGCTCTTA
rs12896399	ACGTTGGATGGACCTGTGTGAGACCCAGT	ACGTTGGATGCTGGCGATCCAATTCTTTGT	145	TTTAGGTCAGTATATTTTGGG
rs12821256	ACGTTGGATGGGAGCCAAAGGGCATGTTACTA	ACGTTGGATGATGCCCAAAGGATAAGGAAT	138	GTTTAGGGTTTTTTTCCTTTAG
rs683	ACGTTGGATGCACAAAACCACTGGTTGAA	ACGTTGGATGTGAAAAGGTCCTCCAGCTT	158	AAAAGTATGCCTAGAACTTTAAT
rs1545397	ACGTTGGATGCTGACAATGGTTGTACAAC	ACGTTGGATGGGAGATATAGAATTCACAC	116	ATCTTGCAAAATTATATCATTCAG

Annexe 3 : Description et références des populations modernes utilisées pour les comparaisons des haplotypes mitochondriaux

Populations ou pays	Nombre de séquences	Références
Asie Centrale		
Kazakhes	55	(Comas et al., 1998)
Kirghizes	94	(Comas et al., 1998)
Uigurs	55	(Comas et al., 1998)
Total	149	
Asie du Sud		
Inde du Centre-Est	161	(Metspalu et al., 2004)
Inde du Centre-Ouest	166	(Metspalu et al., 2004)
Inde du Nord	239	(Metspalu et al., 2004)
Inde du Sud	187	(Metspalu et al., 2004; Rajkumar and Kashyap, 2003)
Pathans	230	(Rakha et al., 2011)
Sri-Lankais	132	(Metspalu et al., 2004)
Total	788	
Asie du Sud-Est		
Laotiens	214	(Bodner et al., 2011)
Vietnamiens	187	(Irwin et al., 2008)
Total	401	
Asie de l'Est		
Japonais	162	(Imaizumi et al., 2002)
Coréens	185	(Jin et al., 2006)
Total	347	
Chine		
Daur	80	(Kong et al., 2003; Powell et al., 2007)
Ewenki	70	(Kong et al., 2003; Powell et al., 2007)
Gelao	102	(Liu et al., 2011)
Hainan	285	(Peng et al., 2011)
Han G	106	(Chen et al., 2008)
Han Harbin	29	(Powell et al., 2007)
Han Inner Mongolia	17	(Powell et al., 2007)
Han Xinjiang	29	(Powell et al., 2007)
Hmong	167	(Wen et al., 2005)
Hezhen	43	(Powell et al., 2007)
Inner Mongolian	39	(Powell et al., 2007)
Manchu	33	(Powell et al., 2007)
Mien	370	(Wen et al., 2005)
Oroqen	74	(Kong et al., 2003; Powell et al., 2007)
Daheyuan population	58	(Cui et al., 2010)
Uigurs	63	(Powell et al., 2007)
Xibe	38	(Powell et al., 2007)
Total	1603	
Mongolie		
Mongols	193	(Gokcumen et al., 2008; Kolman et al., 1996)
Total	193	

Sibérie de l'Ouest		
Kets	38	(Derbeneva et al., 2002a)
Khantys	106	(Pimenoff et al., 2008)
Nganasans	63	(Derbeneva et al., 2002a; Volodko et al., 2008)
Tatars	197	(Malyarchuk et al., 2010)
Total	404	
Sibérie du Sud		
Altaians	110	(Derenko et al., 2003)
Altaiian Kazakhs	237	(Gokcumen et al., 2008)
Altaiian Kizhis	337	(Dulik et al., 2012; Phillips-Krawczak et al., 2006)
Chelkans	91	(Dulik et al., 2012)
Khakassians	53	(Derenko et al., 2003)
Koumandins	52	(Dulik et al., 2012)
Sojots	30	(Derenko et al., 2003)
Todjins	48	(Derenko et al., 2003)
Tofalars	104	(Derenko et al., 2003; Starikovskaya et al., 2005)
Tubalars	287	(Dulik et al., 2012; Starikovskaya et al., 2005; Sukernik et al., 2012)
Tuvinians	245	(Derenko et al., 2003; Pakendorf et al., 2006; Starikovskaya et al., 2005)
Total	1594	
Sibérie de l'Est		
Aleuts	36	(Volodko et al., 2008)
Bouriates	116	(Derenko et al., 2003; Starikovskaya et al., 2005)
Chukchis	182	(Volodko et al., 2008)
Eskimos	208	(Saillard et al., 2000; Volodko et al., 2008)
Evenks	110	(Pakendorf et al., 2006; Starikovskaya et al., 2005)
Evens	87	(Sukernik et al., 2012)
Itelmen	46	(Schurr et al., 1999)
Koryaks	147	(Schurr et al., 1999)
Mansis	161	(Derbeneva et al., 2002b; Pimenoff et al., 2008)
Negidals	33	(Starikovskaya et al., 2005)
Nivkhis	56	(Starikovskaya et al., 2005)
Udegeys	46	(Starikovskaya et al., 2005)
Ulchis	247	(Starikovskaya et al., 2005; Sukernik et al., 2012)
Yakoutes	592	(Fedorova et al., 2003; Pakendorf et al., 2006; Puzyrev et al., 2003; Tarskaia and Melton, 2006)
Yakoutes parlant Evenk	32	(Pakendorf et al., 2006)
Yukaghirs	154	(Pakendorf et al., 2006; Volodko et al., 2008)
Total	2253	
Moyen-Orient		
Iran	436	(Metspalu et al., 2004)

Arabie Saoudite	553	(Abu-Amero et al., 2008)
Syrie	69	(Richards et al., 2000)
Iraq	116	(Richards et al., 2000)
Druze	356	(Macaulay et al., 1999; Shlush et al., 2008)
Kurdes	82	(Comas et al., 2000; Richards et al., 2000)
Dubaï	249	(Alshamali et al., 2008)
Palestine	117	(Richards et al., 2000)
Turques	340	(Calafell et al., 1996; Comas et al., 1996; Di Benedetto et al., 2001; Quintana-Murci et al., 2004; Richards et al., 2000)
Total	2318	
Caucase		
Russie de l'Ouest	358	(Grzybowski et al., 2007; Malyarchuk et al., 2002)
Populations du nord du Caucase	246	(Macaulay et al., 1999; Quintana-Murci et al., 2004; Richards et al., 2000)
Géorgie	64	(Comas et al., 2000; Quintana-Murci et al., 2004)
Arménie	191	(Richards et al., 2000)
Azerbaïdjan	48	(Richards et al., 2000)
Total	907	
Europe de l'Ouest		
Grande Bretagne	100	(Piercy et al., 1993)
France	172	(Dubut et al., 2004)
Irlande	300	(McEvoy et al., 2004)
Total	572	
Nord-centre de l'Europe		
Allemagne	409	(Brandstatter et al., 2006; Lutz et al., 1998; Pfeiffer et al., 1999)
République Tchèque	179	(Malyarchuk et al., 2006)
Pologne	849	(Grzybowski et al., 2007; Malyarchuk et al., 2002)
Slovaquie	207	(Malyarchuk et al., 2008)
Suisse	154	(Dimo-Simonin et al., 2000)
Lettonie	299	(Pliss et al., 2006)
Autriche	101	(Parson et al., 1998)
Total	2198	
Scandinavie		
Norvège	74	(Passarino et al., 2002)
Finlande	603	(Hedman et al., 2007; Meiniola et al., 2001)
Total	677	
Sud-est de l'Europe		
Bulgarie	30	(Calafell et al., 1996)
Bosnie	144	(Malyarchuk et al., 2003)
Roumanie	105	(Bosch et al., 2006)
Total	279	
Ouest de la méditerranée		
Portugal	847	(Gonzalez et al., 2003; Pereira et al., 2004)
Sud de la France	37	(Dubut et al., 2004)
Galice	43	(Gonzalez et al., 2003)
Total	927	

Méditerranée Centrale		
Italie	263	(Thomas et al., 2008)
Slovénie	104	(Malyarchuk et al., 2003)
Total	367	
Est de la méditerranée		
Macédoine	37	(Bosch et al., 2006)
Albanie	42	(Bosch et al., 2006)
Grèce	25	(Bosch et al., 2006)
Total	104	
Total général	16081	

Annexe 4 : Description et références des individus anciens utilisés pour les comparaisons des haplotypes mitochondriaux et Y

Référence	Description (origine, culture)	Période	ADN mitochondrial		Chromosome Y	
			Taille des séquences	N	SNP/STR	N
(Bramanti et al., 2009)	Allemagne, Lituanie, Pologne, Russie, chasseurs cueilleurs	13400-2300 BC	15997-16409	20	-	
(Haak et al., 2010; Haak et al., 2005)	Allemagne, LBK culture	Néolithique (5500-4900 BC)	15997-16409	42	SNP	3
(Der Sarkissian et al., 2013)	Nord-ouest de la Russie	7500-7000 BP	16056-16409	11	-	
(Sampietro et al., 2007)	Espagne	Néolithique (3500-3000 BC)	16022-16378	11	-	
(Malmstrom et al., 2009)	Suède, PWC	Néolithique (5500-4500 BP)	16051-16383	19	-	
	Suède, FBC	Néolithique (4800-4000 BP)	16051-16383	3	-	
(Lacan et al., 2011a)	France, culture des Treilles	Fin du Néolithique (3000 BC)	16024-16383	29	STR et SNP	22
(Haak et al., 2008)	Allemagne, culture Cordée	Fin du Néolithique (2700-2400BC)	15997-16409	9	STR et SNP	3
(Chikisheva et al., 2007)	Altaï, Russie	Néolithique (4 ^{ème} mill BC)	16024-16401	1	-	
(Mooder et al., 2006)	lac Baïkal, Russie	Néolithique (6125–2020 BC)	16191-16367	26	-	
(Ricaud et al., 2005)	Yakoutie, Sibérie	Néolithique (3600 BP)	16015-16391	1	-	
(Lee et al., 2012)	Allemagne, culture Bell Beaker	Fin du Néolithique	15995 -16429	6	SNP	2
(Gonzalez-Ruiz et al., 2012)	Altaï, Mongolie	âge du Bronze	16051-16400	3	-	
(Keyser et al., 2009)	Sud de la Sibérie, cultures d'Andronovo et de Karasuk	âge du Bronze (1800-800 BC)	16009-16390	11	STR et SNP	3
(Lalueza-Fox et al., 2004)	Kazakhstan	âge du Bronze (3400-1000 BC)	-	8	-	
(Molodin et al., 2012)	Sibérie de l'Ouest, cultures d'Ust-Tartas, d'Odinovo, de Krotovo, d'Andronovo, de Baraba, d'Irmen	âge du Bronze (4 ^{ème} mill -8 ^{ème} siècle BC)		92	-	
(Chikisheva et al., 2007)	Altaï, Russie	âge du Bronze (milieu 2 ^{ème} mill BC)	16024-16401	3	-	
(Li et al., 2010)	Chine, Bassin du Tarim	âge du Bronze (3900 BP)	16035-16409	20	SNP	7
(Zhang et al., 2010)	Chine, Xinjiang, site de Gavaerk	2800-2100 BP	16053-16366	35	-	
(Li et al., 2011)	Nord-est de la Chine	âge du Bronze (3600 BP)	16035-16409	14	SNP	5
(Gonzalez-Ruiz et al., 2012)	Altaï, Mongolie, culture de Pazyryk	âge du Fer	16051-16400	16	-	
(Keyser et al., 2009)	Sud de la Sibérie, cultures de Tagar et de Tachtyk	âge du Fer (800 BC-400 AD)	16009-16390	15	STR et SNP	7
(Chikisheva et al., 2007)	Altaï, Russie, culture de Pazyryk	âge du Fer (4 ^{ème} -3 ^{ème} siècle BC)	16024-16401	7	-	
(Der Sarkissian et al., 2013)	Nord-ouest de la Russie	3500 BP	16056-16409	23	-	
(Lalueza-Fox et al., 2004)	Kazakhstan	âge du Fer (800-100 BC)	-	18	-	

ANNEXES

Références	Description (origine, culture)	Période	ADN mitochondrial		Chromosome Y	
			Taille des séquences	N	SNP/STR	N
(Clisson et al., 2002)	Kazakhstan, site de Berel, Scythes	âge du Fer (3 ^{ème} siècle BC)	16021-16417	2	-	
(Ricaud et al., 2004b)	Altaï, Russie, Scythes	âge du Fer (5 ^{ème} siècle BC)	16018-16383	2	-	
(Ricaud et al., 2004a)	Altaï, Russie, Scythes	âge du Fer (4 ^{ème} – 2 ^{ème} siècle BC)	16047-16385	1	-	
(Bennett and Kaestle, 2010)	Sibérie de l'Ouest, culture de Sargat	âge du Fer (5 ^{ème} siècle BC- 5 ^{ème} siècle AD)	16038-16356	4	-	
(Gao et al., 2008)	Chine, désert du Taklamakan	2000-2500 BP	16035-16378	16	-	
(Keyser-Tracqui et al., 2003a)	Mongolie, nécropole d'Egyin Gol, Xiongnu	3 ^{ème} siècle BC- 2 ^{ème} siècle AD	16009-16390	46	STR	27
(Kim et al., 2010)	Nord-est de la Mongolie, Xiongnu	âge du Fer (300 BC-100 AD)	15997-16380	3	SNP	2
(Zhao et al., 2010)	Chine, Mongolie Intérieure	770 BC-56 AD	16050-16409	6	SNP	4

BP : Before Present (avant aujourd'hui (1950)) ; AD : Anno Domini (après Jésus-Christ) ; BC : Before Christ (avant Jésus-Christ); mill : millénaire

Annexe 5 : Description et références des populations modernes utilisées pour les comparaisons des haplotypes Y

Populations ou pays	Nombre de séquences	Références
Asie Centrale		
Kazakhes	49	(Perez-Lezaun et al., 1999)
Kirghizes	84	(Perez-Lezaun et al., 1999)
Uigurs	157	(Perez-Lezaun et al., 1999; Zhong et al., 2011)
Afghans	518	(Achakzai et al., 2012; Haber et al., 2012; Lacau et al., 2011)
Total	808	
Asie du Sud		
Indiens	998	(Frank et al., 2008; Nair et al., 2011; Ramana et al., 2001; Sengupta et al., 2006)
Pathans	718	(Qamar et al., 2002)
Bangladais	216	(Alam et al., 2010)
Pakistanaï	177	(Sengupta et al., 2006)
Tibétains	470	(Gayden et al., 2011; Zhang et al., 2006; Zhong et al., 2011)
Total	2579	
Asie du Sud-Est		
Cambodgiens	6	(Sengupta et al., 2006)
Malaisiens	334	(Chang et al., 2007)
Mon-Khmer	790	(Cai et al., 2011)
Taiwanais	200	(Huang et al., 2008)
Total	1330	
Asie de l'Est		
Japonais	1189	(Hashiyada et al., 2008; Sengupta et al., 2006)
Coréens	526	(Kim et al., 2008)
Total	1715	
Chine		
Chinois	939	(Cai et al., 2011; Li et al., 2004; Sengupta et al., 2006; Zhong et al., 2011)
Uigurs	107	(Zhu et al., 2005)
Total	1046	
Mongolie		
Mongols	139	(Kayser et al., 2001; Kwak et al., 2006; Malyarchuk et al., 2011; Zhong et al., 2011)
Total	139	
Sibérie de l'Ouest		
Komis	99	(Mirabal et al., 2009)
Khantys	25	(Mirabal et al., 2009)
Kalmyks	145	(Malyarchuk et al., 2011; Nasidze et al., 2005; Roewer et al., 2007)
Total	269	
Sibérie du Sud		
Altaiian Kazakhs	119	(Dulik et al., 2011)
Tuviniens	63	(Derenko et al., 2006; Malyarchuk et al., 2011; Pakendorf et al., 2006)

Todjins	11	(Derenko et al., 2006; Malyarchuk et al., 2011)
Shors	28	(Derenko et al., 2006; Malyarchuk et al., 2011)
Khakassians	275	(Derenko et al., 2006; Khar'kov et al., 2011; Malyarchuk et al., 2011)
Sojots	9	(Derenko et al., 2006; Malyarchuk et al., 2011)
Teleuts	59	(Derenko et al., 2006; Khar'kov et al., 2009; Malyarchuk et al., 2011)
Altaian Kizhis	155	(Derenko et al., 2006; Dulik et al., 2012)
Tofalars	5	(Derenko et al., 2006; Malyarchuk et al., 2011)
Chelkans	25	(Dulik et al., 2012)
Koumandins	17	(Dulik et al., 2012)
Tubalars	27	(Dulik et al., 2012)
Altaians	24	(Malyarchuk et al., 2011)
Total	817	
Sibérie de l'Est		
Evenks	17	(Derenko et al., 2006; Malyarchuk et al., 2011; Pakendorf et al., 2007)
Bouriates	225	(Derenko et al., 2006; Malyarchuk et al., 2011; Wozniak et al., 2006)
Yakoutes	394	(Pakendorf et al., 2006; Puzyrev et al., 2003; Sengupta et al., 2006; Theves et al., 2010)
Russes de l'Est	152	(Lessig et al., 2006)
Evens	48	(Malyarchuk et al., 2011; Pakendorf et al., 2007)
Koryaks	4	(Malyarchuk et al., 2011)
Yukaghirs	4	(Pakendorf et al., 2006)
Total	844	
Moyen-Orient		
Iran	340	(Nasidze et al., 2003; Roewer et al., 2009)
Liban	587	(Zalloua et al., 2008b)
Syrie	161	(Zalloua et al., 2008a)
Palestine	364	(Zalloua et al., 2008a)
Druze	283	(Shlush et al., 2008)
Turques	815	(Alakoc et al., 2010; Cakir et al., 2004; Cinnioglu et al., 2004; Nasidze et al., 2003)
Total	2550	
Caucase		
Azéris	119	(Nasidze et al., 2003; Roewer et al., 2009)
Arménie	100	(Nasidze et al., 2003)
Géorgie	77	(Nasidze et al., 2003)
Autres populations du Nord du Caucase	253	(Nasidze et al., 2003)
Total	549	
Europe de l'Ouest		
Irlande	155	(Ballard et al., 2006)
France	100	(Keyser-Tracqui et al., 2003b)
Belgique	113	(De Maesschalck et al., 2005)
Pays-Bas	72	(Rodig et al., 2008)
Angleterre	250	(Ballard et al., 2005)
Total	690	

Nord-centre de l'Europe		
Pologne	1794	(Pepinski et al., 2004; Ploski et al., 2002; Rebala and Szczerkowska, 2005; Soltyszewski et al., 2007; Wozniak et al., 2007)
Allemagne	1741	(Hohoff et al., 2007; Ploski et al., 2002; Rodig et al., 2008)
Autriche	135	(Berger et al., 2005)
Danemark	185	(Hallenberg et al., 2005)
République Tchèque	1744	(Zastera et al., 2010)
Slovaquie	629	(Petrejckikova et al., 2011)
Suisse	150	(Haas et al., 2006)
Nord de l'Italie	155	(Turrina et al., 2006)
Estonie	251	(Lappalainen et al., 2008; Ploski et al., 2002)
Lettonie	258	(Lappalainen et al., 2008; Ploski et al., 2002)
Lituanie	440	(Lappalainen et al., 2008; Pepinski et al., 2005; Ploski et al., 2002)
Ouest de la Russie	814	(Derenko et al., 2006; Mirabal et al., 2009; Ploski et al., 2002; Roewer et al., 2008)
Total	8296	
Scandinavie		
Norvège	1766	(Dupuy et al., 2006)
Suède	868	(Holmlund et al., 2006; Lappalainen et al., 2008)
Finlande	536	(Lappalainen et al., 2008)
Total	3170	
Sud-est de l'Europe		
Hongrie	215	(Volgyi et al., 2009)
Roumanie	297	(Egyed et al., 2006; Stanciu et al., 2010)
Bosnie-Herzégovine	181	(Klaric et al., 2005)
Monténégro	404	(Mirabal et al., 2010)
Serbie	179	(Mirabal et al., 2010)
Total	1276	
Ouest de la méditerranée		
Portugal	313	(Adams et al., 2008; Pontes et al., 2007)
Espagne	1374	(Adams et al., 2008; Rodriguez et al., 2009; Young et al., 2011)
Total	1687	
Méditerranée Centrale		
Italie du Sud	899	(Di Gaetano et al., 2009; Ghiani et al., 2009; Robino et al., 2006; Rodriguez et al., 2009)
Croatie	200	(Haliti et al., 2009)
Malte	50	(Zalloua et al., 2008a)
Total	1149	
Est de la méditerranée		
Grèce	557	(King et al., 2011; King et al., 2008; Kovatsi et al., 2009; Martinez et al., 2007; Robino et al., 2004)
Macédoine	28	(Robino et al., 2004)
Chypre	163	(Zalloua et al., 2008a)
Total	748	
Total général	29662	

Annexe 6 : Génotypes obtenus et résultats de la prédiction de la couleur des cheveux et des yeux pour les individus modernes

Echantillons																									Valeurs obtenues avec l'outil de prédiction									phénotype prédit (couleur la plus probable)	phénotype observé																		
																									couleur des cheveux				nuance		couleur des yeux																						
	C/A	N29	rs11547464	rs885479	rs1805008	rs1805005	rs1805006	rs1805007	rs1805009*	C/A	Y1520CH	A/G	rs2228479	T/C	rs1110400	C/A	rs28777	G/C	rs16891982	C/T	rs12821256	C/A	rs4959270	C/T	rs12203592	C/A	rs1042602	C/T	rs1800407	G/A	rs2402130	G/A	rs12913832	G/A	rs2378249	G/T	rs12896399	G/A	rs1393350	C/A	rs683	Brun	Roux	Noir	Blond	Clair	Foncé	Bleu	Int.	Brun	cheveux	yeux	cheveux
M1	C	G	G	C/T	G	C	C	-	C	G	T	CA	GC	T	A	C	C	C	G	G	A	G	G	CA	0,255	0,055	0,146	0,545	0,779	0,221	0,678	0,131	0,191	brun clair - brun foncé	bleu	brun	bleu																
M2	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	C	C	CA	C	GA	GA	A	G	G	C	0,378	0,002	0,338	0,282	0,43	0,57	0,097	0,148	0,758	brun clair - brun foncé	brun	gris	brun																
M3	C	G	G	C	G	C	C	-	C	G	T	A	G	CT	CA	C	CA	C	A	A	A	GT	GA	CA	0,32	0	0,413	0,267	0,368	0,632	0,002	0,034	0,964	brun clair - brun foncé	brun	brun	brun																
M4	C	G	G	C	G	C	C	-	C	G	T	CA	GC	CT	CA	CT	CA	C	A	GA	A	GT	G	A	0,324	0	0,615	0,06	0,076	0,924	0,065	0,147	0,787	brun foncé - noir	brun	brun foncé	brun																
M5	C	G	G	C	G	C	C	G	C	G	T	A	C	T	C	C	C	C	G	A	A	G	G	C	0,16	0	0,832	0,008	0,018	0,982	0	0,006	0,994	noir	brun	noir	brun																
M6	C	G	G	C	G	C	C	-	C	G	T	A	GC	T	CA	C	A	C	A	G	A	G	G	CA	0,228	0,002	0,163	0,608	0,838	0,162	0,678	0,131	0,191	brun clair-brun foncé	bleu	brun	gris/vert																
M7	C	G	G	C	G	C	C	-	C	G	T	A	G	C	CA	C	A	C	GA	GA	A	GT	G	A	0,235	0,001	0,074	0,689	0,866	0,134	0,15	0,14	0,711	brun clair-brun foncé	brun	brun	brun																
M8	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	C	C	CA	C	GA	GA	A	G	G	CA	0,348	0,002	0,328	0,321	0,475	0,525	0,094	0,148	0,758	brun clair-brun foncé	brun	brun	brun																
M9	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	CA	CT	CA	C	GA	GA	A	G	GA	C	0,445	0,002	0,476	0,078	0,154	0,846	0,19	0,271	0,539	brun clair-brun foncé	brun	blond foncé	gris/bleu																
M10	C	G	G	C/T	G	C	C	G	C	G	T	A	G	T	A	C	C	C	GA	G	A	G	GA	A	0,064	0,052	0,021	0,863	0,982	0,018	0,899	0,066	0,035	blond - brun clair	bleu	roux	bleu																
M11	C	G	G	C/T	G	C	C/T	G	C	G	T	A	G	T	CA	CT	C	C	A	G	GA	GT	GA	CA	0,048	0,8	0,008	0,145	0,972	0,028	0,943	0,045	0,012	roux	bleu	roux	bleu																
M12	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	CA	C	C	C	G	G	A	G	GA	CA	0,22	0,006	0,086	0,686	0,896	0,104	0,899	0,066	0,035	brun clair - brun foncé	bleu	brun	bleu																
M13	C	G	G	C/T	G	C	C	G	C	G	T	A	G	CT	CA	C	CA	CT	GA	G	GA	G	G	CA	0,074	0,055	0,008	0,862	0,989	0,011	0,91	0,072	0,018	blond	bleu	roux	bleu																
M14	C	G	G	C	GT	C	C	-	C	AG	T	A	G	T	CA	C	C	T	A	A	A	GT	GA	A	0,286	0,001	0,407	0,305	0,566	0,434	0,019	0,218	0,763	brun clair - brun foncé	brun	brun	brun																
M15	C	G	G	C	G	C	C	-	C	G	T	A	G	CT	C	C	A	C	GA	G	A	G	G	A	0,142	0,002	0,044	0,812	0,931	0,069	0,87	0,076	0,053	blond	bleu	blond	bleu																
M16	C	G	G/A	C	G	C	C	-	C	G	T	C	C	T	C	C	C	CT	GA	A	GA	G	G	C	0,127	0	0,871	0,003	0,002	0,998	0	0,017	0,983	noir	brun	noir	brun																
M17	C	G	G/A	C	G	C	C	-	C	G	T	CA	GC	T	C	C	C	C	GA	A	G	G	G	C	0,211	0	0,778	0,011	0,009	0,991	0	0,013	0,987	noir	brun	noir	brun																

Les SNP surlignés en gris ont été analysés par séquençage.

Annexe 7 : Génotypes obtenus pour les 5 SNP AIM et résultats obtenus après leur analyse avec le logiciel STRUCTURE

Echantillons		rs1545397	rs2814778	rs1426654	rs16891982	rs3827760	Valeurs obtenues avec STRUCTURE *		
		T/A	G/A	G/A	G/C	C/T	Cluster 1	Cluster 2	Cluster 3
Echantillons modernes									
Européen	M1	A	A	A	GC	T	0.949	0.030	0.020
Européen	M2	A	A	A	G	T	0.990	0.005	0.005
Européen	M3	A	A	A	G	T	0.992	0.004	0.004
Européen	M4	TA	A	A	GC	T	0.914	0.022	0.065
Africain	M5	A	G	GA	C	T	0.028	0.967	0.005
Européen	M6	A	A	A	GC	T	0.955	0.028	0.017
Européen	M8	A	A	A	G	T	0.991	0.006	0.003
Européen	M9	A	A	A	G	T	0.991	0.005	0.004
Européen	M10	A	(A)	A	G	T	0.989	0.007	0.004
Européen	M11	A	A	A	G	T	0.991	0.005	0.004
Européen	M12	TA	A	A	G	T	0.984	0.005	0.011
Européen	M13	A	A	A	G	T	0.990	0.006	0.004
Européen	M14	A	A	A	G	T	0.991	0.005	0.004
Européen	M15	A	A	A	G	T	0.991	0.005	0.004
Asiatique	M16	TA	A	G	C	C	0.010	0.015	0.975
Asiatique	M17	TA	A	G	GC	C	0.157	0.014	0.830
Echantillons anciens									
Kaf	Kh19	A	(A)	A	GC	T	0.961	0.024	0.015
	Kh20	A	(A)	A	G	T	0.991	0.005	0.004
	Kh21	A	(A)	A	GC	T	0.968	0.020	0.011
	Kh22	A	(A)	A	C	T	0.857	0.091	0.051
Ok	Kh1	-	(A)	(A)	(C)	(T)	-	-	-
	Kh2	(TA)	(A)	(GA)	(C)	(CT)	0.260	0.036	0.705
	Kh3	-	-	(A)	-	(CT)	-	-	-
	Kh7	A	(A)	A	(GC)	C	0.626	0.008	0.366
	Kh8	(A)	(A)	(A)	(C)	(CT)	0.649	0.021	0.329
	Kh12	(TA)	(A)	(G)	(C)	(T)	0.085	0.225	0.690
	Kh13	(T)	(A)	(GA)	(GC)	(C)	0.486	0.012	0.503
	Kh14	(A)	(A)	(A)	(GC)	(T)	0.967	0.020	0.013
	Kh15	-	-	(A)	(C)	(CT)	-	-	-
	Kh17	(A)	-	(A)	(C)	(CT)	-	-	-
Bol	11	-	-	(A)	(C)	-	-	-	-
Aaf	2	A	(A)	A	G	T	0.990	0.005	0.005
	3	-	-	-	(C)	-	-	-	-
El	6	(T)A	(A)	A	C	-	-	-	-
	7	(T)	-	(GA)	(G)	(CT)	-	-	-
	8	-	-	(A)	(GC)	(CT)	-	-	-
	9	(AT)	(A)	(A)	C	(CT)	0.496	0.011	0.493
AT_626	AT_626			(G)	(C)	(C)	-	-	-
	AT_628	(A)	(A)	(GA)	(C)	-	-	-	-
	AT_635	(T)	(A)	(GA)	(C)	(C)	0.045	0.007	0.948
Sag	TA4	T	(A)	A	G	C	0.519	0.004	0.477
	TA8	A	(A)	GA	G	T	0.899	0.060	0.041
	TA9	TA	(A)	GA	GC	CT	0.486	0.014	0.499
	TA10	A	(A)	GA	G	CT	0.751	0.018	0.231
	TA11	TA	(A)	GA	C	T	0.458	0.083	0.459
	TA12	A	(A)	A	C	T	0.867	0.086	0.047
	TA14	A	A	GA	G	CT	0.761	0.016	0.222
	TU15	T	A	GA	G	T	0.679	0.011	0.310
	TU17	A	A	A	GC	CT	0.829	0.013	0.158
Maf	TU34	A	A	A	G	T	0.991	0.005	0.004
	TU31	A	(A)	GA	GC	T	0.752	0.164	0.084
Chem	13	(A)	(A)	(GA)	(C)	(T)	0.540	0.298	0.162
	15	(A)	(A)	(A)	(C)	(T)	0.865	0.097	0.037
	16	(T)	(A)	(A)	(C)	(T)	0.546	0.019	0.435
MK	AT_614	(T)	(A)	(GA)	(C)	(CT)	0.111	0.011	0.878
	AT_861	-	-	(G)	(C)	-	-	-	-

* Fractions de génome de chaque individu provenant de la population K.

Annexe 8 : Génotypes obtenus avec des quantités décroissantes d'ADN

Echantillons	quantité d'ADN	N29insA	rs1042602	rs12203592	rs12821256	rs12896399	rs12913832	rs1393350	rs1426654	rs1545397	rs16891982	rs1800407	rs1805005	rs1805006	rs885479	rs1805009	rs2378249	rs2402130	rs2814778	rs28777	rs3827760	rs4959270	rs683
Ind1	10ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	10ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	5ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	5ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	2ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	2ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	1ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	1ng	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	500pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	500pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	250pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	250pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	100pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	100pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	50pg	C	CA	C	CT	G	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind1	50pg	C	CA	C	CT	GT	A	GA	A	A	G	C	G	C	G	-	A	A	A	A	T	CA	CA
Ind2	10ng	C	C	C	T	G	A	G	GA	A	C	C		C	G	-	A	G	G	A	T	C	C
Ind2	10ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	5ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	5ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	2ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	2ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	1ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	1ng	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	500pg	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	500pg	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	250pg	C	C	C	T	G	A	G	GA	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	250pg	C	C	C	T	G	A	G	A	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	100pg	C	C	C	T	G	A	G	A	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	100pg	C	C	C	T	G	A	G	A	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	50pg	C	C	C	T	G	A	G	G	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind2	50pg	C	C	C	T	G	A	G	A	A	C	C	G	C	G	-	A	G	G	A	T	C	C
Ind3	10ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	10ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	5ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	5ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	2ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	2ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	1ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	1ng	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	500pg	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	500pg	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	250pg	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	250pg	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	100pg	C	A	C	CT	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A
Ind3	100pg	C	A	C	C	G	G	G	A	A	G	C	G	C	G	-	A	G	A	A	T	C	A
Ind3	50pg	-	A	C	T	G	G	G	A	A	G	C	G	C	G	-	A	GA	-	A	T	C	A
Ind3	50pg	-	A	C	T	G	G	G	A	A	G	C	G	C	G	-	A	GA	A	A	T	C	A

Les résultats surlignés en gris sont différentes du génotype attendu.

Annexe 9 : Génotypes et résultats obtenus pour la prédiction de la couleur des yeux et des cheveux pour les individus anciens

Echantillons		N29msA	rs11547464	rs885479	rs1805008	rs1805005	rs1805006	rs1805007	rs1805009	Y1520CH	rs2228479	rs1110400	rs28777	rs16891982	rs12821256	rs4959270	rs12203592	rs1042602	rs1800407	rs2402130	rs12913832	rs2378249	rs12896399	rs1393350	rs683	Valeurs obtenues avec l'outil de prédiction									phénotype prédit (couleur la plus probable)		
																										Cheveux				nuance		yeux					
																										C/A	G/A	G/A	C/T	G/T	C/A	C/T	G/C	C/A	A/G	T/C	C/A
Kaf	Kh19	C	-	-	-	G	C	-	-	-	G	-	CA	GC	T	A	CT	C	C	A	(A)	A	GT	G	C	0,15	0	0,845	0,005	0,007	0,993	0,001	0,025	0,974	brun foncé - noir	brun	
	Kh20	C	-	-	-	G	C	-	-	-	G	-	(A)	G	(CT)	A	T	(C)	C	A	(G)A	A	(GT)	G	(C)A	0,432	0,001	0,503	0,064	0,157	0,843	0,28	0,315	0,405	brun clair - brun foncé	Int.	
	Kh21	C	-	-	-	G	C	-	-	-	G	-	(C)A	GC	T	CA	C	(C)	CT	A	G	A	G	G	CA	0,221	0,001	0,191	0,586	0,759	0,241	0,791	0,138	0,071	brun clair -brun foncé	bleu	
	Kh22	C	-	-	-	G	C	-	-	-	G	-	C	C	T	A	C	(C)	C	A	A	A	G	G	C(A)	0,104	0	0,889	0,007	0,006	0,994	0	0,006	0,994	brun foncé - noir	brun	
	Kh1	-	-	-	-	(G)	(C)	-	-	-	-	(G)	-	(C)	(C)	(T)	(C)	(C)	(C)	(G)	(A)	(A)	(G)	(G)	(C)	(C)	-	-	-	-	0	0,006	0,994	-	-	brun	
Ok	Kh2	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(C)	(C)	(C)	(C)	(GA)	(A)	(A)	(GT)	(G)	(C)	0,087	0	0,91	0,002	0,001	0,999	0	0,006	0,994	brun foncé - noir	brun	
	Kh3	-	-	-	-	(G)	(C)	-	-	-	(G)	-	-	-	(T)	-	(C)	(C)	(C)	-	(A)	-	(G)	(G)	(C)	-	-	-	-	-	-	(brun)	-	-	(brun)		
	Kh7	-	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(A)	(C)	(C)	(C)	(A)	(G)	(A)	(T)	(G)	(CA)	-	-	-	-	-	0,623	0,094	0,283	-	-	bleu	
	Kh8	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(C)	(C)	(C)	(CT)	(A)	(A)	(GA)	(G)	(G)	(C)	0,128	0	0,867	0,005	0,003	0,997	0	0,017	0,983	brun foncé - noir	brun	
	Kh12	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(CA)	(C)	(T)	(CA)	(C)	(C)	(C)	(GA)	(A)	(GA)	(G)	(G)	(C)	0,159	0	0,833	0,007	0,008	0,992	0	0,006	0,994	brun foncé - noir	brun	
	Kh13	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(A)	(GC)	(T)	(C)	(C)	(C)	(C)	(A)	(GA)	(GA)	(G)	(G)	(CA)	0,285	0	0,488	0,227	0,381	0,619	0,024	0,083	0,892	brun clair - brun foncé	brun	
	Kh14	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(CA)	(GC)	(T)	(A)	(C)	(CA)	(C)	(A)	(A)	(A)	(T)	(G)	(A)	0,189	0	0,764	0,047	0,051	0,949	0,001	0,013	0,987	brun foncé - noir	brun	
	Kh15	-	-	-	-	(G)	(C)	-	-	-	(A)	-	-	(C)	(T)	(C)	(C)	(A)	(C)	(A)	(A)	(A)	(G)	(G)	(C)	-	-	-	-	0	0,006	0,994	-	-	brun		
	Kh17	-	-	-	-	-	-	-	-	-	-	-	-	(C)	-	-	(C)	(C)	(C)	(A)	(A)	-	(G)	(G)	(C)	-	-	-	-	-	-	(brun)	-	-	(brun)		
	Bol	11	-	-	-	-	(G)	(C)	-	-	-	(G)	-	-	(C)	-	(A)	(T)	(C)	(C)	(C)	(A)	(A)	(A)	-	-	-	-	-	-	-	(brun)	-	-	(brun)		
Aaf	2	-	-	-	-	G	C	-	-	-	G	-	A	G	T	C(A)	C	(C)	C	A	GA	A	GT	G	A	-	-	-	-	0,15	0,14	0,711	-	-	brun		
	3	-	-	-	-	-	-	-	-	-	-	-	-	(C)	(T)	-	-	-	-	(A)	-	(G)	-	-	-	-	-	-	-	-	-	-	-	-	-		
El	6	-	-	-	-	(G)	C	-	-	-	G	-	(CA)	C	T	-	C	C	C	(A)	A	(G)	(G)T	G	(C)	-	-	-	-	-	-	0	0,006	0,994	-	-	brun
	7	-	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(G)	(T)	(A)	-	(C)	(C)	-	(A)	(GA)	(G)	-	(C)	-	-	-	-	-	-	(brun)	-	-	(brun)		
	8	-	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(GC)	(T)	-	(C)	(C)	(C)	(A)	-	-	(T)	(G)	(C)	-	-	-	-	-	-	-	-	-	-		
	9	-	-	-	-	G	C	-	-	-	(A)	-	(C)	C	T	(C)	C	(C)	(C)	-	(A)	A	(G)	(G)	(C)	-	-	-	-	-	0	0,006	0,994	-	-	brun	
ChemN	AT_626	-	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(CA)	(C)	(C)	(C)	(A)	(A)	(GA)	(T)	(G)	(A)	-	-	-	-	-	0	0,006	0,994	-	-	brun	
	AT_628	-	-	-	-	(G)	(C)	-	-	-	(G)	-	-	(C)	(T)	(A)	(C)	(C)	(C)	(A)	(GA)	(GA)	(G)	(G)	(C)	-	-	-	-	-	0,006	0,043	0,952	-	-	brun	
	AT_635	-	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(C)	-	(C)	(C)	(A)	(A)	(A)	(G)	(G)	(C)	-	-	-	-	-	(brun)	-	-	(brun)			
Sag	TA4	C	G	G/A	C	G	C	C	-	C	G	T	A	G	T	CA	CT	(C)	C	A	A	A	G	G	A	0,192	0	0,785	0,023	0,047	0,953	0,002	0,05	0,948	brun foncé - noir	brun	
	TA8	C	-	-	-	G	C	-	-	-	G	-	A	G	CT	A	CT	(C)	CT	A	(G)A	GA	G(T)	G	C	-	-	-	-	-	0,363	0,336	0,3	-	-	int.	
	TA9	C	G	G	C	G	C	C	-	C	G	T	CA	GC	T	C	C	CA	C	GA	GA	G	GT	G	0,375	0,001	0,549	0,075	0,061	0,939	0,04	0,083	0,877	brun foncé - noir	brun		
	TA10	(C)	G	(A)	C	G	C	C	-	C	G	T	A	G	T	C	C	C	CT	A	A	A	G	G	C	0,261	0	0,658	0,081	0,177	0,823	0,003	0,071	0,936	brun clair - brun foncé	brun	
	TA11	C	G	A/G	C	G	C	C	-	C	G	T	C	C	CT	C	(C)	C	GA	A	A	G	GT	G	A	0,171	0	0,824	0,005	0,003	0,997	0	0,006	0,994	brun foncé - noir	brun	
	TA12	C	G	G	C	G	C	C	-	C	G	T	C	C	T	A	C	C	C	G	A	A	G	G	(C)A	0,117	0	0,878	0,005	0,003	0,997	0	0,006	0,994	brun foncé - noir	brun	
	TA14	C	G	G/A	C	G	C	C	-	C	G	T	A	G	T	CA	CT	C	C	GA	G	GA	G	G	C	0,455	0,003	0,288	0,255	0,534	0,466	0,885	0,085	0,03	brun clair - brun foncé	bleu	
	TU15	C	(G)	(G)	(C)	G	C	(C)	-	(C)	G	(T)	A	G	T	C	C	C	(A)	A	A	T	G	CA	0,223	0	0,65	0,217	0,171	0,829	0,003	0,026	0,971	brun foncé - noir	brun		
	TU17	C	G	A/G	C	GT	C	C	-	C	G	T	CA	GC	T	C	C	C	C	A	GA	A	T	G	CA	0,256	0	0,66	0,083	0,118	0,882	0,066	0,081	0,853	brun clair - brun foncé	brun	
	TU34	C	G	-	C	G	C	C	-	C	G	T	A	G	T	C	C	C	C	GA	GA	GA	G	G	CA	0,31	0,002	0,322	0,366	0,49	0,51	0,094	0,148	0,758	brun clair - brun foncé	brun	
Maf	TU31	C	-	A/G	C	G	C	C	-	C	G	T	C(A)	GC	T	A	C	C	C	(A)	A	GA	GT	GA	CA	0,203	0	0,765	0,031	0,039	0,961	0,001	0,016	0,983	brun foncé - noir	brun	
Chem	13	(C)	-	-	-	(T)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(CA)	(C)	(C)	(C)	(A)	(A)	(A)	(G)	(GA)	(CA)	0,089	0	0,906	0,005	0,004	0,996	0	0,008	0,992	brun foncé - noir	brun	
	15	(C)	-	-	-	(G)	(C)	-	-	-	(AG)	-	(A)	(C)	(T)	(A)	(C)	(C)	(C)	(A)	(A)	(A)	(G)	(G)	(A)	0,135	0	0,833	0,033	0,111	0,889	0	0,006	0,994	brun foncé - noir	brun	
	16	(C)	-	-	-	(G)	(C)	-	-	-	(A)	-	(C)	(C)	(T)	(A)	(C)	(C)	(CT)	(A)	(A)	(A)	(G)	(G)	(C)	0,142	0	0,849	0,009	0,009	0,991	0	0,017	0,983	brun foncé - noir	brun	
MK	AT_614	(C)	-	-	-	(G)	(C)	-	-	-	(G)	-	(C)	(C)	(T)	(CA)	(C)	(C)	(C)	(A)	(A)	(A)	(GT)	(G)	(C)	0,094	0	0,902	0,004	0,003	0,997	0	0,006	0,994	brun foncé - noir	brun	
	AT_861	-	-	-	-	-	-	-	-	-	-	-	-	(C)	-	-	(C)	-	(C)	-	-	(A)	(G)	(C)	-	-	-	-	-	-	-	-	-	-	-		

Les SNP surlignés en gris ont été analysés par séquençage. Les allèles entre parenthèses n'ont été observés qu'une seule fois. La couleur des yeux indiquée entre parenthèse a été prédite uniquement à partir du SNP rs12913832 (bleu : A/A ; brun : G/G ou G/A). Int : Intermédiaire.

VIII PUBLICATIONS

**Application of the iPLEX™ Gold SNP genotyping method for the analysis of
Amerindian ancient DNA samples: benefits for ancient population studies**

Mendisco F, Keyser C, **Hollard C**, Seldes V, Nielsen AE, Crubézy E, Ludes B.

Electrophoresis. 2011 Feb;32(3-4):386-93

Fanny Mendisco^{1,2}
Christine Keyser^{1,2}
Clémence Hollard¹
Veronica Seldes³
Axel E. Nielsen³
Eric Crubézy²
Bertrand Ludes^{1,2}

¹Laboratoire d'Anthropologie Moléculaire, Université de Strasbourg, Institut de Médecine Légale, Strasbourg, France

²Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), CNRS FRE 2960, Toulouse, France

³INAPL, CONICET, Ciudad de Buenos Aires, Argentina

Received September 20, 2010

Revised November 16, 2010

Accepted November 17, 2010

Research Article

Application of the iPLEXTM Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: Benefits for ancient population studies

Important developments in the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) technique have generated new perspectives regarding SNP genotyping, which are particularly promising for ancient population-based studies. The main aim of the present study was to investigate the application of a MALDI-TOF MS-based SNP genotyping technique, called iPLEX[®] Gold, to analyze Amerindian ancient DNA samples. The first objective was to test the sensitivity of the method, which is recommended for DNA quantities between 10 and 5 ng, for ancient biological samples containing DNA molecules that were degraded and present in minute quantities. The second objective was to detail the advantages of this technique for studies on ancient populations. Two multiplexes were designed, allowing the major Amerindian mitochondrial and Y haplogroups to be determined simultaneously. This analysis has never been described before. Results demonstrated the reliability and accuracy of the method; data were obtained for both mitochondrial and nuclear DNA using picogram (pg) quantities of nucleic acid. This technique has the advantages of both MS and minisequencing techniques; thus, it should be included in the protocols for future ancient DNA studies.

Keywords:

Ancient DNA / MALDI-TOF MS / Mitochondrial haplogroup / SNP genotyping / Y haplogroup
DOI 10.1002/elps.201000483

1 Introduction

Mass spectrometry (MS), initially developed for proteomics, has rapidly become widespread for the analysis of nucleic acids [1, 2]. Indeed, several SNP genotyping methods based on MS have been developed using different platforms, including electrospray ionization time-of-flight (ESI-TOF) [3, 4] or matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF). SNP genotyping by MALDI-TOF MS emerges as a very powerful and attractive approach [5, 6], due to the speed of signal acquisition, direct measurement of the intrinsic properties of molecules and, finally, the possibility of a complete automation from the sample

preparation to data acquisition [7]. Therefore, numerous genotyping methods, based on diverse types of allelic discriminations, have been developed [8–10]. Among them, protocols based on a single base primer extension were targeted and incessantly improved [4, 5, 8]. Today, SNP genotyping turn-key methods like the iPLEX[®] Gold technology, developed by the Sequenom company, are available [11]. This method has major advantages, including those of MS, as well as the high capacity of multiplexing (up to 40 SNPs can be analyzed simultaneously) [12], and an ability to work with small-sized amplicons (around 100 bp). These advantages are very important when working on ancient or degraded DNA material.

Ancient or degraded DNA molecules present characteristics that make their analysis challenging: they are most often fragmented, present in minute quantities and co-extracted with inhibitive substances. The preparation and analysis of ancient DNA samples requires rigorous and appropriate protocols [13, 14]. Most of the studies on past human population are based on the sequencing of the first and sometimes the second hypervariable segment (HVS-1, HVS-2) of the mitochondrial DNA (mtDNA) control region. In some of these studies, sequencing is completed by the

Correspondence: Fanny Mendisco, Institut de Médecine Légale, Laboratoire d'Anthropologie Moléculaire EA 4438, 11 rue Humann, 67085 Strasbourg Cedex, France
E-mail: fanny.mendisco@neuf.fr
Fax: +33-3-68-85-33-62

Abbreviations: HVS-1, mitochondrial first hypervariable segment; mtDNA, mitochondrial DNA; mt-SNP, mitochondrial DNA SNP

analysis of haplogroup-tagging SNPs of the mtDNA coding region. Until recently, this typing was done mainly by the RFLP method [15], which is a time- and sample-consuming method, or sometimes by SNaPshot® [16]. It is important to find alternative techniques, which allow greater degrees of discrimination than those achieved by the analysis of a restricted portion of the mitochondrial genome. In this framework, the recent development of high-throughput sequencing techniques [17, 18] has brought new perspectives concerning the ancient DNA problematic [19]. However, complete genome sequencing of ancient human samples remains limited to the analysis of very precious samples whose evolutionary history could be highly relevant [20]. The high cost of the complete genome sequencing approach does not allow a large number of ancient samples to be tested; thus, it cannot be used for population studies. Moreover, due to the limited quantity of endogenous DNA available in the majority of ancient samples this approach is not necessarily appropriate. An alternative to complete genome sequencing can be to target SNPs by MALDI-TOF MS. This approach may prove to be promising for ancient DNA studies since it is described as an accurate, cost-effective, fast and sensitive method, ideally suited for the analysis of short DNA fragments. The only question concerns the quantity of DNA required. Indeed, all the genotyping technologies based on MALDI-TOF MS, and in particular the iPLEX® Gold method, require starting quantities of DNA between 5 and 10 ng. Such quantities are difficult to obtain with ancient DNA extracts; thus, the aim was to determine the validity of using this method for ancient DNA.

Various aspects were addressed during this study. First, we tested the sensitivity and accuracy of the method for analyzing very low quantities of DNA, by comparing the results with those obtained by traditional methods (sequencing and STR typing). The second aspect relates to the resolution achieved in the studies on ancient DNA. For this, we tried to develop assays which allowed Y and mitochondrial phylogenies to be analyzed at the same time, with the highest possible degree of discrimination. We also wanted to evaluate the benefits of the method for ancient population studies. For this purpose, we selected ancient samples from northwest Argentina in South America. The aim was to characterize the genetic diversity of an ancient population, the Omaguacas, in order to gain a better understanding of their origin and their genetic affinities with ancient and modern populations.

2 Materials and methods

2.1 Samples

The ancient samples analyzed come from the archaeological site of Los Amarillos (LA), located in the Humahuaca Valley in northwest Argentina (Supporting Information Fig. 1). This site was occupied for several centuries, from the

beginning of the 11th century until the arrival of the Europeans during the 16th century [21]. The structure of the ancient Omaguacas population in the Humahuaca valley remains relatively unknown. Archaeological evidence suggests that this valley was located in the middle of a regional circuit for the traffic of goods, from approximately the 9th century [22]. Indeed, groups coming from various regions, such as south Bolivia, the Puna region and the eastern valleys and forests (Supporting Information Fig. 1), could have contributed to the composition of this population.

Genetic analyses were carried out on teeth samples carefully taken from 23 ancient skeletons (at least two teeth from each individual), which were excavated, from 1992 onwards, within the framework of the P.I.O. (Programa de Investigacion Omaguaca) project. The sex of each individual was not morphologically determined, but the analysis of the amelogenin sex marker allowed us to determine that the sample was composed of 12 males, 6 females and 5 individuals whose sex remained unknown. All these individuals are dated between the 11th and 15th centuries, due to radiocarbon dating [21].

To validate the results and to test the assays, DNA extracts of modern Amerindian samples were included in the study: four male samples and one female sample. Each of the four male samples belongs to the Y paragroup Q1a3a*, typical of Native American populations [23], and to a different major Amerindian mitochondrial lineage: A2, B2, C1 and D1 [24–26]. The female sample belongs to the B2 mitochondrial haplogroup.

2.2 DNA extraction and quantification

DNA extraction was preceded by a decontamination step in order to eliminate surface contamination. Each tooth was cleaned with bleach, rinsed with deionized water and, finally, irradiated under UV light for 30 min on each face. Powder was generated by grinding the tooth under liquid nitrogen with a 6870 SamplePrep Freezer Mill® (Fischer Bioblock, Illkirch, France). DNA extraction was carried out using the NucleoSpin® Extract II kit (Macherey-Nagel, Düren, Germany), following the manufacturer's protocol, except for the buffer volumes. The first step of this extraction protocol consisted of an overnight incubation at 50°C of 150–200 mg of tooth powder in a lysis buffer containing 500 µL of EDTA (0.5 µM), 50 µL of proteinase K (20 mg/mL) and 5 µL of DTT (1 M). Second, the supernatant was purified onto NucleoSpin® Extract II silica columns (Macherey-Nagel) in three steps: (i) 1 mL of binding NT buffer was added so that the DNA bound to the silica membrane; (ii) 600 µL of NT3 ethanolic buffer was added in the washing step; and (iii) purified DNA was eluted on 250 µL using the NE buffer (5 mM Tris-HCl, pH 8.5). Finally, the extracts were concentrated using Amicon® Ultra-0.5 30 KDa columns (Millipore, Billerica, USA), in a final volume of 40 µL.

Quantification of all DNA extracts was made using the Quantifiler[®] Human DNA Quantification Kit (Applied Biosystems, Foster City, CA, USA), on an ABI PRISM[®] 7000 Sequence Detection System (Applied Biosystems, Courtaboeuf, France), following the recommended protocol of the company.

Precautions were taken to avoid contamination with modern DNA, as previously described in Carnese et al. [27].

2.3 Standard STR and mtDNA typing

Fifteen autosomal STRs (D8S1179, D7S820, D3S1358, D13S317, D16S539, D2S1338, D19S433, D5S818, D21S11, CSF1PO, vWA, TH01, TPOX, D18S51 and FGA) and the sex-determining marker amelogenin were amplified using the AmpFlSTR[®] Identifiler[™] kit (Applied Biosystems). Seventeen Y-chromosomal STR loci (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and Y GATA H4) included in the AmpFlSTR[®] Y-filer[™] Kit (Applied Biosystems) were analyzed for the DNA of the ancient male samples. The experimental conditions were those recommended by the manufacturer, except that 34 PCR cycles were used instead of 28 and 30 for the Identifiler and Y-filer amplification kit, respectively. The amplified autosomal and Y-chromosomal STR products were analyzed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems) using the GeneMapper software, version 3.2.1 (Applied Biosystems).

Sequencing of the mitochondrial HVS-1 portion was carried out using two sets of primers as described previously in Carnese et al. [27]. To evaluate the accuracy of the method, mitochondrial haplogroups were determined by MS genotyping and were compared to those determined from the sequencing of HVS-1.

2.4 SNP selection

In an attempt to develop an assay based on the simultaneous typing of mitochondrial and Y-chromosome SNPs in the population studied, we selected from the literature [23–26, 28, 29] a set of 24 mitochondrial and Y-chromosome markers that define the major mitochondrial and Y-chromosome Amerindian haplogroups (Fig. 1).

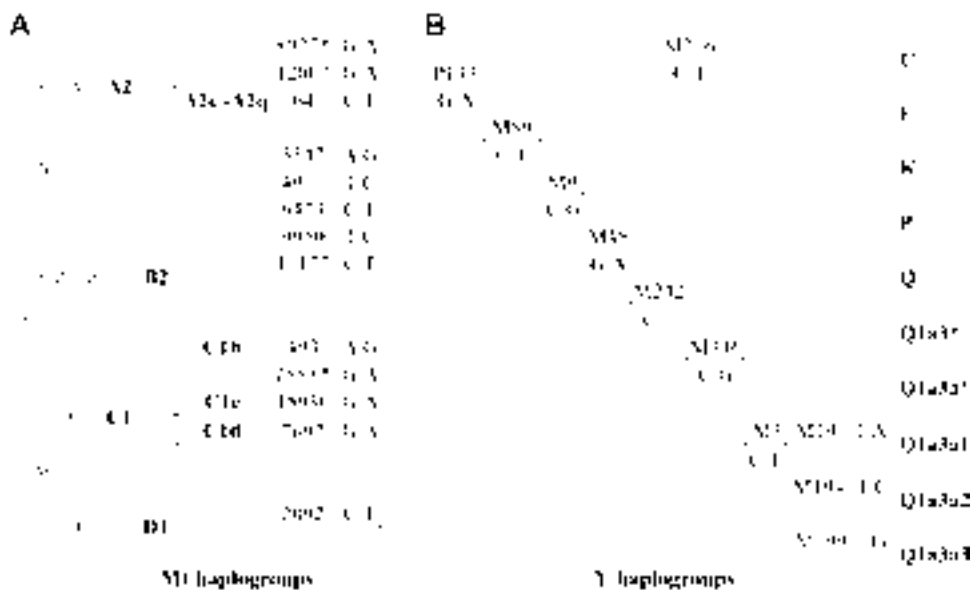
Concerning the mitochondrial lineages, the 13 selected SNPs permit the definition and refinement of the A2, B2, C1 and D1 haplogroups, which are the most frequently observed haplogroups in Amerindian populations [24–26]. Since the iPLEX[®] Gold method has not been described for low quantities of DNA, we decided to select several markers, which are characteristic of the same haplogroup, in order to optimize the probability of an interpretable result.

Regarding the Y haplogroups, the Amerindian populations present a low genetic diversity. Two major haplogroups are described: the Q haplogroup, which is the most prevalent, and the C haplogroup. We selected 11 SNPs to define these two major haplogroups and also to refine sub-haplogroups of the Q lineage, including the Q1a3a sub-haplogroup, which is considered characteristic of Native American populations [23, 28, 29]. For these markers, the nomenclature used follows that of the Y-chromosome consortium [30].

2.5 iPLEX[®] Gold SNP genotyping method

The genotyping analysis was performed using the MassARRAY[®] compact system (Sequenom, San Diego, USA) associated with the iPLEX[®] Gold SNP genotyping kit (Sequenom).

The manufacturer's protocol was used [11, 31], except that the quantity of DNA used in the reaction was lower



than the recommended measure. With the aim of having the highest possible quantity of DNA, the volume of ancient DNA extract added in the reaction was increased from 2 to 3.4 μL . PCR amplifications were performed in a total reaction volume of 5 μL with 1 $\times$ PCR buffer, MgCl_2 (2 mM), dNTPs (500 μM), PCR primer (0.1 μM each) and HotStar-Taq[®] (0.5 U). Prior to the mini-sequencing reaction, in order to digest unincorporated dNTPs, 2 μL of a solution containing SAP (0.5 U) was added to the PCR reaction and incubated at 37°C for 40 min and 85°C for 5 min. The single base extension reaction was made by adding 2 μL of an iPLEX Gold extension reaction cocktail containing: 10 $\times$ iPLEX buffer, 1 $\times$ iPLEX termination mix, 1 $\times$ iPLEX enzyme and primer mix. After a cleanup step with the SpectroCLEAN resin, extended products were arrayed onto a SpectroCHIP using the RS1000 nanodispenser. Finally, the products were detected with the MassARRAY mass spectrometer, and the data were acquired in real time with the MassARRAY[®] RT software. Alleles were automatically assigned by the software and associated to a reliable value.

Before genotyping valuable ancient samples, we undertook a preliminary step to test the sensitivity of the technique, consisting of a dilution range (from 10 ng to 5 pg of nuclear DNA per reaction) from modern DNA samples. We used one assay of 17 autosomal and Y-chromosomal SNPs, designed previously in the laboratory for another project (unpublished data).

To test the reproducibility of the iPLEX[®] method we analyzed two different extracts (obtained from two teeth), for each of the 23 ancient individuals. As recommended in the protocol, the deposits onto the SpectroCHIP were made in duplicate; we obtained four spectra for each ancient individual. From these four spectra, we determined a consensus genotype, keeping the alleles that appeared at least twice. Positive (modern Amerindian DNA) and negative controls (extraction and PCR controls, modern African DNA) were added during the genotyping process in order to detect any problems (e.g. contamination, false positive).

2.6 Primer design

The design of the amplification and extension primers was made using the MassARRAY[®] assay Design 4.1 software, associated with the MassARRAY[®] compact system. We tried to incorporate all the markers in the same assay; however, according to the software, the 24 selected SNPs were not compatible to be grouped into a single multiplex. Two mt-SNPs (mitochondrial SNPs) (A2_8027 and C1c_1888) were excluded from the assay. Finally, the software designed two multiplexes containing 11 mitochondrial and Y-chromosome SNPs each. All the amplification and extension primers used are listed in Supporting Information Table 1. Both multiplexes were applied to each sample, except the ancient female samples, for which only the second multiplex (plex 2) was analyzed.

3 Results

3.1 Quantification of the ancient DNA extracts and results of the standard STR and mitochondrial typing

The full set of quantification results, HVS-1 sequencing and STR typing obtained is reported in Supporting Information Table 2.

The quantifications carried out on the ancient DNA extracts showed concentrations ranging from 1.4 ng/ μL (LA13) to values below the detection limit of the kit (1.07 pg/ μL for LA1). DNA extracts of six individuals (LA2, LA5, LA6, LA11, LA19 and LA22) could not be quantified, since they were apparently too degraded.

The sequencing of HVS-1 resulted in the maternal lineages of 18 of the 23 analyzed individuals to be determined. Indeed, our sample was composed of 13 individuals belonging to the A2 lineage, three (LA9, LA16 and LA4) to the D1 lineage, one (LA17) to the B2 lineage and one (LA18) to the C1 lineage. It was not possible to determine the mitochondrial haplotypes of five samples: LA2, LA5, LA6, LA11 and LA19.

As shown in Supporting Information Table 2, we obtained relatively complete profiles with regard to autosomal and Y-chromosome STRs. For six of the ancient samples (LA1, LA2, LA5, LA6, LA11 and LA19), none of the loci, autosomal or Y-chromosome, were amplified. Partial profiles (with less than half of the loci typed) were obtained for five samples (LA7, LA10, LA12, LA22 and LA24). Regarding the remaining 12 individuals, we obtained complete or nearly complete STR profiles using the Identifier and Y-filer kits.

Several data revealed that the DNA extracted from the ancient samples was endogenous: (i) the STR profiles of each ancient sample were unique and different from those of the operators and excavators; and (ii) the mitochondrial haplotypes obtained on sequencing of HVS-1 were concordant with the area under study.

3.2 Validation of the SNP genotyping assays

The dilution range performed prior to this study showed that the iPLEX[®] technique was reliable and sensitive (with a call rate greater than 85%) until quantities of 100 pg of nuclear DNA per reaction (unpublished data).

Following this sensitivity study, we chose to test the two multiplex systems designed for this analysis with modern Amerindian DNA extracts. For both multiplexes, all the selected markers were successfully amplified and the results of the SNP genotyping were concordant with the expected haplogroups of the modern individuals (Supporting Information Table 2). In addition, this first test with modern DNA underlined one problem: unexpected peaks appeared systematically with the modern female samples and the negative controls for the Y-chromosome markers CF_P143

and Q1a3a2_M194 (Supporting Information Table 2). Several hypotheses can be postulated to explain the appearance of these unexpected peaks. The first hypothesis is that of contamination. However, in the negative PCR controls (without DNA), only these two unexpected peaks appeared. If it was a contamination with exogenous DNA or a cross-contamination between samples, other peaks would have been appeared. The second hypothesis is that of non-specific primers. However, all the primers used were compared against the human DNA sequence in GenBank and appeared to be specific for the Y-chromosome. The third and most probable hypothesis is that of an artifact induced by the biochemistry of the kit [32]. For ancient female samples or ancient samples of unknown sex, the appearance of such unexpected peaks can cause doubts in the interpretation of results. It is therefore important to test the designed assays to locate these artifacts, in order to be able to interpret the results correctly.

3.3 SNP genotyping of ancient samples

The genotyping call rate for the mt-SNPs, calculated from all the spectra obtained for the 23 ancient individuals, was 96.4%. The results obtained for the Y markers were more differentiated. Indeed, the call rate was about 74% for the 12 ancient male samples, decreasing to 62% when the results of the ancient samples of unknown sex were added.

No discordant alleles were observed in the four different spectra obtained for each ancient specimen, which concludes that the iPLEX[®] Gold method is highly reproducible. The consensus genotypes obtained are described in Table 1. With this method, the determination of both maternal and paternal lineages was possible from the same manipulation. An example of spectra obtained is shown in Fig. 2, illustrating the differences between a male and a female spectrum.

As we can see in Table 1, the determination of both mitochondrial and Y haplogroups was possible for each quantified extract. The non-quantified extracts presented relatively complete spectra; ranging from two markers out of 11 (LA5) to a complete spectra for some individuals (LA2, LA6 and LA22). Mitochondrial haplogroups determined from the SNPs genotyping were concordant with the expected haplogroups deduced from the HVS-1 sequencing (Supporting Information Table 2), showing the reliability of the method. In addition, SNP genotyping has allowed the refinement of some haplogroups (A2a–A2b for LA7, and C1b for LA18), and the determination of the maternal lineages for two samples (A2 for LA2 and LA6) whose HVS-1 portion could not be sequenced.

3.4 Genetic variation observed in the Los Amarillos ancient samples

We observed in this ancient population a high prevalence of the mitochondrial haplogroup A2 (15 individuals of

the 23 studied). It is interesting to note that the specimen LA7 presents the allele C for the A2_64 marker (Table 1) and thus, it belongs to a different maternal lineage from those of the other 14 samples belonging to the A2 haplogroup.

The diversity of the Y lineages is limited. Indeed, of the 12 male samples, 8 belong to the Q1a3a* paragroup, 2 belong to Q1a3* paragroup, and the remaining 2 probably also belong to the Q1a3a* paragroup, but 1 marker (Q1a3a1_M19) is lacking to confirm this.

4 Discussion

4.1 Behavior of the iPLEX[®] technology with ancient samples

The SNP genotyping of ancient Amerindian samples, for which DNA is fragmented, chemically modified and present in a low copy number, was an interesting means to test the sensibility of the iPLEX[®] method. The results obtained show the possibility of genotyping SNPs by MALDI-TOF MS from ancient biological samples containing few molecules of degraded DNA. Concerning mtDNA, we have not reached the detection limit for alleles. Thus, the use of Sequenom technology allows information on ancient degraded samples to be collected [33]. However, for the reliable detection of nuclear DNA, more than 100 pg of DNA per reaction is required to obtain complete genotypes.

The two multiplexes that we developed allowed the discrimination of the major Amerindian maternal and paternal lineages at the same time, and thus, permitted the quantities of DNA extract used (3.4 µL) to be optimized. The quantity of DNA used in analyses is often a limiting factor in many ancient population studies; thus, the possibility to determine mitochondrial and Y haplogroups with one assay is an interesting advancement and perspective. We can plan, in the future, to increase the multiplexing degree of the assays (especially for mtDNA) and to integrate new markers, allowing the branches of the Amerindian phylogeny to be detailed further.

With regard to our study sample, this method allowed the genetic diversity of the Los Amarillos site to be characterized. The diversity of maternal lineages described in our sample is different from that commonly observed in ancient Andean populations, which are usually characterized by a strong dominance of the B2 haplogroup [34–36]. The paternal lineages, however, correspond to the expected diversity. Indeed, the dominance of the Q1a3a* paragroup has been reported in other Amerindian populations [37, 38]. Nevertheless, the reduced sample size of 23 individuals could introduce a bias in the study, modifying the real genetic diversity of the Omaguacas. Indeed, more samples are required to better understand the genetic affinities of the Omaguacas.

Table 1. Consensus results obtained for the 23 ancient samples of the Los Amarillos site. The markers in grey correspond to the Y-chromosome SNPs

PLEX_1											PLEX_2													
CF_P143	K_M9	P_M45	Q_M242	Q1a3_M346	Q1a3_M3	Q1a3a1_M19	Q1a3a2_M194	Q1a3a3_M199	C1c_15930	C1d_7697	A2_12007	A2_64	B2_3547	B2_4977	B2_6473	B2_9950	B2_11177	C1b_493	D1_2092	F_M89	C_M216	Haplogroups	mt	Y
Ancient male samples																								
LA3	A	G	A	T	G	T	T	T	DEL	G	G	C	A	T	C	T	C	A	A	T	T	C	D1	Q1a3a*
LA4	A	G	A	T	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	A2	C	A2	Q1a3a
LA7	A	G	A	T	G	T	T	DEL	G	A	C	A	A	T	C	T	C	A	A	T	T	C	A2a-A2b	Q1a3a
LA9	A	G	A	T	G	T	T	DEL	G	G	C	A	T	T	C	T	C	A	A	T	T	C	D1	Q1a3a*
LA10	A	G	A	T	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3a*	
LA12	A	G	A	T	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3a*	
LA16	A	G	A	T	G	T	T	DEL	G	G	C	A	T	T	C	T	C	A	A	T	T	D1	Q1a3a*	
LA17	A	A	A	C	G	C	T		G	G	C	G	C	T	C	T	C	A	A	C	T	B2	Q1a3*	
LA20	A	G	A	T	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3a*	
LA21	A	A	T	G	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3a*	
LA22	A	A	T	G	G	C	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3*	
LA24	A	G	A	G	G	T	T	DEL	G	A	T	A	T	T	C	T	C	A	A	C	T	A2	Q1a3a	
Ancient female samples																								
LA1										A	T	A	T	T	C	T	C	A	A	C		A2		
LA8										A	T	A	T	T	C	T	C	A	A	C		A2		
LA13										A	T	A	T	T	C	T	C	A	A	C		A2		
LA14										A	T	A	T	T	C	T	C	A	A	C		A2		
LA18										A	T	A	T	T	C	T	C	A	A	C		A2		
LA23										G	C	A	T	T	C	T	C	G	C	C		C1b		
Ancient samples of unknown sex																								
LA2	A					T			G	A	T	A	T	T	C	T	C	A	A	C		A2	/	
LA5	A					T			G	A	T	A	T	T	C	T	C	A	A	C		/	/	
LA6						T			G	A	T	A	T	T	C	T	C	A	A	C		A2	/	
LA11						T			G	A	T	A	T	T	C	T	C	A	A	C		/	/	
LA19	A					T			G	A	T	A	T	T	C	T	C	A	A	C		/	/	

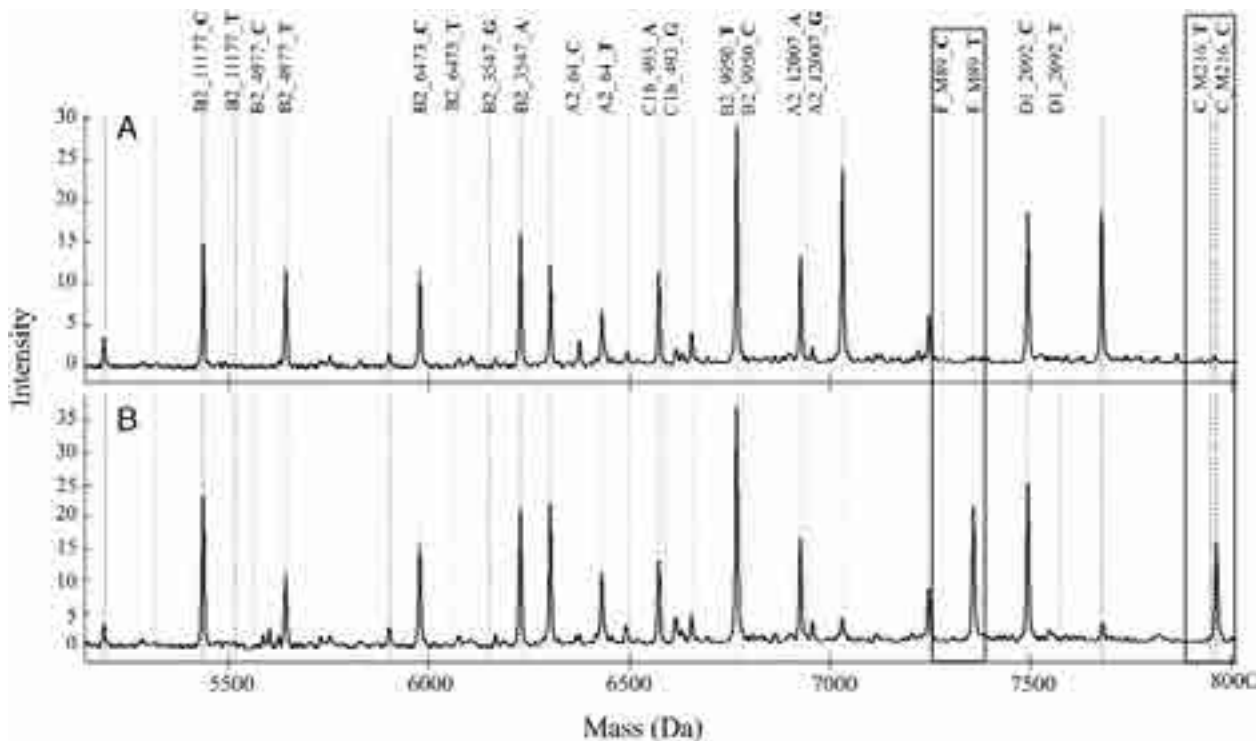


Figure 2. Example of the spectra obtained for the ancient female sample LA2 (A), and the ancient male sample LA4 (B), showing the peaks corresponding to the alleles of the “plex 2.” Both bordered markers are Y-chromosome SNP.

4.2 Benefits for ancient population studies

This study allowed us to emphasize several advantages of the iPLEX[®] method for ancient population studies. The first advantage of the iPLEX[®] technique is that it is a turn-key method. During analysis, we noted that it was a very easy and fast method to implement in the laboratory. Indeed, it allowed a quick set-up of assays and generation of results. In addition, the mt-SNP genotyping by MALDI-TOF MS provided the possibility, in a fast and effective way and by using small quantities of DNA extract: (i) to confirm the results obtained from the sequencing of the HV1 segment; (ii) to deepen the mitochondrial phylogenies determined from the HVS-1; (iii) to obtain information from very degraded samples even when the classical techniques (sequencing, STR analysis) did not work; and (iv) to authenticate the results obtained from ancient DNA extracts, by comparing different markers.

Concerning the Y SNPs, the results were more variable. However, it is also a very promising technique that allows a rapid definition of the Y-chromosome haplogroups. For ancient DNA studies, Y haplogroups are rarely analyzed, and, thus, we can imagine that such a technique could be used to explore the history of paternal lineages in ancient populations.

To conclude, the iPLEX[®] method is very promising for population-based studies, and in particular for ancient populations. Indeed, this technique is a good alternative to new technologies such as complete genome sequencing,

since it optimizes the ancient sample resource and results in a greater degree of discrimination than that obtained by techniques currently routinely used in ancient DNA studies.

The authors have declared no conflict of interest.

5 References

- [1] Griffin, T. J., Tang, W., Smith, L. M., *Nat. Biotechnol.* 1997, 15, 1368–1372.
- [2] Tost, J., Gut, I. G., *Mass Spectrom. Rev.* 2002, 21, 388–418.
- [3] Hall, T. A., Budowle, B., Jiang, Y., Blyn, L., Eshoo, M., Sannes-Lowery, K. A., Sampath, R., Drader, J. J., Hannis, J. C., Harrell, P., Samant, V., White, N., Ecker, D. J., Hofstadler, S. A., *Anal. Biochem.* 2005, 344, 53–69.
- [4] Hall, T. A., Sannes-Lowery, K. A., McCurdy, L. D., Fisher, C., Anderson, T., Henthorne, A., Gioeni, L., Budowle, B., Hofstadler, S. A., *Anal. Chem.* 2009, 81, 7515–7526.
- [5] Sauer, S., Gut, I. G., *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 2002, 782, 73–87.
- [6] Petkovski, E., Keyser-Tracqui, C., Hienne, R., Ludes, B., *J. Forensic Sci.* 2005, 50, 535–541.
- [7] Griffin, T. J., Smith, L. M., *Trends Biotechnol.* 2000, 18, 77–84.
- [8] Jurinke, C., van den Boom, D., Jacob, A., Tang, K., Worl, R., Koster, H., *Anal. Biochem.* 1996, 237, 174–181.
- [9] Cullinan, A., Cantor, C., *Pharmacogenomics* 2008, 9, 1211–1215.

- [10] Paracchini, S., Arredi, B., Chalk, R., Tyler-Smith, C., *Nucleic Acids Res.* 2002, **30**, e27.
- [11] Gabriel, S., Ziaugra, L., Tabbaa, D., *Curr. Protoc. Hum. Genet.* 2009, Chapter 2, Unit 2.12.
- [12] Cerezo, M., Cerny, V., Carracedo, A., Salas, A., *Electrophoresis* 2009, **30**, 3665–3673.
- [13] Cooper, A., Poinar, H. N., *Science* 2000, **289**, 1139.
- [14] Willerslev, E., Cooper, A., *Proc. Biol. Sci.* 2005, **272**, 3–16.
- [15] González, A. M., García, O., Larruga, J. M., Cabrera, V. M., *BMC Genomics* 2006, **23**, 124–131.
- [16] Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V. G., Montagnon, D., Reis, T., Ludes, B., *Hum. Genet.* 2009, **126**, 395–410.
- [17] Mardis, E. R., *Annu. Rev. Genomics Hum. Genet.* 2008, **9**, 387–402.
- [18] Maricic, T., Pääbo, S., *Biotechniques* 2009, **46**, 51–54.
- [19] Shapiro, B., Hofreiter, M., *Bioessays* 2010, **32**, 388–391.
- [20] Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., Pääbo, S., *Science* 2010, **7**, 723–728.
- [21] Nielsen, A. E., in: Berberian, E. E., Nielsen, A. E. (Eds), *Historia Argentina Prehispánica, Tomo 1*, Editorial Brujas, Córdoba 2001, pp. 171–264.
- [22] Cocilovo, J. A., Varela, H. H., Valdano, S. G., in: Berberian, E. E., Nielsen, A. E. (Eds), *Historia Argentina Prehispánica, Tomo 1*, Editorial Brujas, Córdoba 2001, pp. 265–288.
- [23] Bailliet, G., Ramallo, V., Muzzio, M., Garcia, A., Santos, M. R., Alfaro, E. L., Dipierri, J. E., Salceda, S., Carnese, F. R., Bravi, C. M., Bianchi, N. O., Demarchi, D. A., *Am. J. Phys. Anthropol.* 2009, **140**, 578–582.
- [24] Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q. P., Woodward, S. R., Salas, A., Torroni, A., Bandelt, H. J., *PLoS One* 2008, **3**, e1764.
- [25] Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., Bravi, C. M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E. K., Fedorova, S. A., Golubenko, M. V., Stepanov, V. A., Gubina, M. A., Zhadanov, S. I., Ossipova, L. P., Damba, L., Voevoda, M. I., Dipierri, J. E., Villems, R., Malhi, R. S., *PLoS One* 2007, **2**, e829.
- [26] van Oven, M., Kayser, M., *Hum. Mutat.* 2009, **30**, 386–394.
- [27] Carnese, F. R., Mendisco, F., Keyser, C., Dejean, C. B., Dugoujon, J. M., Bravi, C. M., Ludes, B., Crubezy, E., *Am. J. Phys. Anthropol.* 2010, **141**, 452–462.
- [28] Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., Hammer, M. F., *Genome Res.* 2008, **18**, 830–838.
- [29] Zegura, S. L., Karafet, T. M., Zhivotovsky, L. A., Hammer, M. F., *Mol. Biol. Evol.* 2004, **21**, 164–175.
- [30] The Y Chromosome Consortium, *Genome Res.* 2002, **12**, 339–348.
- [31] Oeth, P., del Mistro, G., Marnellos, G., Shi, T., van den Boom, D., *Methods Mol. Biol.* 2009, **43**, 578–587.
- [32] Schaeffeler, E., Zanger, U. M., Eichelbaum, M., Asante-Poku, S., Shin, J. G., Schwab, M., *Clin. Chem.* 2008, **54**, 1637–1647.
- [33] Hughes-Stamm, S. R., Ashton, K. J., van Daal, A., *Int. J. Legal Med.* 2010, DOI 10.1007/s00414-010-0455-3.
- [34] Shinoda, K., Adachi, N., Guillen, S., Shimada, I., *Am. J. Phys. Anthropol.* 2006, **131**, 98–107.
- [35] Fuselli, S., Taorazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D., Pettener, D., *Mol. Biol. Evol.* 2003, **20**, 1682–1691.
- [36] Moraga, M., Santoro, C. M., Standen, V. G., Carvallo, P., Rothhammer, F., *Am. J. Phys. Anthropol.* 2005, **127**, 170–181.
- [37] Bortolini, M.-C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. K., Bau, C. H. D., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., Hill, K., Hurtado, A. M., Castro-de-Guerra, D., Torres, M. M., Groot, H., Michalski, R., Nymadawa, P., Bedoya, G., Bradman, N., Labuda, D., Ruiz-Linarez, A., *Am. J. Hum. Genet.* 2003, **73**, 524–539.
- [38] Blanco-Verea, A., Jaime, J. C., Brion, M., Carracedo, A., *Forensic Sci. Int. Genet.* 2010, **4**, 187–193.

First application of the Investigator DIPplex indels typing kit for the analysis of ancient DNA samples

Hollard C, Mendisco F, Keyser C, Crubézy E, Ludes B

Forensic Science International: Genetics Supplement Series. 2011 Dec; 3 (1): 393–94



First application of the Investigator DIPplex indels typing kit for the analysis of ancient DNA samples

C. Hollard^{a,*}, F. Mendisco^{a,b}, C. Keyser^{a,b}, E. Crubézy^b, B. Ludes^a

^a Laboratoire d'Anthropologie Moléculaire, Institut de Médecine Légale, Université de Strasbourg, CNRS UMR 5288, 11 rue Humann, 67085 Strasbourg Cedex, France

^b Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), Université Paul Sabatier, CNRS UMR 5288, Toulouse, France

ARTICLE INFO

Article history:

Received 29 August 2011

Accepted 15 September 2011

Keywords:

Ancient DNA

Insertion/deletion polymorphism

Kinship

Multiplex PCR

ABSTRACT

Kinship testing and identification of human remains in forensic and anthropological fields is commonly based on the analysis of short tandem repeats (STRs) or more seldom single nucleotide polymorphisms (SNPs). In this study, another type of genetic variations, insertion/deletion polymorphisms (indels) was used through the DIPplex indels typing kit (Qiagen) for such purpose. The objective was to evaluate the performance of the assay for the analysis of ancient DNA samples as well as the informativeness of the kit on kinship investigation. Although some limitations have been noticed, this work shows that indels are well suited for the analysis of degraded samples and that they might be an interesting strategy in addition to STRs typing.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Nowadays, short tandem repeats (STRs) typing is considered as the method of choice in forensic science. The benefits of STRs are multiples and well-known. However, the analysis of STRs from highly degraded DNA sometime failed to resolve informative profile due to the relatively large size of amplicons. In order to solve this problem, miniSTR kit (Applied Biosystems) has been developed to reduce amplicon size. Nevertheless, the power of discrimination is low (10^{-11}) compared to STR typing kits (10^{-18} at least). As an alternative to STRs, SNPs have been successfully employed for highly degraded DNA due to the small size of amplicons (less than 150 bp), although their use in forensic kinship investigation is discussed [1]. Another form of genetic variation on the human genome has recently received an increasing focus: insertion/deletion polymorphisms (indels). Their characteristics seem particularly interesting in forensic analysis: (i) their wide distribution throughout the genome [2]; (ii) the small size of amplicons which facilitates multiplex amplifications in highly degraded DNA; (iii) the possibility of typing them with a simple dye labeling electrophoretic approach.

In the present study, indels were analyzed through the Investigator DIPplex indels typing kit (Qiagen), in order to evaluate the performance of the kit for the analysis of ancient DNA samples and the informativeness of this kit for kinship investigation in anthropological fields where the trio mother–father–son is not

always available. First tested on modern samples, the assay was secondarily applied on ancient specimens selected due to their putative kinship deduced from classical markers previously analyzed such as autosomal and Y-chromosomal STRs and mitochondrial DNA.

2. Materials and methods

Four modern DNA samples (buccal swabs), for which an appropriate informed consent was obtained, and nine ancient samples (teeth) from different age and location were analyzed (Table 1). Previous genetic analyses (unpublished observation) to estimate close parental relationship were carried as published in Lacan et al. [3].

The indels analyses were performed using the Investigator DIPplex kit (Qiagen) allowing the amplification of 30 indels distributed over 19 autosomes as well as the Amelogenin locus. The manufacturer's protocol was used except the final volume of reaction that was divided by two (13 μ l). For ancient DNA samples, the following modifications were applied: 5 μ l of DNA was used for the PCR reaction and the number of PCR cycles was increased from 30 to 32.

3. Results and discussion

The results obtained on modern DNA samples are successful since the 30 indels loci included in this kit were amplified. Nonetheless, the analysis of the DNA profiles revealed that the intensity of the peak differs depending on the dye used. Indeed, the intensity was always the lowest with the BTR dye (Fig. 1).

* Corresponding author. Tel.: +33 03 68 85 33 48; fax: +33 03 68 85 33 62.

E-mail address: clemence.hollard@etu.unistra.fr (C. Hollard).

Table 1

Data on the ancient specimens tested.

Sample name	Date	Putative kinship
I1	1700–1800 A.D	Parent offspring
I2		
O1		
O2	1800–1900 A.D	Parent offspring
L9	1000–1500 A.D	Parent offspring
L20		
N1		
N2	3000 B.C	N1/N2: full sibling; N2/N3: parent offspring
N3		

Moreover, these samples showed an imbalanced allele for heterozygous genotype which could be a significant limit for the use of this kit in DNA mixture analysis.

Regarding ancient DNA samples, a clear correlation between DNA quantity in the samples and the number of amplified loci was observed. The sensitivity of this assay allows obtaining workable profiles from degraded samples, although, as described by the manufacturer, the profiles are often partial and the results inconsistent, with less than 100 pg of DNA. In addition, the loss of signal intensity with the high molecular weight markers from standard STRs typing kit is reduced due to the small size of amplicons (<150 bp).

Then, the objective was to study if the results of indels typing obtained from ancient DNA samples confirmed the putative kinships deduced from previous autosomal/Y-STRs typing and mtDNA sequencing. In three cases (I1 and I2, O1 and O2, L9 and L20), the two individuals under study seemed linked by a parent/child relationship (Table 1). Indels typing gave confirmation to the hypothesis in two cases (O1/O2 and L9/L20) since the two linked individuals shared a common allele at each indels locus typed. The hypothesis is challenged in one case (I1 and I2). Indeed, the two individuals did not share any allele for one indels locus. This result allows excluding a parent/child relationship, reinforcing the hypothesis given by the statistical analysis conducted with the DNA view software [4] ($LR = 104 \times 10^{12}$). For these individuals, thanks to the analysis of the different types of markers used, the brother/sister link appears as the privileged hypothesis.

For the oldest samples (N1, N2 and N3), especially N1 and N2, the amount of DNA (<60 pg) was very limiting for the analysis. The results varied between experiments and the typing of some markers has failed. The increase of the number of tested loci did not allow resolving the determination of parentage relationship in this case.

In conclusion, although this study as just focused on a few number of archaeological samples, it provides a first idea of the value of the investigator DIPplex for the analysis of ancient DNA

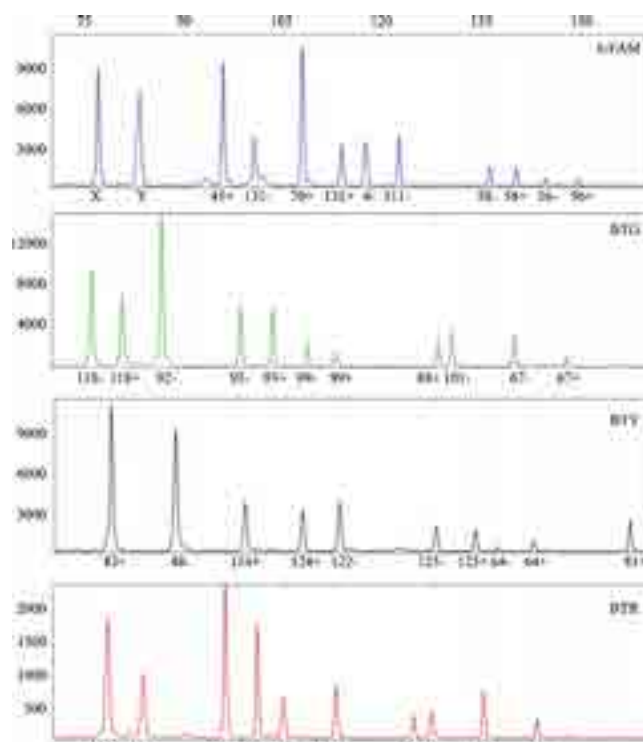


Fig. 1. Electropherogram obtained with the Investigator DIPplex kit using 332 pg DNA sample from an ancient specimen.

samples. Indels typing was successful and highly informative for most of our samples. This kit in combination with standard markers improves the discriminatory power and provides potential supplementation for anthropological studies.

Conflict of interest

None.

References

- [1] A. Amorim, L. Pereira, Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs, *Forensic Sci. Int.* 150 (2005) 17–21.
- [2] J.L. Weber, D. David, J. Heil, et al., Human diallelic insertion/deletion polymorphisms, *Am. J. Hum. Genet.* 71 (2002) 854–862.
- [3] M. Lacan, C. Keyser, F.X. Ricaut, et al., Ancient DNA reveals male diffusion through the Neolithic Mediterranean route, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 9788–9791.
- [4] C.H. Brenner, Symbolic kinship program, *Genetics* 145 (1997) 535–542.

**Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental
and ancestry informative markers**

Hollard C, Keyser C, Giscard P-H, Tsagaan T, Bayarkhuu N, Bemmman J, Crubézy E,
Ludes B

Soumis à Forensic Science International: Genetics

**Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental and ancestry
informative markers**

Abstract

The Altai Mountains have been a long term boundary zone between the Eurasian Steppe populations and South and East Asian populations. To disentangle some of the historical population movements in this area, 14 ancient human specimens excavated in the westernmost part of the Mongolian Altai were studied. Thirteen of them were dated from the Middle to the End of the Bronze Age and one of them to the Eneolithic period. The environmental conditions encountered in this region led to the good preservation of DNA in the human remains. Therefore, a multi-markers approach was adopted for the genetic analysis of identity, ancestry and phenotype markers. Mitochondrial DNA analyses revealed that the ancient Altaians studied carried both Western (H, U, T) and Eastern (A, C, D) Eurasian lineages. In the same way, the patrilineal gene pool revealed the presence of different haplogroups (Q1a2a1-L54, R1a1a1b2-Z93 and C), probably marking different origins for the male paternal lineages. To go further in the search of the origin of these ancient specimens, phenotypical characters (ie: hair and eye color) were determined. For this purpose, we adapted the HIRISplex assay recently described to MALDI-TOF mass spectrometry. In addition, some ancestry informative markers were analyzed with this assay. The results revealed mixed phenotypes among this group confirming the probable admixed ancestry of the studied Altaian population at the Middle Bronze Age. The good results obtained from ancient DNA samples suggest that this approach might be relevant for forensic casework too.

Keywords: ancient DNA, admixture, Bronze Age, Altai Mountains

Introduction

The Eurasian Steppe stretches from Moldavia through Ukraine to Siberia as far as the Baikal area, therefore connecting Europe, Central Asia, China, South Asia and the Middle East. This steppe experienced many ancient population movements from diverse geographical sources over an extended time period. Some archeological evidence suggests that populations spread eastward from Western Eurasia across this steppe during the Eneolithic and Bronze Age [1]. Later, during the Middle Ages, Turks then Mongols expanded to the west through this same area [2-4]. These multiple population movements explain that present-day populations from this region present both Eastern and Western Eurasian genetic features [5-7]. Direct analysis of the genetic pool of ancient specimens can give important information about the settlement of this area. Indeed, such analysis can show which genetic lineages were present at one place at a particularly period, a result that can't be obtained using modern DNA data. In this context a previous ancient DNA study [8] suggested that some individuals among South Siberian populations carried European features during the Bronze and Iron Ages.

In its East Asian part, the southern border of the Eurasian Steppe is delimited by the Altai Mountains. This 2000 kilometer long mountain range with many peaks between 3000 and 4000 meters high extend southward in Mongolia and in the Xinjiang province of China. In the present work, we studied two necropolises in the Mongolian Altai Mountains: the site of Tsagaan Asga (on the Chinese border) and the site of Takhilgat Uzuur-5 (on the Siberian border) (figure 1). These sites were dated from the Middle Bronze Age, as were subjects of European origin found in the Siberian side of the Eurasian steppe. The cold environment in this area probably helped the recovery of 14 well preserved skeletons excavated in important Kurgans (fig 1).

For each ancient specimen, two uniparentally inherited systems (mitochondrial and Y-chromosome) were studied to trace the maternal and paternal lineages. Autosomal markers were analyzed to determine the molecular sex of the individuals as well as their kinship. They were also used to determine the physical appearance and the biological ancestry of these ancient individuals. Indeed, a SNP assay was developed from a forensic test to trace the history of this population. The aims of this study were (i) to study the funeral recruitment of these two old necropolises, and (ii) to evaluate the biological boundary between West and East Eurasian populations at this period. Moreover, this study highlights the usefulness of a multi marker approach in ancient DNA studies where other studies often only analyze mitochondrial DNA.

Materials and Methods

Samples

Altaiian Middle Bronze age cultures are poorly known. Some necropolises from the Mongolian and Chinese Altai, and sometimes from South Siberia, usually display stone circles with a kurgan at the center. Under this kurgan is a pit filled with large blocks in which rests a buried body positioned on his side and displaying no associated artifact. Russian archeologists sometimes attribute these tombs to the so-called "Mongun Taiga" culture, but Mongolian researchers described them under the term "Sagsay tombs". Radiocarbon dating of these sites (table 1) showed that these kurgans cover a large chronological period considered in this area as the Middle to Late Bronze Age. Nevertheless a more ancient tomb, dated to the Eneolithic, was found at the Takhilgat

Uzuur-5 site. The use of these necropolises over a large period and the prominence of the graves suggests a specific sampling, which could be limited to members of the elite. We studied skeletal remains of 14 individuals from two different partially excavated archeological sites located in the Mongolian Altai. Seven skeletons (TA4, TA8, TA9, TA10, TA11, TA12 and TA14) were excavated from the site of Tsagaan Asga in the Tsengel-Sagsai districts near China by a French –Mongolian archeological mission. Seven other skeletons (TU15, TU17, TU18, TU25, TU32, TU31 and TU34) were excavated in 2011 in the framework of the “Eneolithic and Bronze Age Mongolian Altai” project, by a Mongolian-German expedition, closer to the Siberian border in the site of Takhilgat Uzuur-5. All were adult skeletons except TU18 who was a child. All the skeletal remains used for the analyses were teeth except for one specimen (TA4) for which only a long bone fragment was available.

DNA extraction

The first step consisted in sample decontamination. The teeth were cleaned with bleach, rinsed with ultrapure water and exposed to UV light on each side during 20min. The bone fragment was decontaminated by elimination of the outer surface with a sanding machine. Bone fragment and teeth were then powdered in a grinder mill under liquid nitrogen (6870 FREEZER/MILL[®], SPEX SamplePrep). The DNA extraction and purification were performed from 200mg of the teeth or bone powder obtained as previously described [9].

Autosomal STR analysis

Fifteen autosomal STRs and the sex-determining marker amelogenin were amplified using the AmpFLSTR[®] Identifiler Plus[™] kit (Life Technologies). PCR conditions described by the manufacturer were respected, except for the number of PCR cycles which was increased from 28 to 34. Capillary electrophoresis was run on the 3500 Genetic Analyzer (Life Technologies) and data analysis was performed with the GeneMapper[™] 4.1 software (Life Technologies).

Mitochondrial DNA analysis

The mitochondrial HV1 region (16024-16383) was amplified in two overlapping fragments and sequenced as described in [10]. Haplogroup assignment deduced from the HV1 haplotype from the phylotree system (www.phylotree.org) was confirmed by the typing of SNPs of the mitochondrial coding region. Thirteen SNPs at nucleotide positions 663, 13263, 1715, 5178, 752, 1438, 13086, 7028, 14766, 13708, 4917, 12308, and 9090 were selected, combined in one multiplex PCR reaction and typed using the iPLEX[®] gold technology (Sequenom) as described in [9]. The HV1 haplotypes were quality controlled by EMPOP [11] and the software EMMA [12] was used to check the haplogroup assignment and determine the mitochondrial haplogroups from HV1 sequences for TU31 and TA12 samples.

Y chromosomal analysis

Seventeen Y-chromosomal STRs were amplified using the AmpFLSTR[®] Yfiler[™] PCR amplification kit (Life Technologies) from DNA of the ancient male samples (n=9). The experimental conditions were those recommended by the manufacturer except that the number of PCR cycles was increased from 30 to 34. The capillary electrophoresis and analysis conditions were the same than those described for autosomal STR analysis. To define the Y-chromosomal haplogroups, two sets of Y-chromosomal SNPs were selected according to the

ISOGG nomenclature. The first set was composed of 13 Y-chromosomal SNPs defining the R haplogroup and some R subclades (M207, M173, M420, M516, M17, M458, M434, Z280, Z93, M343, P297, M73 and M269). The second set was constituted of 14 Y-chromosomal SNPs characterizing the main Asian Y haplogroups found in the Altaian populations (M214, M175, M231, Tat, M128, P43, M242, M346, L54, M3, M120, M130, M217 and M174). SNP typing was performed by using the iPLEX[®] gold technology (Sequenom) as described in [9]. A phylogenetic tree of the Y-chromosomal SNPs analyzed in this study for the Q and R haplogroups is shown in figure 2.

Autosomal SNP typing

The 24 SNPs (N29insA, rs11547464, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs1805009, Y1520CH, rs2228479, rs1110400, rs28777, rs16891982, rs12821256, rs4959270, rs12203592, rs1042602, rs1800407, rs2402130, rs12913832, rs2378249, rs12896399, rs1393350 and rs683) described in [13] were used to predict the eye and hair color of ancient individuals. Six SNPs (rs11547464, rs885479, rs1805008, rs1805007, Y1520CH and rs1110400) present on the MC1R gene were analyzed by direct sequencing due to the proximity between them. The 18 remaining SNPs were all typed in one multiplex reaction. Genotyping was done with the Mass Array platform and the iPLEX[®] gold kit (Sequenom). The eye and hair colors were predicted by means of the interactive spreadsheet tool developed with the HirisPlex assay and available as supplemental data in [13].

Four additional ancestry-informative SNPs (rs1545397, rs2814778, rs1426654 and rs3827760) were added to the multiplex reaction described in the previous section, which finally contained 22 autosomal SNPs. They were selected due to their high frequency of different alleles in distinct populations (Asian, European and African) and their usefulness ascertained in the recent AIM (Ancestry Informative Marker) panels published [14-16]. The SNP rs16891982 used in the HirisPlex panel was used as ancestry marker too. From the genotypes obtained from these five AIMs, the biogeographical ancestry of each specimen was evaluated using the STRUCTURE 2.3.4 software [17]. Three populations of reference consisting of 556 individuals were used: Asian (Han Chinese in Beijing, China; Japanese in Tokyo, Japan; Southern Han Chinese), African (Yoruba in Ibadan, Nigeria; Luhya in Webuye, Kenya) and European (Utah Residents (CEPH) with Northern and Western European ancestry). The genetic data were obtained from the ensembl browser (www.ensembl.org). The program was first run without population information to ensure that the pre-defined populations are in rough agreement with the genetic information.

Data analysis

A median Joining network for the Q1a2a1-L54 haplogroup was generated with the Network 4.6 software (Fluxus Technology Ltd) from the ancient Altaian Mongolian (aAM) data and those published in [6, 18]. Fifteen Y-STRs (DYS385a/b loci were deleted) were used to generate the network. Microsatellite loci were weighted according to the observed STR allelic variances as described in [19]. The dataset was post-processed with the Maximum Parsimony option.

The small number of ancient specimens studied cannot be considered as a population. Nevertheless, to analyze the genetic affinities of these ancient specimens with present-day populations and ancient groups, each mitochondrial and Y-chromosomal haplotype obtained from the ancient specimens studied was searched for in

personal databases. The mtDNA database compiles 21332 published HV1 sequences from European and Asian populations. The NRY database comprises 31759 haplotypes published in the literature. A database compiling 306 HV1 sequences obtained from ancient Eurasian individuals published in the literature was used for comparison. To increase the possibilities of finding our haplotypes in current populations, a search in the EMPOP database (<http://empop.org/>) [11] for the mitochondrial data was also performed.

Precautions taken to avoid contamination

Strict precautions were taken to avoid contamination. The pre PCR and post PCR laboratories are located on different floors. The pre PCR steps were performed in a laboratory strictly dedicated to ancient DNA with over pressure and UV light irradiation. Between each manipulation, benches and material were cleaned with bleach, ultrapure water and DNA contamination removal solution and placed under UV light. Sterile aerosol resistant tips and sterile plastics were used. The experimenters wore appropriate equipment: overshoes, facial mask, mobcap, lab coat and gloves. Moreover, genetic profiling of all the people directly in contact with the samples was undertaken and compared with the DNA profile of the ancient specimens. For each sample, independent multiple DNA extractions and PCR amplifications were performed and negative controls were included for each experiment.

Results

Autosomal STRs

Eleven of the fourteen DNA samples analyzed gave full unique STR profiles. For one sample (TU18) a very partial profile was established and for 2 samples (TU25, TU32) no amplifiable product was obtained (table S1). These results could be explained by the mediocre state of the skeletons' conservation. Typing of the amelogenin locus established that 9 ancient specimens were males and 2 were females (table 1). For the other 3 specimens, the sex determination was uncertain (TU18) or impossible (TU25 and TU32) with our experimental setup. Comparison in pairs of complete profiles revealed no first degree relatives.

Mitochondrial DNA

The HV1 sequences obtained for 12 out of the 14 specimens revealed 12 different haplotypes. Two DNA samples (TU25 and TU32) appeared too degraded to give a HV1 sequence. The successful typing (100%) of coding mitochondrial SNPs allowed the classification of 10 haplotypes out of the 12 in 6 different haplogroups (table 1 and table S2). Two samples (TA12 and TU31) could not be assigned to particular haplogroups due to the selected SNPs included in the multiplex reaction. Nevertheless the search for these haplotypes in our database showed that haplotype 16037G, 16311C and 16319A (TU31) was found in one Finnish individual. This individual belongs to the R* super-haplogroup but could not be affiliated to one specific haplogroup based on HV1 and HV2 sequences [20]. Among the 10 individuals affiliated to one haplogroup, 4 fall in West Eurasian haplogroups (U, T, H) and 6 fall in East Eurasian haplogroups (A, C, D). To assess the present distribution of these 12 haplotypes, a search for their occurrence among actual populations was undertaken (table S3, S3bis). Nine out of the 12 haplotypes were found in the databases. Among the western lineages, two samples belonged to the H haplogroup. One is the CRS-sequence, and one presents three substitutions at position 16183C, 16189C

and 16356C. This haplotype was only found in 5 Polish and one Portuguese individuals. Haplogroup U4 (16356C) was carried by one ancient specimen. This haplotype is very frequent among West Eurasian and South Siberian populations. Haplogroup T was represented by the subhaplogroup T1a with the following variants: 16093C-16126C-16163G-16186T-16189C-16294T which is mainly found in Central Asia and in the Near East. Regarding the Eastern lineages, most mtDNA haplotypes (4/6) belong to the D haplogroup. One sample presents the basal haplotype (16223T-16362C), which displays a wide geographic distribution in Northern, Eastern and Central Asian populations. Haplotype 16223T-16311C-16362C, which has just one supplemental substitution, is uncommon in Asian populations. The two other D haplotypes are very rare since only one (TA14) was found in a present-day Korean and for the other (TA9), no exact match was found in our database. The haplotype assigned to haplogroup A was shared mainly with South Siberians, as well as East and Central Asians. Finally, the haplotype sharing analyses found no exact match in comparative modern-day populations for two haplotypes of the aAM TA10 and TA12. The current distribution of the mtDNA haplotypes found in aAM is shown in figure 3A. Central Asia and south Siberia are the overlapping areas where both the eastern and western Eurasian lineages are currently found.

NRY-chromosome

From autosomal DNA analyses, 9 samples were determined as males and were subsequently used for seventeen Y-chromosomal STRs typing. Successful results (complete 17 Y-STRs haplotype) were obtained for all the samples and correspond to six different haplotypes (table S4). Y-chromosomal haplogroup assignment was done from the typing of NRY-chromosomal SNPs. The six haplotypes were grouped in four haplogroups: R1a1a1b2-Z93 (TA4, TA10, TA14, TU34), Q1a2a1-L54 (TA8, TA11, TA12), Q-M242 (TU31) and C-M130 (TU17) (table 1 and table S5). Each haplotype was searched for in our own database. As no exact match was found from the 17 STR loci, we used the minimal haplotype based on 9 STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393). Results are presented in table S6. Among the R1a1a1b2-Z93 haplogroup, one haplotype (TA14/ TU34) was found to present a wide distribution (East Europe, South Siberia and South Asia). The other R1a1a haplotype carried by TA4 and TA10 was less frequently found in the present-day population database (5 individuals: 4 Europeans and 1 individual from China). As the M434, M458, Z280, and Z93 markers, defining sub-clades of the R1a1a haplogroup, have been recently described [21, 22], few population studies deal with this information for R1a1a individuals. Nevertheless a search for our haplotypes among the data available in these two publications showed that our R1a1a1b2 haplotypes matched only R1a1a*(x M458) in Underhill et al., and R1a1a1b2-Z93 in Pamjav et al.

Among the individuals belonging to haplogroup Q, from the minimal haplotype, only one haplotype (TA11/TA12) was shared in our database with 2 individuals (one Tibetan and one Even). To evaluate the phylogenetic relationships of the aAM Q-L54 haplotypes with the current Q-L54 haplotypes found in Asia and America, a median joining network was built (fig S1). The 3 aAM individuals are on the same branch than the South Siberian populations.

Affinities with ancient Eurasian groups

The 12 mitochondrial haplotypes were searched for in a personal database compiling haplotypes of ancient Eurasian individuals published in the literature. As shown in table S3, no exact match was found for 5 aAM

haplotypes (TA9, TA10, TA14, TU15 and TU31) among ancient Eurasian individuals. In the 7 other ones, West and East Eurasian haplotypes were shared with different ancient populations (table S3 and figure 3B). One East Eurasian lineage (16223T-16362C) was present in a Bronze Age group in the west Liao-River valley (China) [23]. Apart from this individual, the East Eurasian lineages matched mainly those from Iron Age Asian populations like the Xiongnu [24, 25], some ancient nomads of Pengyang in China [26] and some ancient inhabitants of the Xinjiang province [27]. Nevertheless, the basal D haplotype carried by aAM TA11 was found in a North Eastern European group dated 3500 BP [28]. Unlike Eastern lineages, the West Eurasian lineages (T and U4) were present at least since the Bronze Age in ancient South Siberians [8] and ancient Kazakhs [29]. The CRS sequence presented a wide distribution in ancient Eurasian populations since it was found from the Bronze Age in populations of South Siberia [8, 30, 31], Central Asia [29] and the Xinjiang province [27, 32]. Finally, one haplotype (16305G - 16311C), which was not found in present-day populations, was found in one ancient Xinjiang individual dated back to 2500 years [27].

Concerning the Y-chromosomal affinities with ancient Eurasian populations, the cross population comparison is very weak as few studies treat this information. Nevertheless, 4 aAM shared the R1a1a haplogroup with ancient south Siberians [8], ancient inhabitants of the Tarim Basin [33] and with one individual found in a Xiongnu cemetery in Northeast Mongolia [25]. The C haplogroup was carried by one Bronze Age South Siberian [8]. Finally, the Q-M242 haplogroup was found in 4 ancient inhabitants of Inner Mongolia dated back to 2500 years [26].

Externally visible characteristics and ancestry

To obtain complementary information about the biogeographical origin of the aAM specimens, externally visible characteristics such as eye and hair color were inferred on the basis of the Hiris Plex assay. The assay, originally based on the SNaPshot technology, was adapted to be used with a mass spectrometry technology. It was first applied to modern samples in order to validate the assignments (table S7). Once validated, the assay was applied to the ancient samples. The phenotype and ancestry of the aAM specimens are indicated in table 1 (genotypes are available in table S8). Eye colors were predicted with confidence due to the low number of SNPs (6) necessary for color prediction and the strong weight of the rs12913832 for color determination. It seems that a majority of the ancient individuals had brown eyes (9/10) and one aAM had blue eyes. The hair colour results have to be interpreted with more caution due to the absence of result for the rs1805009 and the high number of SNPs typed that increase the possibilities of allele drop out. This can turn heterozygote into homozygote genotypes. Nevertheless, 1aAM seemed to have dark blond/brown hair, 1 other had brown hair, 3 had dark brown hair and 5 probably had dark brown/black hair.

Since the self-reported ancestry of all the modern samples was in accordance with biogeographical ancestry (table S7) inferred by the Structure software, the test was considered reliable for ancient samples. The typing of the five AIMs gave us a glimpse into the ancestry of each ancient individual (fig S2, table S9). Three individuals (TA4, TA9 and TA11) seem to have an admixed ancestry with both Asian and European components. The other ancient specimens seem to have a major proportion of European ancestry based on the 5 SNPs used in the assay.

Authenticity of the results

Dealing with ancient DNA required drastic precautions to avoid contamination with exogenous DNA. Despite the fact that not all published authenticity criteria were fulfilled, the methodological strategy adopted supports the authenticity of the results. Indeed, all the negative controls for DNA extraction and PCR were blank. The genetic profiles obtained were all different from each other and from the profiles of the personnel who had been in contact with the ancient remains. The results were obtained from different extractions and different amplifications without discordance. There is a clear inverse correlation between the length of the amplified products and the intensity of the peaks during STR analysis. The typing of SNPs for haplogroup assignment is concordant with the results of the mitochondrial HV1 haplotype and Y-STR haplotype.

Discussion

Biological structure of the funeral sites

No first degree relatives were determined from the aSTR profiles of the buried individuals, a result consistent with the dispersion of the graves over several centuries. Moreover, the high mitochondrial diversity observed (haplotype diversity value of 1) reinforces this statement. Sex typing of the ancient remains revealed that there are more male graves than female graves (9 ancient individuals were males and 2 were females). This imbalanced sex ratio in the buried group could be explained by particular funeral recruitment due to cultural or social practices as described in other Siberian cultures [34]. Indeed, in the Tsagaan Asga site, two paternal lineages were shared by two individuals (TA11/TA12 and TA4/TA10) suggesting that males belonging to the same paternal lineages were buried in the same site at different times. However, the diversity of the paternal lineages suggests that this necropolis was not exclusively dedicated to one lineage for hundreds of years. The high mitochondrial diversity combined with the reduced Y-chromosomal diversity observed within these groups could also suggest a limited paternal gene flow that could be linked to specific cultural practices such as patrilocality. Nevertheless, the low number of studied individuals could also bias our results and therefore lead to misinterpretation.

Biogeographical origin

Concerning the maternal lineages, our data revealed a mixed gene pool of Western (U, T, H) and Eastern (A, C, D) Eurasian lineages. Among the Western lineages, three haplotypes found in aAM were already described in Bronze Age populations of Kazakhstan [29] and South Siberia [8], attesting a genetic flow from the West. This western Eurasian genetic pool may have been added to the East Eurasian lineages probably still present in the Altai Mountains at the Bronze Age. Indeed, Bronze Age individuals of the Mongolian Altai Mountains studied in [35] carried only haplogroup D. Interestingly, one eastern lineage was shared with one Iron Age individual in Kazakhstan [29], which is consistent with a putative genetic flow of eastern lineages to the west side of the Altai Mountains between the Bronze and Iron Ages. Indeed, in Siberian Eurasian steppes and in Kazakhstan, a shift from a West Eurasian genetic pool at the Bronze Age to a mixture of West and East Eurasian lineages at the Iron Age was observed in [8, 29].

The data obtained from the analysis of the paternal lineages of the ancient male individuals suggests a similar scenario. Different haplogroups were found (Q1a2a1-L54, R1a1a1b2-Z93, and C-M130). These haplogroups have been proposed as signatures of different migratory events [36]. As an example, indigenous American

lineages derived mostly from the Q haplogroup and numerous modern population studies have underlined a link between native American Q lineages and the South Siberian populations lineages [6, 37-40]. From the improvement of the phylogeny one marker (L54) was characterized and presented as the marker from which M3, the major Amerindian lineage, arose [6]. This marker is very frequent in South Siberian populations, since among the males belonging to haplogroup Q, more than 90% in the Tuva republic and 100% in South Altai populations belong to the Q-L54 lineage [18]. In our work, the three males belonging to the Q haplogroup, belong to the Q-L54 subclade. This result suggests the presence of an ancient paternal Siberian component among the aAM. Four ancient males belong to the R1a1a1b2-Z93 haplogroup. R1a1a-M17 (Z93 markers being not yet described) was found in ancient south Siberian and Tarim Basin populations. The origin and dispersal of this haplogroup remain controversial given its wide distribution throughout Eurasia [22, 36]. Indeed, the distribution of the minimal Y-STR haplotype of our four R1a1a1b2 lineages suggests a link with Central/Eastern Europe, the Near East and South Asia. From the recent phylogeny improvement, a more detailed scenario seems to emerge. Indeed, new sub-haplogroups occurring at informative frequencies have been described. These subclades are defined by markers for which the derived allele was found in a restricted geographic area, thus dividing the R1a1a haplogroup into a European branch (M458, Z280) and an Asian branch (Z93). As described in [21], both forms of alleles of the Z93 and Z280 markers overlapped in Central Asia. In this context, the origin of R1a1a could be somewhere between South Asia and Eastern Europe (the Eurasian Steppes and the Middle East being presented as potential candidates) and from this area could have split up into two different branches. Although only four R1a1a-M17 individuals were analyzed in this study, all of them were found to belong to Z93 lineages. Therefore, the presence of the Z93 marker in aAM suggests a potential western paternal origin for these ancient individuals. Moreover, the R1a1a-M17 haplogroup was found in Northeast Mongolia in a Xiongnu cemetery [25], suggesting a westward migration of this haplogroup to Northeast Asia 2000 years ago.

As these individuals seem to have diverse origins, eye and hair color were determined. Thanks to the well preserved DNA molecules in the ancient human remains studied it was possible to genotype autosomal SNPs. Nevertheless, working with ancient DNA molecules always mean working with small quantities and often degraded DNA that can lead in some case to allele drop-out and so to a false homozygous state. This problem is avoided when working with haploid markers but is present with diploid systems. The sensitivity of the method we developed is not discussed here, but in order to avoid false genotyping, the ancient DNA analyses were performed twice. The results suggest different phenotypes. Indeed, two individuals seem to present light hair colour, with blue eyes for one of them which could suggest an „Europoid“ phenotype. The other ones present a brown-eyed and brown/black-haired phenotype. Nevertheless, the addition of 5 AIMs seems to show that most of the ancient individuals (7/10) have a major proportion of European ancestry. Obviously, our sample of five AIMs is small compared to the number of markers described in the literature and used in published panels [14]. Increasing the number of AIMs allows a better inference of individual ancestry. Nevertheless, working with ancient DNA limits the amount of DNA available, which is why we chose only five SNPs.

Biological Population history, a first hypothesis

The increasing number of ancient DNA studies available allowed us to perform cross “population” comparisons. In summary, shared haplotype studies highlighted putative links between our ancient individuals and ancient specimens from South Siberia, the Altai Mountains and the Xinjiang province mainly from the Y-chromosomal

data (R1a1a haplogroup) but from the mitochondrial data too. These observations suggest a common genetic pool which may have diffused in these areas. In the Altai Mountains, the West Eurasian genetic pool would mix with East Eurasian individuals still present in this region. Indeed, although partially excavated, the two necropolises studied delivered mixed groups with both East and West Eurasian genetic components. This admixture was notably present at the individual level since some males have mitochondrial and Y chromosomal haplotypes from different geographic origins. Among the ancient specimens studied, two males buried in two different sites carried the same Y-chromosomal haplotype (17 Y-STR). This example shows that males belonging to the same paternal clan were not restricted to one site, but that population movement and genetic flow occurred over time in this region.

In the Takhilgat-Uzuur 5 site, one tomb dated to 2700 BC and contemporary to the Afanasievo culture was excavated. The genetic analyses revealed that the ancient individual was a male who was already admixed as he carried both Western and Eastern genetic components. As one individual is not enough to give conclusions about the Afanasievo culture, a larger study is of course necessary.

Interest of the multi-markers approach

The first interest of this approach is to study close relative kinship and try to reconstruct a partial genealogy of the buried subjects and understand the funeral sampling. Indeed, working with ancient DNA to trace kinship is more challenging so that analyzing the three conventional systems (mtDNA, NRY-chromosome and aSTRs) adds weight to the results (ie. EG necropolis [24]). The second interest of a multi-markers study is to trace back different events. Indeed, the analysis of just one system like the mitochondrial DNA allows only the reconstruction of the maternal history of the individuals. Sexually asymmetric demographic processes can however lead to different histories for maternal and paternal lineages. Thus the analysis of both mitochondrial DNA and NRY-chromosome limits the bias in the reconstruction of population histories.

Besides its interest in molecular anthropology, the strategy employed could be useful in a forensic context. On one hand, SNPs are good candidates in forensic casework when STR amplification failed due to degradation of the DNA molecules. The genotyping method used allows to analyze several SNPs in one reaction from a low volume of DNA which is very interesting for challenging samples. On the other hand, the use of a multi-marker strategy consisting of the analysis of mitochondrial, Y-chromosome and autosomal DNA can give valuable information in events like mass disasters (terrorist attacks, airplane crashes, collective genocides or natural disasters) where people can be from various biogeographical origins and where the human remains are usually severely damaged. This strategy can also be very helpful for the identification of long dead people or ancient war victims (where only bones and teeth remain)[41] for whom reference samples can be missing or for whom no information regarding the sex, geographic origin or physical appearance is available. Molecular data can be added to those given by the physical anthropologist to help the identification of the deceased people. In challenging cases, the multiplication of data obtained is a strategy to facilitate and speed up the identification of the body.

Conclusion

In this study, we demonstrated the usefulness of a multi-markers approach (mitochondrial DNA, Y-chromosome and autosomes) for both ancient DNA studies and forensic casework applications. Indeed, the results obtained from these different genetic systems are consistent with the presence of an “admixed” population from both East and West Eurasia in the Altai Mountains at the middle Bronze Age and allowed us to propose a first hypothesis on the peopling of this area that will need to be compared to a broader synthesis taking into account dozens of ancient individuals.

Acknowledgments

We thank the program HUMAD from IPEV (Institut polaire français Paul Emile Victor), the Gerda Henkel Foundation and l’Institut des Déserts et des Steppes.

References

1. Kuzmina EE: **The Prehistory of the Silk Road**. *University of Pennsylvania Press* 2008.
2. Di Cosmo N, Frank AJ, Golden PB: **The Cambridge history of Inner Asia: The Chinggisid Age**. *Cambridge University Press* 2009.
3. Grousset R: **L'empire des steppes**. *Editions Payot* 1989.
4. Sinor D: **The Cambridge History of Early Inner Asia**. *Cambridge University Press* 1990, **1**.
5. Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M *et al*: **Diversity of mitochondrial DNA lineages in South Siberia**. *Annals of human genetics* 2003, **67**(Pt 5):391-411.
6. Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, Schurr TG: **Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians**. *American journal of human genetics* 2012, **90**(2):229-246.
7. Kharkov VN, Stepanov VA, Medvedeva OF, Spiridonova MG, Voevoda MI, Tadinova VN, Puzyrev VP: **Gene Pool Differences Between Northern and Southern Altaians Inferred from the Data on Y-Chromosomal Haplogroups**. *Russian Journal of Genetics* 2007, **43**(5):551-562.
8. Keyser C, Bouakaze C, Crubezy E, Nikolaev VG, Montagnon D, Reis T, Ludes B: **Ancient DNA provides new insights into the history of south Siberian Kurgan people**. *Human genetics* 2009, **126**(3):395-410.

- 385 9. Mendisco F, Keyser C, Hollard C, Seldes V, Nielsen AE, Crubezy E, Ludes B: **Application of the**
386 **iPLEX(TM) Gold SNP genotyping method for the analysis of Amerindian ancient DNA**
387 **samples: Benefits for ancient population studies.** *Electrophoresis* 2011.
- 388 10. Carnese FR, Mendisco F, Keyser C, Dejean CB, Dugoujon JM, Bravi CM, Ludes B, Crubezy E:
389 **Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina).**
390 *American journal of physical anthropology* 2010, **141**(3):452-462.
- 391 11. Parson W, Dur A: **EMPOP--a forensic mtDNA database.** *Forensic science international Genetics*
392 2007, **1**(2):88-92.
- 393 12. Rock AW, Dur A, van Oven M, Parson W: **Concept for estimating mitochondrial DNA**
394 **haplogroups using a maximum likelihood approach (EMMA).** *Forensic science international*
395 *Genetics* 2013, **7**(6):601-609.
- 396 13. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M: **The**
397 **HLrisPlex system for simultaneous prediction of hair and eye colour from DNA.** *Forensic*
398 *science international Genetics* 2013, **7**(1):98-115.
- 399 14. Fondevila M, Phillips C, Santos C, Freire Aradas A, Vallone PM, Butler JM, Lareu MV, Carracedo A:
400 **Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard**
401 **reference sample genotypes and extended population studies.** *Forensic science international*
402 *Genetics* 2013, **7**(1):63-74.
- 403 15. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen
404 PK, Belmont JW *et al*: **Ancestry informative marker sets for determining continental origin**
405 **and admixture proportions in common populations in America.** *Human mutation* 2009,
406 **30**(1):69-78.
- 407 16. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Calaza M, de Cal MC,
408 Ballard D, Lareu MV *et al*: **Inferring ancestral origin using a single multiplex assay of**
409 **ancestry-informative marker SNPs.** *Forensic science international Genetics* 2007, **1**(3-4):273-
410 280.
- 411 17. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus**
412 **genotype data.** *Genetics* 2000, **155**(2):945-959.
- 413 18. Regueiro M, Alvarez J, Rowold D, Herrera RJ: **On the origins, rapid expansion and genetic**
414 **diversity of Native Americans from hunting-gatherers to agriculturalists.** *American journal of*
415 *physical anthropology* 2013, **150**(3):333-348.
- 416 19. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi
417 SQ: **Y-chromosomal DNA variation in Pakistan.** *American journal of human genetics* 2002,
418 **70**(5):1107-1124.

- 419 20. Hedman M, Brandstatter A, Pimenoff V, Sistonen P, Palo JU, Parson W, Sajantila A: **Finnish**
420 **mitochondrial DNA HVS-I and HVS-II population data.** *Forensic science international* 2007,
421 **172(2-3):171-178.**
- 422 21. Pamjav H, Feher T, Nemeth E, Padar Z: **Brief communication: new Y-chromosome binary**
423 **markers improve phylogenetic resolution within haplogroup R1a1.** *American journal of*
424 *physical anthropology* 2012, **149(4):611-615.**
- 425 22. Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CE, Semino
426 O, Battaglia V *et al*: **Separating the post-Glacial coancestry of European and Asian Y**
427 **chromosomes within haplogroup R1a.** *European journal of human genetics : EJHG* 2010,
428 **18(4):479-484.**
- 429 23. Li H, Zhao X, Zhao Y, Li C, Si D, Zhou H, Cui Y: **Genetic characteristics and migration history of a**
430 **bronze culture population in the West Liao-River valley revealed by ancient DNA.** *Journal of*
431 *human genetics* 2011, **56(12):815-822.**
- 432 24. Keyser-Tracqui C, Crubezy E, Ludes B: **Nuclear and mitochondrial DNA analysis of a 2,000-**
433 **year-old necropolis in the Egyin Gol Valley of Mongolia.** *American journal of human genetics*
434 2003, **73(2):247-260.**
- 435 25. Kim K, Brenner CH, Mair VH, Lee KH, Kim JH, Gelegdorj E, Batbold N, Song YC, Yun HW, Chang EJ *et*
436 *al*: **A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast**
437 **Mongolia.** *American journal of physical anthropology* 2010, **142(3):429-440.**
- 438 26. Zhao YB, Li HJ, Cai DW, Li CX, Zhang QC, Zhu H, Zhou H: **Ancient DNA from nomads in 2500-**
439 **year-old archeological sites of Pengyang, China.** *J Hum Genet* 2010, **55(4):215-218.**
- 440 27. Zhang F, Xu Z, Tan J, Sun Y, Xu B, Li S, Zhao X, Zhou H, Gong G, Zhang J *et al*: **Prehistorical East-**
441 **West admixture of maternal lineages in a 2,500-year-old population in Xinjiang.** *Am J Phys*
442 *Anthropol* 2010, **142(2):314-320.**
- 443 28. Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshel S, Zaporozhchenko
444 V, Gronenborn D, Moiseyev V, Kolpakov E *et al*: **Ancient DNA reveals prehistoric gene-flow from**
445 **siberia in the complex human population history of North East Europe.** *PLoS Genet* 2013,
446 **9(2):e1003296.**
- 447 29. Lalueza-Fox C, Sampietro ML, Gilbert MT, Castri L, Facchini F, Pettener D, Bertranpetit J:
448 **Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central**
449 **Asians.** *Proc Biol Sci* 2004, **271(1542):941-947.**
- 450 30. Chikisheva TA, Gubina MA, Kulikov IV, Karafet TM, Voevoda MI, Romaschenko AG: **A paleogenetic**
451 **study of the prehistoric populations of the Altai.** *Archeology, Ethnology & Anthropology of*
452 *Eurasia* 2007, **4:130-142.**

31. Clisson I, Keyser C, Francfort HP, Crubezy E, Samashev Z, Ludes B: **Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC).** *Int J Legal Med* 2002, **116**(5):304-308.
32. Gao S, Cui Y, Yang Y, Duan R, Abuduresule I, Mair VH, Zhu H, Zhou H: **Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China.** *Sci China C Life Sci* 2008, **51**(3):205-213.
33. Li C, Li H, Cui Y, Xie C, Cai D, Li W, Mair VH, Xu Z, Zhang Q, Abuduresule I *et al*: **Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age.** *BMC biology* 2010, **8**:15.
34. Crubezy E, Alekeev A: **Chamane: Kyys, jeune fille des glaces.** Editions Errance 2007.
35. Gonzalez-Ruiz M, Santos C, Jordana X, Simon M, Lalueza-Fox C, Gigli E, Aluja MP, Malgosa A: **Tracing the origin of the east-west population admixture in the Altai region (Central Asia).** *PloS one* 2012, **7**(11):e48904.
36. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S *et al*: **The Eurasian heartland: a continental perspective on Y-chromosome diversity.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(18):10244-10249.
37. Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT *et al*: **Y-chromosome evidence for differing ancient demographic histories in the Americas.** *American journal of human genetics* 2003, **73**(3):524-539.
38. Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P *et al*: **Ancestral Asian source(s) of new world Y-chromosome founder haplotypes.** *American journal of human genetics* 1999, **64**(3):817-831.
39. Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ: **The central Siberian origin for native American Y chromosomes.** *American journal of human genetics* 1999, **64**(2):619-628.
40. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF: **High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas.** *Molecular biology and evolution* 2004, **21**(1):164-175.
41. Draus-Barini J, Walsh S, Pospiech E, Kupiec T, Glab H, Branicki W, Kayser M: **Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains.** *Investigative genetics* 2013, **4**(1):3.

486

FIGURE CAPTIONS

487 Fig.1

488 A. Geographic location of studied sites (1: Tsagaan Asga; 2: Takhilgat Uzuur-5). B. Graves and human remains
489 of two individuals excavated on the site of Tsagaan Asga.

490 Fig.2

491 Phylogenetic tree of the Y-chromosomal SNPs analyzed in this study for the Q and R haplogroups. Haplogroup
492 names are according to the ISOGG website (February 2014). The lengths of the branches have no value in term
493 of phylogenetic distance.

494 Fig.3

495 Distribution of mitochondrial HV1 haplotypes of aAM (A) in current populations and (B) in ancient populations.
496 Circles and triangles denote West and East Eurasian lineages respectively.

497 Fig.S1

498 Median joining network based on Q-L54 haplotypes found in current Siberian (Northeast Siberia, South Altaian
499 and Tuva Republic), American (Cakchikel and Yucatan) and in the aAM.

500 Fig. S2

501 Triangle plot built with Structure software representing the putative ancestry of aAM.

502

Figure 1



Figure 2

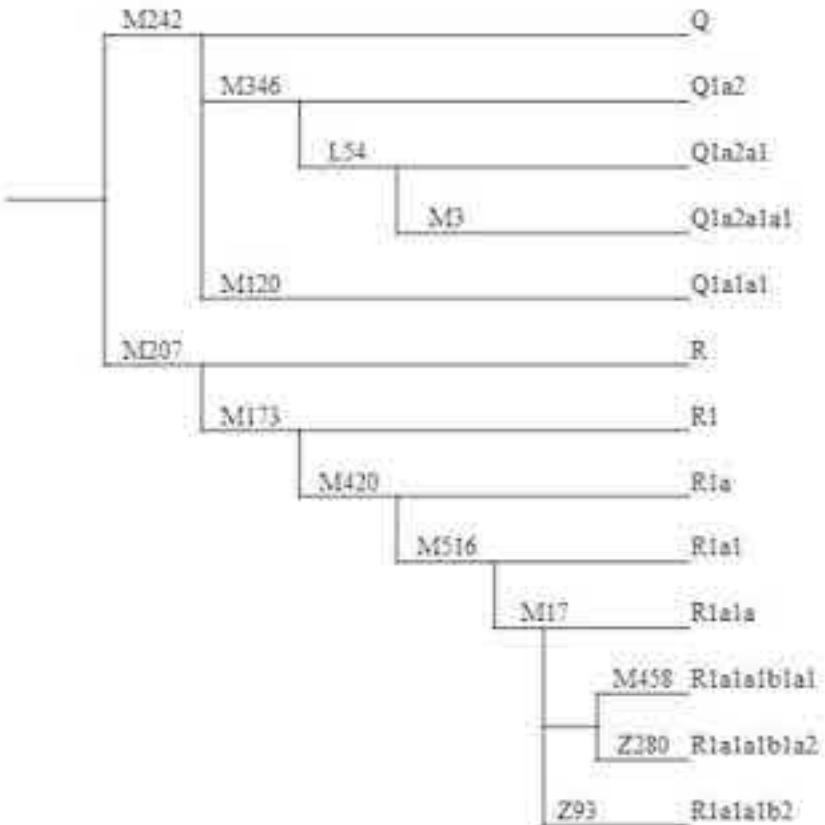


Figure 3

A



B

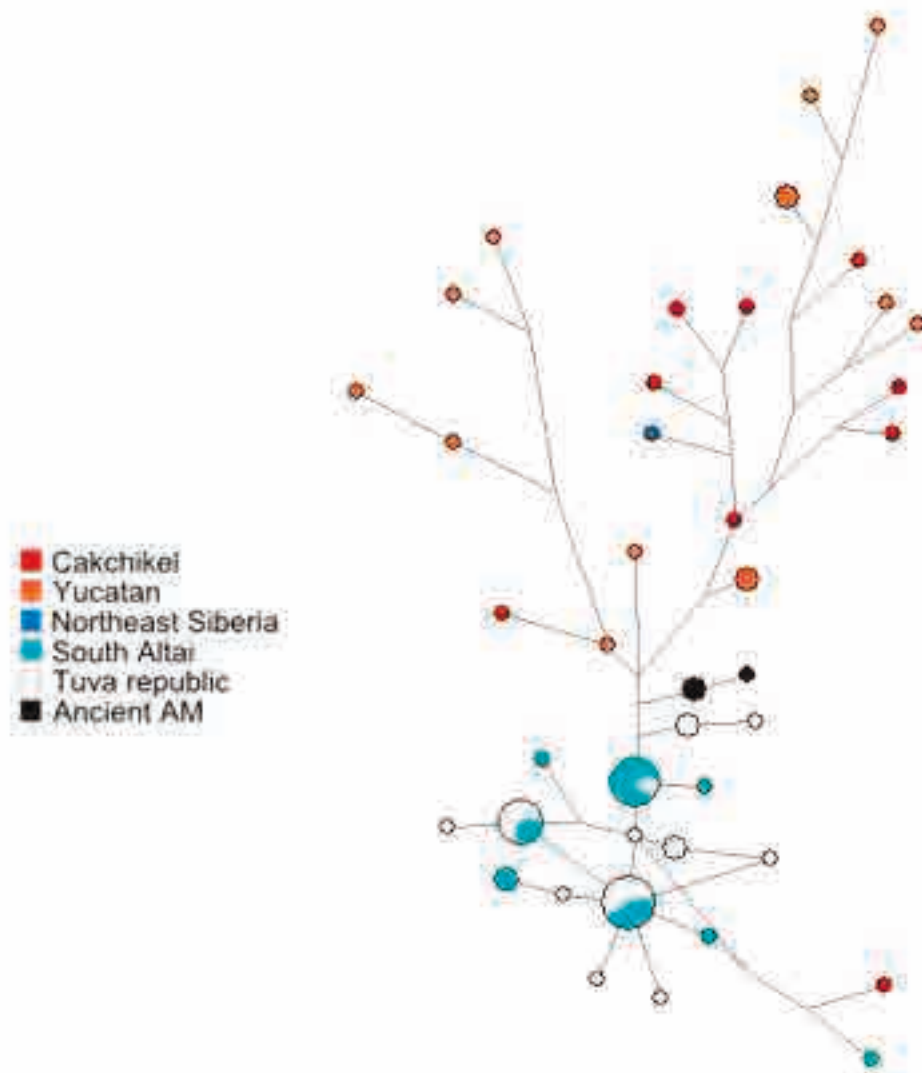


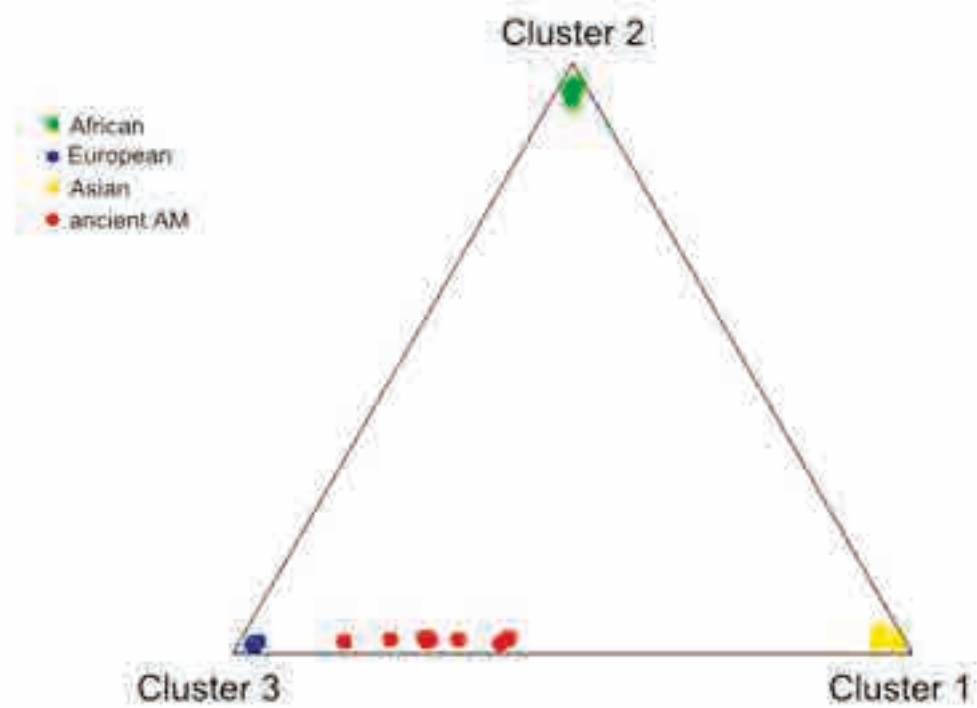
Table 1

Sample	¹⁴ C Datation (years BC)	Molecular Sexe	Mt HV1 haplotype (- 16000)	HV1 haplogroup	Mt SNP haplogroup	Y haplogroup	The most probable eye color	The most probable hair color	Ancestry inferred from STRUCTURE		
TA4	1371 ± 41	M	093C 126C 163G 186T 189C 294T	T1a	T	R1a1a1b2-Z93	Brown	Black / dark brown	Asian	African	European
TA8		M	356C	U4	U	Q1a2a1-L54	/	/	0.381	0.005	0.615
TA9		F	086C 092C 164G 182C 183C 185T 189C 193del 223T 266T 362C	D5	D5a2		Brown	Black / dark brown	0.21	0.01	0.78
TA10		M	093C 129A 223T 260T 263C 298C 327T	C4a1	C4a1	R1a1a1b2-Z93	Brown	Dark brown	0.391	0.009	0.6
TA11	1358 ± 42	M	223T 362C	D	D	Q1a2a1-L54	Brown	Black / dark brown	0.273	0.009	0.718
TA12		M	305G 311C	H7e*	?	Q1a2a1-L54	Brown	Black / dark brown	0.391	0.014	0.595
TA14		M	092C 129A 223T 362C	D2	D	R1a1a1b2-Z93	Blue	Brown	0.26	0.014	0.725
TU15		F	183C 189C 356C	H1b	H		Brown	Dark brown	0.27	0.009	0.721
TU17		M	223T 290T 319A 362C	A	A	C-M130	Brown	Dark brown	0.314	0.008	0.677
TU18		?	223T 311C 362C	D	D		/	/	0.267	0.008	0.725
TU31		M	037G 311C 319A	R1b1*	?	Q-M242	Brown	Black / dark brown	/	/	/
TU34		M	CRS	H/U	H	R1a1a1b2-Z93	Brown	Dark blond / brown	0.265	0.014	0.720
	1011 ± 56								0.141	0.007	0.851

Table 1: Consensus of the results obtained from each ancient Altaian Mongolian studied.

Dash denotes an absence of result. * These haplogroups were determined with the software EMMA.





Sample	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	Amel	D5S818	FGA
Research team																
I1	14/16	30/32.2	10/11	10/12	16/17	9.3/9.3	11/13	9/12	16/25	13.2/15	18/19	8/11	12/15	XX	11/12	21/21
I2	12/14	29/29	9/11	12/12	14/18	9/9.3	12/12	11/11	17/19	13/13	15/15	11/11	14/17	XX	11/11	21/23
I3	8/15	27/31.2	8/10	11/12	14/18	6/9.3	12/12	9/11	22/24	13/14	17/19	8/8	14/15	XY	9/13	20/20
I4	8/13	30/31.2	10/11	12/12	16/17	6/8	12/12	9/12	25/25	13/14	17/17	11/11	11/18	XX	11/13	22.2/23
I5	14/14	29/32.2	7/10	11/11	14/16	7/8	11/12	9/12	19/19	14/16.2	14/14	9/9	16/20	XX	11/13	23/24
I6	10/13	30/32.2	9/12	10/11	15/17	7/9.3	11/14	10/12	24/25	14/15	16/17	8/11	16/19	XY	11/12	20/20
Ancient MA																
TA4	12/13	30/33.2	8/11	11/12	15/16	9/9	11/13	10/13	20/25	15/15.2	17/18	8/11	14/14	XY	11/13	19/23
TA8	12/16	31/31	8/10	11/14	15/16	6/9.3	8/10	10/11	20/23	11/15.2	18/19	8/11	17/20	XY	11/12	22/24
TA9	12/17	29/31.2	11/11	11/13	15/17	7/9	10/12	12/14	20/23	14/15	16/17	8/9	12/17	XX	11/13	23/24
TA10	11/12	29/30	8/10	10/11	17/18	8/9	8/11	9/12	20/(22)	11/14	16/18	8/8	19/19	XY	11/13	22/23
TA11	12/14	30/32.2	8/11	10/11	14/15	7/9.3	10/14	10/12	19/19	13/14.2	14/14	11/12	14/18	XY	11/12	22/26
TA12	13/13	30/30	9/11	10/11	15/17	9.3/9.3	11/14	11/11	17/19	14/14	17/18	9/12	14/18	XY	12/13	22/24
TA14	13/14	33.2/33.2	9/10	11/14	15/15	6/9	11/12	12/14	19/25	15/15	18/20	8/8	13/15	XY	9/11	23/24
TU15	14/16	30/30	11/12	(11)	15/17	6/9.3	10/13	9/11	(19)	13/14	17/17	9/11	17/20	XX	12/13	24/25
TU17	14/14	32.2/32.2	8/11	10/11	17/18	7/9.3	10/12	10/11	17/18	14.2/15.2	17/18	11	16/17	XY	12/12	21/25
TU18	10	-	-	-	-	(9.3)	-	-	-	(13)	-	-	-	(X)	(12)	-
TU25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TU31	10/12	29/33.2	12/12	10/11	14/14	8/9.3	11/12	9/12	18/23	14/15.2	14/17	8/8	13/14	XY	11/12	21/25
TU32	-	-	-	-	-	(7)	-	-	-	-	-	-	-	-	-	-
TU34	11/13	30.2/32.2	8/11	10/10	16/16	6/7	9/12	11/12	(19/25)	13/13.2	18/20	8/11	13/19	XY	10/11	19/20

Table S1: Consensus genetic profiles of the 14 ancient specimens studied and DNA profiles of the members of the research team who have been in contact with the samples.

Alleles in brackets were just observed once. Dash denotes locus for which amplification was null. Consensuses were built after three independent amplifications from three different DNA extracts.

Sample	HV1 haplotype (16024-16383)	HV1 hg	A-663	C-13263	C4a1-1715	D-5178	D5a-752	D5a2-1438	HV-14766	H-7028	T-4917	U-12308	J-13708	Z-9090	SNP hg
TA4	16093C 16126C 16163G 16186T 16189C 16294T	T1a	A	A	C	C	C	G	T	T	G	A	G	T	T
TA8	16356C	U4	A	A	C	C	C	G	T	T	A	G	G	T	U
TA9	16086C 16092C 16164G 16182C 16183C 16185T 16189C 16193del 16223T 16266T 16362C	D5	A	A	C	A	T	A	T	T	A	A	G	T	D5a2
TA10	16093C 16129A 16223T 16260T 16263C 16298C 16327T	C4a1	A	G	T	C	C	G	T	T	A	A	G	T	C4a1
TA11	16223T 16362C	D	A	A	C	A	C	G	T	T	A	A	G	T	D
TA12	16305G 16311C	?	A	A	C	C	C	G	T	T	A	A	G	T	?
TA14	16092C 16129A 16223T 16362C	D2	A	A	C	A	C	G	T	T	A	A	G	T	D
TU15	16183C 16189C 16356C	H1b	A	A	C	C	C	G	C	C	A	A	G	T	H
TU17	16223T 16290T 16319A 16362C	A4	G	A	C	C	C	G	T	T	A	A	G	T	A
TU18	16223T 16311C 16362C	D	A	A	C	A	C	G	T	T	A	A	G	T	D
TU31	16037G 16311C 16319A	?	A	A	C	C	C	G	T	T	A	A	G	T	?
TU34	CRS	H/U	A	A	C	C	C	G	C	C	A	A	G	T	H

Table S2: Mitochondrial HV1 haplotype and haplogroup inferred for each ancient specimen.

Allele in bold is variant for the concerned loci. Question mark corresponds to undetermined haplogroup. del = Deletion.

Sample	Shared haplotype in present-day populations	Shared haplotype in ancient populations
TA4	7 ind: 2 Kazakhs ¹ , 1 Kirghiz ¹ , 1 Slovakian ² , 1 Iranian ³ , 1 Sri Lankan ³ , 1 Turkish ⁴	1 ind: 1 aKaz ⁵
TA8	127 ind: 27 Tubalars ⁶⁻⁸ , 11 Latvians ⁹ , 8 Mansis ^{10; 11} , 7 Khantys ¹¹ , 7 Polishs ^{12; 13} , 5 Bosnian-Herzegovians ¹⁴ , 4 Altaians ¹⁵ , 4 Slovaks ² , 4 Portuguese ^{16; 17} , 4 Indians ³ , 3 Sri Lankans ³ , 3 Iranians ³ , 3 Georgians ¹⁸ , 3 Uighurs (Kazakhstan) ¹ , 3 North Caucasus ^{4; 19} , 3 Italians ²⁰ , 2 Irishs ²¹ , 2 Finnishs ^{22; 23} , 2 Macedonians ²⁴ , 2 Germans ^{25; 26} , 2 Tatars ²⁷ , 2 Western Russians ^{12; 13} , 2 Turkishs ⁴ , 2 Palestinians ⁴ , 1 Khakassian ¹⁵ , 1 Chelkan ⁶ , 1 Altaian Kazakh ²⁸ , 1 Slovenian ¹⁴ , 1 British ²⁹ , 1 Armenian ⁴ , 1 Azeri ⁴ , 1 Iraqi ⁴ , 1 Tuvanian ⁷ , 1 Albanian ²⁴ , 1 Romanian ²⁴ , 1 Kazakh ¹	3 ind: 2 aSouth Siberians ³⁰ , 1aPo ³¹
TA9	No match	No match
TA10	No match	No match
TA11	153 ind: 26 Chinese ³²⁻³⁸ , 15 Tubalars ⁶⁻⁸ , 15 Altaians Kizhis ^{6; 39} , 14 Koreans ⁴⁰ , 14 Japaneses ⁴¹ , 6 Kirghizs ¹ , 6 Indians ³ , 6 Evens ⁸ , 5 Sojots ¹⁵ , 5 Buryats ^{15; 42} , 4 Tuvinians ^{7; 15; 43} , 4 Pathans ⁴⁴ , 3 Kazakhs ¹ , 3 Uighurs (Kazakhstan) ¹ , 3 Ewenks ^{7; 37} , 3 Chelkans ⁶ , 2 Altaians ¹⁵ , 2 Khakassians ¹⁵ , 2 Todjins ¹⁵ , 2 Altaians Kazakhs ²⁸ , 2 Mongols ⁴⁵ , 2 Yakuts ⁴³ , 2 Laotians ⁴⁶ , 1 German ⁴⁷ , 1 Uighur (China) ³⁷ , 1 Turkish ⁴ , 1 North Caucasus ⁴ , 1 Ulchi ⁸ , 1 Georgian ¹⁸	12 ind: 5 Xiongnus ^{48; 49} , 1 aXinjiang ⁵⁰ , 3 aBOO ³¹ , 2 aChina ⁵¹ , 1aChina ⁵²
TA12	No match	1 ind: 1 aXinjiang ⁵⁰
TA14	1 ind: 1 Korean ⁴⁰	No match
TU15	6 ind: 5 Polishs ^{12; 13} , 1 Portuguese ¹⁷	No match
TU17	29 ind: 11 chineses ^{32-35; 37; 38} , 4 Altaians Kizhis ⁶ , 4 Altaians Kazakhs ²⁸ , 2 Khakassians ¹⁵ , 2 Koreans ⁴⁰ , 1 Kazakh ¹ , 1 Kirghiz ¹ , 1 Tuvanian ⁴² , 1 Todjin ¹⁵ , 1 Vietnamese ⁵³ , 1 Mongol ⁴⁵ , 1 Czech ⁵⁴ , 1 Tuvanian ⁴³ , 1 Uighur (China) ³⁷	4 ind: 3 Xiongnus ⁴⁸ , 1 aKaz ⁵
TU18	13 ind: 3 chineses ^{32; 37} , 3 Buryats ¹⁵ , 3 Mongols ⁴⁵ , 2 Altaians Kazakhs ²⁸ , 1 Kazakh ¹ , 1 Turkish ⁴	3 ind: 2 Xiongnus ^{48; 49} , 1 aChina ⁵²
TU31	1 ind: 1 Finnish ²²	No match
TU34	Very frequent	16 ind: 1 aAltaian ⁵⁵ , 2 Pazyryks ⁵⁵ , 1aBerel ⁵⁶ , 1 aTaklamakan ⁵⁷ , 2 aSouth Siberians ³⁰ , 1 aKaz ⁵ , 4 aKaz ⁵ , 3 aXinjiang ⁵⁰ , 2 a west Siberia ⁵⁸

Table S3: Distribution of aAM mitochondrial haplotypes within present-day and ancient populations. Results in bold correspond to pre-Iron Age samples.

1. Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al. (1998). Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 63, 1824-1838.
2. Malyarchuk, B.A., Perkova, M.A., Derenko, M.V., Vanecek, T., Lazur, J., and Gomolcak, P. (2008). Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet* 72, 228-240.
3. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26.
4. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67, 1251-1276.
5. Lalueza-Fox, C., Sampietro, M.L., Gilbert, M.T., Castri, L., Facchini, F., Pettener, D., and Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proceedings Biological sciences / The Royal Society* 271, 941-947.
6. Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T.G. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *American journal of human genetics* 90, 229-246.
7. Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., et al. (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69, 67-89.
8. Sukernik, R.I., Volodko, N.V., Mazunin, I.O., Eltsov, N.P., Dryomov, S.V., and Starikovskaya, E.B. (2012). Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *Am J Phys Anthropol* 148, 123-138.
9. Pliss, L., Tambets, K., Loogvali, E.L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., and Villems, R. (2006). Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet* 70, 439-458.
10. Derbeneva, O.A., Starikovskaya, E.B., Wallace, D.C., and Sukernik, R.I. (2002). Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70, 1009-1014.
11. Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., and Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur J Hum Genet* 16, 1254-1264.
12. Grzybowski, T., Malyarchuk, B.A., Derenko, M.V., Perkova, M.A., Bednarek, J., and Wozniak, M. (2007). Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet* 1, 141-147.

13. Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Wozniak, M., and Miscicka-Sliwka, D. (2002). Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66, 261-283.
14. Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Drobniak, K., and Miscicka-Sliwka, D. (2003). Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67, 412-425.
15. Derenko, M.V., Grzybowski, T., Malyarchuk, B.A., Dambueva, I.K., Denisova, G.A., Czarny, J., Dorzhu, C.M., Kakpakov, V.T., Miscicka-Sliwka, D., Wozniak, M., et al. (2003). Diversity of mitochondrial DNA lineages in South Siberia. *Annals of human genetics* 67, 391-411.
16. Gonzalez, A.M., Brehm, A., Perez, J.A., Maca-Meyer, N., Flores, C., and Cabrera, V.M. (2003). Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* 120, 391-404.
17. Pereira, L., Cunha, C., and Amorim, A. (2004). Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med* 118, 132-136.
18. Comas, D., Calafell, F., Bendukidze, N., Fananas, L., and Bertranpetit, J. (2000). Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol* 112, 5-16.
19. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74, 827-845.
20. Thomas, M.G., Barnes, I., Weale, M.E., Jones, A.L., Forster, P., Bradman, N., and Pramstaller, P.P. (2008). New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *Eur J Hum Genet* 16, 124-134.
21. McEvoy, B., Richards, M., Forster, P., and Bradley, D.G. (2004). The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* 75, 693-702.
22. Hedman, M., Brandstatter, A., Pimenoff, V., Sistonen, P., Palo, J.U., Parson, W., and Sajantila, A. (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci Int* 172, 171-178.
23. Meinila, M., Finnila, S., and Majamaa, K. (2001). Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered* 52, 160-170.
24. Bosch, E., Calafell, F., Gonzalez-Neira, A., Flaiz, C., Mateu, E., Scheil, H.G., Huckenbeck, W., Efremovska, L., Mikerezi, I., Xirotiris, N., et al. (2006). Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70, 459-487.
25. Brandstatter, A., Klein, R., Duftner, N., Wiegand, P., and Parson, W. (2006). Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int J Legal Med* 120, 310-314.

26. Lutz, S., Weisser, H.J., Heizmann, J., and Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111, 67-77.
27. Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., Dambueva, I., and Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PloS one* 5, e15214.
28. Gokcumen, O., Dulik, M.C., Pai, A.A., Zhadanov, S.I., Rubinstein, S., Osipova, L.P., Andreenkov, O.V., Tabikhanova, L.E., Gubina, M.A., Labuda, D., et al. (2008). Genetic variation in the enigmatic Altaian Kazakhs of South-Central Russia: insights into Turkic population history. *Am J Phys Anthropol* 136, 278-293.
29. Piercy, R., Sullivan, K.M., Benson, N., and Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106, 85-90.
30. Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V.G., Montagnon, D., Reis, T., and Ludes, B. (2009). Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Human genetics* 126, 395-410.
31. Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., Zaporozhchenko, V., Gronenborn, D., Moiseyev, V., Kolpakov, E., et al. (2013). Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet* 9, e1003296.
32. Chen, F., Wang, S.Y., Zhang, R.Z., Hu, Y.H., Gao, G.F., Liu, Y.H., and Kong, Q.P. (2008). Analysis of mitochondrial DNA polymorphisms in Guangdong Han Chinese. *Forensic Sci Int Genet* 2, 150-153.
33. Cui, Y., Li, C., Gao, S., Xie, C., and Zhou, H. (2010). Early Eurasian migration traces in the Tarim Basin revealed by mtDNA polymorphisms. *Am J Phys Anthropol* 142, 558-564.
34. Kong, Q.P., Yao, Y.G., Liu, M., Shen, S.P., Chen, C., Zhu, C.L., Palanichamy, M.G., and Zhang, Y.P. (2003). Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 113, 391-405.
35. Liu, C., Wang, S.Y., Zhao, M., Xu, Z.Y., Hu, Y.H., Chen, F., Zhang, R.Z., Gao, G.F., Yu, Y.S., and Kong, Q.P. (2011). Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. *Forensic Sci Int Genet* 5, e4-10.
36. Peng, M.S., He, J.D., Liu, H.X., and Zhang, Y.P. (2011). Tracing the legacy of the early Hainan Islanders--a perspective from mitochondrial DNA. *BMC Evol Biol* 11, 46.
37. Powell, G.T., Yang, H., Tyler-Smith, C., and Xue, Y. (2007). The population history of the Xibe in northern China: a comparison of autosomal, mtDNA and Y-chromosomal analyses of migration and gene flow. *Forensic Sci Int Genet* 1, 115-119.
38. Wen, B., Li, H., Gao, S., Mao, X., Gao, Y., Li, F., Zhang, F., He, Y., Dong, Y., Zhang, Y., et al. (2005). Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22, 725-734.

39. Phillips-Krawczak, C., Devor, E., Zlojutro, M., Moffat-Wilson, K., and Crawford, M.H. (2006). MtDNA variation in the Altai-Kizhi population of southern Siberia: a synthesis of genetic variation. *Hum Biol* 78, 477-494.
40. Jin, H.J., Kwak, K.D., Hong, S.B., Shin, D.J., Han, M.S., Tyler-Smith, C., and Kim, W. (2006). Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: an expanded Korean population database. *Forensic Sci Int* 158, 125-130.
41. Imaizumi, K., Parsons, T.J., Yoshino, M., and Holland, M.M. (2002). A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med* 116, 68-73.
42. Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., Shaikhaev, G.O., Dorzhu, C.M., Nimaev, D.D., and Zakharov, I.A. (2000). Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. *Hum Biol* 72, 945-973.
43. Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., and Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120, 334-353.
44. Rakha, A., Shin, K.J., Yoon, J.A., Kim, N.Y., Siddique, M.H., Yang, I.S., Yang, W.I., and Lee, H.Y. (2011). Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int J Legal Med* 125, 841-848.
45. Kolman, C.J., Sambuughin, N., and Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142, 1321-1334.
46. Bodner, M., Zimmermann, B., Rock, A., Kloss-Brandstatter, A., Horst, D., Horst, B., Sengchanh, S., Sanguansermisri, T., Horst, J., Kramer, T., et al. (2011). Southeast Asian diversity: first insights into the complex mtDNA structure of Laos. *BMC Evol Biol* 11, 49.
47. Pfeiffer, H., Brinkmann, B., Huhne, J., Rolf, B., Morris, A.A., Steighner, R., Holland, M.M., and Forster, P. (1999). Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112, 291-298.
48. Keyser-Tracqui, C., Crubezy, E., and 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.
49. Kim, K., Brenner, C.H., Mair, V.H., Lee, K.H., Kim, J.H., Gelegdorj, E., Batbold, N., Song, Y.C., Yun, H.W., Chang, E.J., et al. (2010). A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *Am J Phys Anthropol* 142, 429-440.
50. Zhang, F., Xu, Z., Tan, J., Sun, Y., Xu, B., Li, S., Zhao, X., Zhou, H., Gong, G., Zhang, J., et al. (2010). Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am J Phys Anthropol* 142, 314-320.
51. Li, H., Zhao, X., Zhao, Y., Li, C., Si, D., Zhou, H., and Cui, Y. (2011). Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J Hum Genet* 56, 815-822.

52. Zhao, Y.B., Li, H.J., Cai, D.W., Li, C.X., Zhang, Q.C., Zhu, H., and Zhou, H. (2010). Ancient DNA from nomads in 2500-year-old archeological sites of Pengyang, China. *J Hum Genet* 55, 215-218.
53. Irwin, J.A., Saunier, J.L., Strouss, K.M., Diegoli, T.M., Sturk, K.A., O'Callaghan, J.E., Paintner, C.D., Hohoff, C., Brinkmann, B., and Parsons, T.J. (2008). Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med* 122, 257-259.
54. Malyarchuk, B.A., Vanacek, T., Perkova, M.A., Derenko, M.V., and Sip, M. (2006). Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol* 78, 681-696.
55. Chikisheva, T.A., Gubina, M.A., Kulikov, I.V., Karafet, T.M., Voevoda, M.I., and Romaschenko, A.G. (2007). A paleogenetic study of the prehistoric populations of the Altai. *Archeology, Ethnology & Anthropology of Eurasia* 4, 130-142.
56. Clisson, I., Keyser, C., Francfort, H.P., Crubezy, E., Samashev, Z., and Ludes, B. (2002). Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC). *International journal of legal medicine* 116, 304-308.
57. Gao, S., Cui, Y., Yang, Y., Duan, R., Abuduresule, I., Mair, V.H., Zhu, H., and Zhou, H. (2008). Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China. *Sci China C Life Sci* 51, 205-213.
58. Sato, T., Razhev, D., Amano, T., and Masuda, R. (2011). Genetic features of ancient West Siberian people of the Middle Ages, revealed by mitochondrial DNA haplogroup analysis. *J Hum Genet* 56, 602-608.

Sample	Number of shared haplotype in present-day populations										
	Total	Admixed	African	Afro-asiatic	Afroeurasian	Central Asia	East Asia	Native American	North Asia	South Asia	West Eurasian
TA4	6	0	0	0	0	0	0	0	0	2	4
TA8	114	4	4	0	1	6	1	3	0	9	82
TA9	0	0	0	0	0	0	0	0	0	0	0
TA10	0	0	0	0	0	0	0	0	0	0	0
TA11	207	11	2	0	1	38	141	1	0	7	5
TA12	0	0	0	0	0	0	0	0	0	0	0
TA14	1	0	0	0	0	0	1	0	0	0	0
TU15	14	0	0	0	0	0	0	0	0	0	14
TU17	162	66	1	0	0	18	44	33	0	0	0
TU18	29	2	5	0	1	3	15	0	0	2	1
TU31	0	0	0	0	0	0	0	0	0	0	0
TU34											

very frequent

Table S3 bis: Distribution of aAM mitochondrial haplotypes within present-day populations.

Data obtained from the EMPOP database.

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATAH4	DYS437	DYS438	DYS448
Research team																
I3	15	12	24	28	17	14	11/14	14	11	13	23	13	11	14	12	18
I6	15	13	23	30	16	17	15/16	13	10	12	23	11	11	13	10	19
Ancient MA																
TA4	16	14	24	32	16	16	11/14	13	11	10	23	11	12	14	11	20
TA8	15	13	24	29	16	13	15/18	13	10	14	22	14	11	13	11	19
TA10	16	14	24	32	16	16	11/14	13	11	10	23	11	12	14	11	20
TA11	15	13	24	29	16	13	15/(17)	13	10	13	22	14	11	13	11	(19)
TA12	15	13	24	29	16	13	15/17	13	10	13	22	14	11	13	11	19
TA14	16	13	25	31	15	16	11/14	13	11	10	23	11	13	14	11	20
TU17	15	12	23	27	17	16	11/16	14	10	11	21	11	11	14	10	21
TU31	17	12	22	28	17	13	14/17	14	11	13	23	16	10	14	11	19
TU34	16	13	25	31	15	16	11/14	13	11	10	23	11	13	14	11	20

Table S4: Y-STR haplotypes determined from the ancient males specimens and the male members of the research team who have been in contact with the samples.

Alleles in brackets were just observed once. Loci in bold composed the minimal haplotype.

Sample	R-M207	R1-M173	R1a-M420	R1a1-M516	R1a1a-M17	M434	M458	Z280	Z93	R1b-M343	C-M130	C3-M217	Q-M242	Q1a2-M346	Q1a2a1-L54	Q1a2a1a1-M3	Y Haplogroup
I6	A	A	T	A	G	G	A	C	G	C	C	A	C	C	G	G	
TA4	G	C	A	G	DEL	G	A	C	A	C	ND	ND	ND	ND	ND	ND	R1a1a1b2-Z93
TA8	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	C	-	T	G	A	G	Q1a2a1-L54
TA10	G	-	A	G	DEL	G	A	C	A	C	ND	ND	ND	ND	ND	ND	R1a1a1b2-Z93
TA11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	C	-	T	G	A	G	Q1a2a1-L54
TA12	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	C	-	T	G	A	G	Q1a2a1-L54
TA14	G	C	A	G	DEL	G	A	C	A	C	ND	ND	ND	ND	ND	ND	R1a1a1b2-Z93
TU17	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	T	-	C	C	G	G	C-M130
TU31	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	C	-	T	C	G	G	Q-M242
TU34	G	C	A	G	DEL	G	A	C	A	C	ND	ND	ND	ND	ND	ND	R1a1a1b2-Z93

Table S5: Y-Chromosomal SNP typed from ancient male specimens and one researcher.

Allele in bold is variant for the concerned loci. Dash denotes loci for which null result was obtained. ND = Not determined.

Sample	minimal haplotype (19, 389I, 389II, 390, 391, 392, 393, 385a, 385b)	Shared minimal haplotype in present-day populations
TA4 / TA10	16 14 32 24 11 11 13 11 14	5 ind: 2 Polishs ⁵⁹ , 1 German ⁶⁰ , 1 Slovak ⁶¹ , 1 Mon-Khmer ⁶²
TA11 / TA12	13 13 29 24 10 14 13 15 17	2 Ind: 1 Tibetan ⁶³ , 1 Even ⁶⁴
TA8	13 13 29 24 10 14 13 15 18	No match
TA14 / TU34	16 13 31 25 11 11 13 11 14	33 ind: 5 Polishs ^{59; 65; 66} , 4 Western Russians ⁶⁷⁻⁶⁹ , 3 Germans ^{59; 60; 70} , 2 Slovaks ⁶¹ , 2 Latvians ^{59; 71} , 2 Afghans ^{72; 73} , 2 Tuvinians ⁶⁷ , 2 Indians ⁷⁴ , 1 Altaian-Kizhi ⁶⁷ , 1 Croatian ⁷⁵ , 1 Czech ⁷⁶ , 1 Eastern Russian ⁷⁷ , 1 Evenk ⁶⁷ , 1 Iranian ⁷⁸ , 1 Malay ⁷⁹ , 1 Swedish ⁸⁰ , 1 Turkish ⁸¹ , 1 Uighur ⁸² , 1 Yakut ⁸³
TU17	16 12 27 23 10 11 14 11 16	1 Ind: 1 Japanese ⁸⁴
TU31	13 12 28 22 11 16 14 14 17	No match

Table S6: Distribution of Y STR minimal haplotypes of aAM within present day-population

1. Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al. (1998). Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 63, 1824-1838.
2. Malyarchuk, B.A., Perkova, M.A., Derenko, M.V., Vanecek, T., Lazur, J., and Gomolcak, P. (2008). Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet* 72, 228-240.
3. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26.
4. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67, 1251-1276.
5. Lalueza-Fox, C., Sampietro, M.L., Gilbert, M.T., Castri, L., Facchini, F., Pettener, D., and Bertranpetit, J. (2004). Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proceedings Biological sciences / The Royal Society* 271, 941-947.
6. Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T.G. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *American journal of human genetics* 90, 229-246.
7. Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K., et al. (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69, 67-89.
8. Sukernik, R.I., Volodko, N.V., Mazunin, I.O., Eltsov, N.P., Dryomov, S.V., and Starikovskaya, E.B. (2012). Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *Am J Phys Anthropol* 148, 123-138.

9. Pliss, L., Tambets, K., Loogvali, E.L., Pronina, N., Lazdins, M., Krumina, A., Baumanis, V., and Villems, R. (2006). Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet* 70, 439-458.
10. Derbeneva, O.A., Starikovskaya, E.B., Wallace, D.C., and Sukernik, R.I. (2002). Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70, 1009-1014.
11. Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., and Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur J Hum Genet* 16, 1254-1264.
12. Grzybowski, T., Malyarchuk, B.A., Derenko, M.V., Perkova, M.A., Bednarek, J., and Wozniak, M. (2007). Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet* 1, 141-147.
13. Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Wozniak, M., and Miscicka-Sliwka, D. (2002). Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66, 261-283.
14. Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Drobniak, K., and Miscicka-Sliwka, D. (2003). Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67, 412-425.
15. Derenko, M.V., Grzybowski, T., Malyarchuk, B.A., Dambueva, I.K., Denisova, G.A., Czarny, J., Dorzhu, C.M., Kakpakov, V.T., Miscicka-Sliwka, D., Wozniak, M., et al. (2003). Diversity of mitochondrial DNA lineages in South Siberia. *Annals of human genetics* 67, 391-411.
16. Gonzalez, A.M., Brehm, A., Perez, J.A., Maca-Meyer, N., Flores, C., and Cabrera, V.M. (2003). Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol* 120, 391-404.
17. Pereira, L., Cunha, C., and Amorim, A. (2004). Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med* 118, 132-136.
18. Comas, D., Calafell, F., Bendukidze, N., Fananas, L., and Bertranpetit, J. (2000). Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol* 112, 5-16.
19. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74, 827-845.
20. Thomas, M.G., Barnes, I., Weale, M.E., Jones, A.L., Forster, P., Bradman, N., and Pramstaller, P.P. (2008). New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *Eur J Hum Genet* 16, 124-134.
21. McEvoy, B., Richards, M., Forster, P., and Bradley, D.G. (2004). The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* 75, 693-702.

22. Hedman, M., Brandstatter, A., Pimenoff, V., Sistonen, P., Palo, J.U., Parson, W., and Sajantila, A. (2007). Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci Int* 172, 171-178.
23. Meinila, M., Finnila, S., and Majamaa, K. (2001). Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered* 52, 160-170.
24. Bosch, E., Calafell, F., Gonzalez-Neira, A., Flaiz, C., Mateu, E., Scheil, H.G., Huckenbeck, W., Efremovska, L., Mikerezi, I., Xiotiris, N., et al. (2006). Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70, 459-487.
25. Brandstatter, A., Klein, R., Duftner, N., Wiegand, P., and Parson, W. (2006). Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int J Legal Med* 120, 310-314.
26. Lutz, S., Weisser, H.J., Heizmann, J., and Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111, 67-77.
27. Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., Dambueva, I., and Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PloS one* 5, e15214.
28. Gokcumen, O., Dulik, M.C., Pai, A.A., Zhadanov, S.I., Rubinstein, S., Osipova, L.P., Andreenkov, O.V., Tabikhanova, L.E., Gubina, M.A., Labuda, D., et al. (2008). Genetic variation in the enigmatic Altaian Kazakhs of South-Central Russia: insights into Turkic population history. *Am J Phys Anthropol* 136, 278-293.
29. Piercy, R., Sullivan, K.M., Benson, N., and Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106, 85-90.
30. Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V.G., Montagnon, D., Reis, T., and Ludes, B. (2009). Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Human genetics* 126, 395-410.
31. Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., Zaporozhchenko, V., Gronenborn, D., Moiseyev, V., Kolpakov, E., et al. (2013). Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet* 9, e1003296.
32. Chen, F., Wang, S.Y., Zhang, R.Z., Hu, Y.H., Gao, G.F., Liu, Y.H., and Kong, Q.P. (2008). Analysis of mitochondrial DNA polymorphisms in Guangdong Han Chinese. *Forensic Sci Int Genet* 2, 150-153.
33. Cui, Y., Li, C., Gao, S., Xie, C., and Zhou, H. (2010). Early Eurasian migration traces in the Tarim Basin revealed by mtDNA polymorphisms. *Am J Phys Anthropol* 142, 558-564.
34. Kong, Q.P., Yao, Y.G., Liu, M., Shen, S.P., Chen, C., Zhu, C.L., Palanichamy, M.G., and Zhang, Y.P. (2003). Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 113, 391-405.

35. Liu, C., Wang, S.Y., Zhao, M., Xu, Z.Y., Hu, Y.H., Chen, F., Zhang, R.Z., Gao, G.F., Yu, Y.S., and Kong, Q.P. (2011). Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. *Forensic Sci Int Genet* 5, e4-10.
36. Peng, M.S., He, J.D., Liu, H.X., and Zhang, Y.P. (2011). Tracing the legacy of the early Hainan Islanders--a perspective from mitochondrial DNA. *BMC Evol Biol* 11, 46.
37. Powell, G.T., Yang, H., Tyler-Smith, C., and Xue, Y. (2007). The population history of the Xibe in northern China: a comparison of autosomal, mtDNA and Y-chromosomal analyses of migration and gene flow. *Forensic Sci Int Genet* 1, 115-119.
38. Wen, B., Li, H., Gao, S., Mao, X., Gao, Y., Li, F., Zhang, F., He, Y., Dong, Y., Zhang, Y., et al. (2005). Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22, 725-734.
39. Phillips-Krawczak, C., Devor, E., Zlojutro, M., Moffat-Wilson, K., and Crawford, M.H. (2006). MtDNA variation in the Altai-Kizhi population of southern Siberia: a synthesis of genetic variation. *Hum Biol* 78, 477-494.
40. Jin, H.J., Kwak, K.D., Hong, S.B., Shin, D.J., Han, M.S., Tyler-Smith, C., and Kim, W. (2006). Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: an expanded Korean population database. *Forensic Sci Int* 158, 125-130.
41. Imaizumi, K., Parsons, T.J., Yoshino, M., and Holland, M.M. (2002). A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med* 116, 68-73.
42. Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., Shaikhaev, G.O., Dorzhu, C.M., Nimaev, D.D., and Zakharov, I.A. (2000). Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. *Hum Biol* 72, 945-973.
43. Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., and Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120, 334-353.
44. Rakha, A., Shin, K.J., Yoon, J.A., Kim, N.Y., Siddique, M.H., Yang, I.S., Yang, W.I., and Lee, H.Y. (2011). Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int J Legal Med* 125, 841-848.
45. Kolman, C.J., Sambuughin, N., and Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142, 1321-1334.
46. Bodner, M., Zimmermann, B., Rock, A., Kloss-Brandstatter, A., Horst, D., Horst, B., Sengchanh, S., Sanguansermisri, T., Horst, J., Kramer, T., et al. (2011). Southeast Asian diversity: first insights into the complex mtDNA structure of Laos. *BMC Evol Biol* 11, 49.
47. Pfeiffer, H., Brinkmann, B., Huhne, J., Rolf, B., Morris, A.A., Steighner, R., Holland, M.M., and Forster, P. (1999). Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112, 291-298.

48. Keyser-Tracqui, C., Crubezy, E., and 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.
49. Kim, K., Brenner, C.H., Mair, V.H., Lee, K.H., Kim, J.H., Gelegdorj, E., Batbold, N., Song, Y.C., Yun, H.W., Chang, E.J., et al. (2010). A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *Am J Phys Anthropol* 142, 429-440.
50. Zhang, F., Xu, Z., Tan, J., Sun, Y., Xu, B., Li, S., Zhao, X., Zhou, H., Gong, G., Zhang, J., et al. (2010). Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am J Phys Anthropol* 142, 314-320.
51. Li, H., Zhao, X., Zhao, Y., Li, C., Si, D., Zhou, H., and Cui, Y. (2011). Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J Hum Genet* 56, 815-822.
52. Zhao, Y.B., Li, H.J., Cai, D.W., Li, C.X., Zhang, Q.C., Zhu, H., and Zhou, H. (2010). Ancient DNA from nomads in 2500-year-old archeological sites of Pengyang, China. *J Hum Genet* 55, 215-218.
53. Irwin, J.A., Saunier, J.L., Strouss, K.M., Diegoli, T.M., Sturk, K.A., O'Callaghan, J.E., Paintner, C.D., Hohoff, C., Brinkmann, B., and Parsons, T.J. (2008). Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med* 122, 257-259.
54. Malyarchuk, B.A., Vanecek, T., Perkova, M.A., Derenko, M.V., and Sip, M. (2006). Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol* 78, 681-696.
55. Chikisheva, T.A., Gubina, M.A., Kulikov, I.V., Karafet, T.M., Voevoda, M.I., and Romaschenko, A.G. (2007). A paleogenetic study of the prehistoric populations of the Altai. *Archeology, Ethnology & Anthropology of Eurasia* 4, 130-142.
56. Clisson, I., Keyser, C., Francfort, H.P., Crubezy, E., Samashev, Z., and Ludes, B. (2002). Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, Early 3rd Century BC). *International journal of legal medicine* 116, 304-308.
57. Gao, S., Cui, Y., Yang, Y., Duan, R., Abuduresule, I., Mair, V.H., Zhu, H., and Zhou, H. (2008). Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang, China. *Sci China C Life Sci* 51, 205-213.
58. Sato, T., Razhev, D., Amano, T., and Masuda, R. (2011). Genetic features of ancient West Siberian people of the Middle Ages, revealed by mitochondrial DNA haplogroup analysis. *J Hum Genet* 56, 602-608.
59. Ploski, R., Wozniak, M., Pawlowski, R., Monies, D.M., Branicki, W., Kupiec, T., Kloosterman, A., Dobosz, T., Bosch, E., Nowak, M., et al. (2002). Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Hum Genet* 110, 592-600.
60. Hohoff, C., Dewa, K., Sibbing, U., Hoppe, K., Forster, P., and Brinkmann, B. (2007). Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *Int J Legal Med* 121, 359-363.

61. Petrejčikova, E., Sotak, M., Bernasovska, J., Bernasovsky, I., Rebala, K., Sovicova, A., Boronova, I., Bozikova, A., Gabrikova, D., Svickova, P., et al. (2011). Allele frequencies and population data for 11 Y-chromosome STRs in samples from Eastern Slovakia. *Forensic Sci Int Genet* 5, e53-62.
62. Cai, X., Qin, Z., Wen, B., Xu, S., Wang, Y., Lu, Y., Wei, L., Wang, C., Li, S., Huang, X., et al. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 6, e24282.
63. Gayden, T., Bukhari, A., Chennakrishnaiah, S., Stojkovic, O., and Herrera, R.J. (2011). Y-chromosomal microsatellite diversity in three culturally defined regions of historical Tibet. *Forensic Sci Int Genet* 6, 437-446.
64. Malyarchuk, B., Derenko, M., Denisova, G., Maksimov, A., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2011). Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *Journal of human genetics* 56, 583-588.
65. Soltyszewski, I., Pepinski, W., Spolnicka, M., Kartasinska, E., Konarzewska, M., and Janica, J. (2007). Y-chromosomal haplotypes for the AmpFISTR Yfiler PCR Amplification Kit in a population sample from Central Poland. *Forensic Sci Int* 168, 61-67.
66. Wozniak, M., Grzybowski, T., Starzynski, J., and Marciniak, T. (2007). Continuity of Y chromosome haplotypes in the population of Southern Poland before and after the Second World War. *Forensic Sci Int Genet* 1, 134-140.
67. Derenko, M., Malyarchuk, B., Denisova, G.A., Wozniak, M., Dambueva, I., Dorzhu, C., Luzina, F., Miscicka-Sliwka, D., and Zakharov, I. (2006). Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet* 118, 591-604.
68. Mirabal, S., Regueiro, M., Cadenas, A.M., Cavalli-Sforza, L.L., Underhill, P.A., Verbenko, D.A., Limborska, S.A., and Herrera, R.J. (2009). Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur J Hum Genet* 17, 1260-1273.
69. Roewer, L., Willuweit, S., Kruger, C., Nagy, M., Rychkov, S., Morozowa, I., Naumova, O., Schneider, Y., Zhukova, O., Stoneking, M., et al. (2008). Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. *Int J Legal Med* 122, 219-223.
70. Rodig, H., Roewer, L., Gross, A., Richter, T., de Knijff, P., Kayser, M., and Brabetz, W. (2008). Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Sci Int* 174, 182-188.
71. Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.L., and Lahermo, P. (2008). Migration waves to the Baltic Sea region. *Ann Hum Genet* 72, 337-348.
72. Achakzai, N.M., Rahman, Z., Shahzad, M.S., Daud, S., Zar, M.S., Israr, M., Husnain, T., Willuweit, S., and Roewer, L. (2011). Y-chromosomal STR analysis in the Pashtun population of Southern Afghanistan. *Forensic Sci Int Genet* 6, e103-105.
73. Lacau, H., Bukhari, A., Gayden, T., La Salvia, J., Regueiro, M., Stojkovic, O., and Herrera, R.J. (2011). Y-STR profiling in two Afghanistan populations. *Leg Med (Tokyo)* 13, 103-108.

74. Frank, W.E., Ralph, H.C., and Tahir, M.A. (2008). Y chromosome STR haplotypes and allele frequencies in a southern Indian male population. *J Forensic Sci* 53, 248-251.
75. Haliti, N., Carapina, M., Masic, M., Strinovic, D., Klaric, I.M., and Kubat, M. (2009). Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croats. *Forensic Sci Int Genet* 3, e137-138.
76. Zastera, J., Roewer, L., Willuweit, S., Sekerka, P., Benesova, L., and Minarik, M. (2010). Assembly of a large Y-STR haplotype database for the Czech population and investigation of its substructure. *Forensic Sci Int Genet* 4, e75-78.
77. Lessig, R., Edelmann, J., Kleemann, W.J., and Kozhemyako, V. (2006). Population data of Y-chromosomal STRs in Russian males of the Primorye region population. *Forensic Sci Int* 159, 71-76.
78. Roewer, L., Willuweit, S., Stoneking, M., and Nasidze, I. (2009). A Y-STR database of Iranian and Azerbaijanian minority populations. *Forensic Sci Int Genet* 4, e53-55.
79. Chang, Y.M., Perumal, R., Keat, P.Y., and Kuehn, D.L. (2007). Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci Int* 167, 70-76.
80. Holmlund, G., Nilsson, H., Karlsson, A., and Lindblom, B. (2006). Y-chromosome STR haplotypes in Sweden. *Forensic Sci Int* 160, 66-79.
81. Cakir, A.H., Celebioglu, A., and Yardimci, E. (2004). Y-STR haplotypes in Central Anatolia region of Turkey. *Forensic Sci Int* 144, 59-64.
82. Zhu, B.F., Wang, Z.Y., Yang, C.H., Li, X.S., Zhu, J., Yang, G., Huang, P., and Liu, Y. (2005). Y-chromosomal STR haplotypes in Chinese Uigur ethnic group. *Int J Legal Med* 119, 306-309.
83. Theves, C., Balaesque, P., Evdokimova, L.E., Timofeev, I.V., Alekseev, A.N., Sevin, A., Crubezy, E., and Gibert, M. (2010). Population genetics of 17 Y-chromosomal STR loci in Yakutia. *Forensic Sci Int Genet* 4, e129-130.
84. Hashiyada, M., Umetsu, K., Yuasa, I., Tamura, A., Matsusue, A., Suzuki, K., Kashimura, S., and Funayama, M. (2008). Population genetics of 17 Y-chromosomal STR loci in Japanese. *Forensic Sci Int Genet* 2, e69-70.

Additional Files TableS7

Table S7. Genotypes obtained from the typing of 28 autosomal SNPs for modern samples.

Sample	N29insA		rs11547464	rs885479	rs1805008	rs1805005	rs1805006	rs1805007	rs1805009*	Y1520CH	rs2228479	rs1110400	rs28777	rs16891982	rs12821256	rs4959270	rs12203592	rs1042602	rs1800407	rs2402130	rs12913832	rs2378249	rs12896399	rs1393350	rs683	rs1545397	rs2814778	rs1426654	rs3827760	hair color				shade		eye color			Predicted		
	C/A	G/A	G/A	C/T	G/T	C/A	C/T	C/A	G/C	C/A	A/G	T/C	C/A	G/C	C/T	C/A	C/T	C/A	C/T	G/A	G/A	G/A	G/T	G/A	C/A	T/A	G/A	G/A	C/T	Brown	Red	Black	Blond	Light	Dark	Blue	Int	Brown	hair	eye	ancestry
M1	C	G	G	C/T	G	C	C	-	C	G	T	CA	GC	T	A	C	C	C	G	G	A	G	G	G	CA	A	A	A	T	0.255	0.055	0.146	0.545	0.779	0.221	0.678	0.111	0.191	D-brown/D-blond	blue	E
M2	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	C	C	CA	C	GA	GA	A	G	G	C	A	A	A	T	0.378	0.002	0.338	0.282	0.43	0.57	0.097	0.148	0.758	D-brown/L-brown	brown	E	
M3	C	G	G	C	G	C	C	-	C	G	T	A	G	CT	CA	C	CA	C	A	A	A	GT	GA	CA	A	A	A	T	0.32	0	0.413	0.267	0.368	0.632	0.002	0.034	0.964	D-brown/L-brown	brown	E	
M4	C	G	G	C	G	C	C	-	C	G	T	CA	GC	CT	CA	CT	CA	C	A	GA	A	GT	G	A	TA	A	A	T	0.324	0	0.615	0.06	0.076	0.924	0.065	0.147	0.787	Black/D-brown	brown	E	
M5	C	G	G	C	G	C	C	G	C	G	T	A	C	CT	C	C	C	C	G	A	A	G	G	C	A	G	GA	T	0.16	0	0.832	0.008	0.018	0.982	0	0.006	0.994	Black/D-brown	brown	Afr	
M6	C	G	G	C	G	C	C	-	C	G	T	A	GC	T	CA	C	A	C	A	G	A	G	G	CA	A	A	A	T	0.228	0.002	0.163	0.608	0.838	0.162	0.678	0.111	0.191	D-brown/L-brown	blue	E	
M7	C	G	G	C	G	C	C	-	C	G	T	A	G	C	CA	C	A	C	GA	GA	A	GT	G	A	A	A	A	T	0.235	0.001	0.074	0.689	0.866	0.134	0.15	0.14	0.711	D-brown/L-brown	brown	E	
M8	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	C	C	CA	C	GA	GA	A	G	G	CA	A	A	A	T	0.348	0.002	0.328	0.321	0.475	0.525	0.094	0.148	0.758	D-brown/L-brown	brown	E	
M9	C	G	G	C	GT	C	C	G	C	G	T	A	G	T	CA	CT	CA	C	GA	GA	A	G	GA	C	A	A	A	T	0.445	0.002	0.476	0.078	0.154	0.846	0.19	0.271	0.539	D-brown/L-brown	brown	E	
M10	C	G	G	C/T	G	C	C	G	C	G	T	A	G	T	A	C	C	C	GA	G	A	G	GA	A	A	(A)	A	T	0.064	0.052	0.021	0.863	0.982	0.018	0.899	0.066	0.035	blond	blue	E	
M11	C	G	G	C/T	G	C	C	C/T	-	C	G	T	A	G	T	CA	CT	C	C	A	G	GA	GT	GA	CA	A	A	T	0.048	0.8	0.008	0.145	0.972	0.028	0.943	0.045	0.012	ref	blue	E	
M12	C	G	G	C	GT	C	C	-	C	G	T	A	G	T	CA	C	C	C	G	G	A	G	GA	CA	TA	A	A	T	0.22	0.006	0.086	0.686	0.896	0.104	0.899	0.066	0.035	D-brown/D-blond	blue	E	
M13	C	G	G	C/T	G	C	C	G	C	G	T	A	G	CT	CA	C	CA	CT	GA	G	GA	G	G	CA	A	A	A	T	0.074	0.055	0.008	0.862	0.989	0.011	0.91	0.072	0.018	blond	blue	E	
M14	C	G	G	C	GT	C	C	-	C	AG	T	A	G	T	CA	C	C	T	A	A	A	A	GT	GA	A	A	A	T	0.286	0.001	0.407	0.305	0.566	0.434	0.019	0.218	0.763	D-brown/L-brown	brown	E	
M15	C	G	G	C	G	C	C	-	C	G	T	A	G	CT	C	C	A	C	GA	G	A	G	G	A	A	A	A	T	0.142	0.002	0.044	0.812	0.931	0.069	0.87	0.076	0.053	blond	blue	E	
M16	C	G	G	C	G/A	C	C	-	C	G	T	C	C	C	T	C	C	C	GA	A	A	GT	G	C	TA	A	G	C	0.079	0	0.919	0.001	0.001	0.999	0	0.006	0.994	Black/D-brown	brown	As	
M17	C	G	G	C/A	C	G	C	-	C	G	T	C	C	C	T	C	C	C	CT	GA	A	GA	G	G	C	TA	A	G	C	0.127	0	0.871	0.003	0.002	0.998	0	0.017	0.983	Black/D-brown	brown	As
M18	C	G	G	C/A	C	G	C	-	C	G	T	CA	GC	T	C	C	C	C	C	GA	A	G	G	G	C	TA	A	G	C	0.211	0	0.778	0.011	0.009	0.991	0	0.013	0.987	Black/D-brown	brown	As

* The MALDI TDF mass spectrometry typing of this SNP was null
SNPs typed by direct sequencing.

Additional Files TableS8

Table S8. Genotypes obtained from the typing of 24 autosomal SNPs for ancient specimens.

Sample	N29InA	r111547464	rs85479	r1805008	r1805005	r1805006	r1805007	r1805009*	V1520CH	r2228479	r1110400	r28777	r116891982	r112821256	rs4959270	r112203592	r11042602	r1800407	r2402130	r112913832	r12378249	r112896399	r11393350	rs683	Brown	Red	hair color	Black	Blond	Light	shade	Dark	Blue	eye color	Brown	
	C/A	G/A	G/A	C/T	G/T	C/A	C/T	G/C	C/A	A/G	T/C	C/A	G/C	C/T	C/A	C/T	C/A	C/T	G/A	G/A	G/A	G/T	G/A	C/A												
TA4	C	-	G/A	C	G	C	C	-	C	G	T	A	G	T	CA	CT	(C)	C	A	A	A	G	G	A	C	0.192	0	0.785	0.023	0.047	0.953	0.002	0.05	0.548		
TA8	C	-	-	-	G	C	-	-	-	G	-	A	G	CT	A	CT	(C)	CT	A	(G)A	GA	(G)T	8	C	-	-	-	-	-	-	-	-	-	0.872		
TA9	C	G	G	C	G	C	C	-	C	G	T	CA	GC	T	C	C	GA	C	GA	GA	G	GT	G	C	0.375	0.001	0.549	0.075	0.061	0.939	0.04	0.083	0.872			
TA10	(C)	G	(A)	C	G	C	C	-	C	G	T	A	G	T	C	C	C	CT	A	A	A	G	G	C	0.261	0	0.658	0.081	0.177	0.823	0.003	0.071	0.956			
TA11	C	G	A/G	C	G	C	C	-	C	G	T	C	C	CT	C	C	(C)	C	GA	A	G	GT	G	A	0.171	0	0.824	0.005	0.003	0.997	0	0.006	0.994			
TA12	C	G	G	C	G	C	C	-	C	G	T	C	C	C	T	A	C	C	G	A	A	G	G	(C)A	0.117	0	0.878	0.005	0.003	0.997	0	0.006	0.994			
TA14	C	G	G/A	C	G	C	C	-	C	G	T	A	G	T	CA	CT	C	C	GA	G	GA	G	G	C	0.455	0.003	0.288	0.255	0.534	0.466	0.885	0.085	0.03			
TU15	(C)	(G)	(G)	(C)	G	C	(C)	-	(C)	G	(T)	A	G	T	C	C	C	(A)	A	A	A	T	G	CA	0.223	0	0.65	0.217	0.171	0.829	0.003	0.026	0.971			
TU17	C	G	A/G	C	GT	C	C	-	C	G	T	CA	GC	T	C	C	C	A	GA	A	T	G	CA	0.256	0	0.66	0.083	0.118	0.882	0.066	0.081	0.913				
TU31	C	-	A/G	C	G	C	C	-	C	G	T	(A)	GC	T	A	C	C	C	(A)	A	GA	GT	GA	CA	0.203	0	0.765	0.031	0.039	0.961	0.001	0.016	0.983			
TU34	C	G	-	C	G	C	C	-	C	G	T	A	G	T	C	C	C	C	GA	GA	GA	G	G	CA	0.31	0.002	0.322	0.366	0.49	0.51	0.094	0.148	0.758			

* The typing of this SNP was null.

SNPs typed by direct sequencing:

Alleter in brackets were just observed once.

Dash denotes locus for which amplification was null

Table S9. Genotypes of AIM for the aAM samples

Sample	rs1545397	rs2814778	rs1426654	rs16891982	rs3827760	Inferred ancestry of individual		
	T/A	G/A	G/A	G/C	C/T	cluster 1	cluster 2	cluster 3
TA4	T	(A)	A	G	C	0.381	0.005	0.615
TA8	A	(A)	GA	G	T	0.21	0.01	0.78
TA9	TA	(A)	GA	GC	CT	0.391	0.009	0.6
TA10	A	(A)	GA	G	CT	0.273	0.009	0.718
TA11	TA	(A)	GA	C	T	0.391	0.014	0.595
TA12	A	(A)	A	C	T	0.26	0.014	0.725
TA14	A	A	GA	G	CT	0.27	0.009	0.721
TU15	T	A	GA	G	T	0.314	0.008	0.677
TU17	A	A	A	GC	CT	0.267	0.008	0.725
TU18	-	-	-	-	-	-	-	-
TU31	A	(A)	GA	GC	T	0.265	0.014	0.72
TU34	A	A	A	G	T	0.141	0.007	0.851

Alleles in brackets were just observed once

Publications annexes au sujet de thèse

Participations aux exercices du groupe EDNAP (European DNA Profiling group)

RNA/DNA co-analysis from human saliva and semen stains: results of a third collaborative EDNAP exercise

Haas C, Hanson E, Anjos MJ, Banemann R, Berti A, Borges E, Carracedo A, Carvalho M, Courts C, De Cock G, Dötsch M, Flynn S, Gomes I, **Hollard C**, Hjort B, Hoff-Olsen P, Hřebíková K, Lindenberg A, Ludes B, Maroñas O, McCallum N, Moore D, Morling N, Niederstätter H, Noel F, Parson W, Popielarz C, Rapone C, Roeder AD, Ruiz Y, Sauer E, Schneider PM, Sijen T, Court DS, Sviežená B, Turanská M, Vidaki A, Zatkalíková L, Ballantyne J.

Forensic Sci Int Genet. 2013 Feb;7(2):230-9



RNA/DNA co-analysis from human saliva and semen stains – Results of a third collaborative EDNAP exercise

C. Haas^{a,*}, E. Hanson^b, M.J. Anjos^k, R. Banemann^r, A. Bertiⁿ, E. Borges^j, A. Carracedo^t, M. Carvalho^k, C. Courts^h, G. De Cock^m, M. Dötsch^r, S. Flynn^u, I. Gomes^g, C. Hollard^c, B. Hjort^l, P. Hoff-Olsen^q, K. Hríbková^w, A. Lindenbergh^d, B. Ludes^c, O. Maroñas^t, N. McCallum^u, D. Moore^p, N. Morling^l, H. Niederstätter^o, F. Noel^m, W. Parson^o, C. Popielarz^j, C. Raponeⁿ, A.D. Roeder^s, Y. Ruiz^t, E. Sauer^h, P.M. Schneider^g, T. Sijen^d, D. Syndercombe Court^v, B. Sviežená^e, M. Turanská^f, A. Vidaki^v, L. Zatkalíková^f, J. Ballantyne^b

^a Institute of Legal Medicine, University of Zurich, Switzerland

^b National Center for Forensic Science, University of Central Florida, Orlando, USA

^c Institute of Legal Medicine, University of Strasbourg, France

^d Netherlands Forensic Institute, The Hague, The Netherlands

^e Institute of Forensic Science, Department of Criminalistic Biology and Genetic Analysis, Bratislava, Slovakia

^f Institute of Forensic Science, Ministry of the Interior, Department of Biology and DNA Analysis, Slovenská Lupca, Slovakia

^g Institute of Legal Medicine, Medical Faculty, University of Cologne, Germany

^h Institute of Legal Medicine, University of Bonn, Germany

^j Institut National de Police Scientifique, Laboratoire de Police Scientifique de Lyon, France

^k Forensic Genetic Service, Centre Branch, National Institute of Legal Medicine, Coimbra, Portugal

^l Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

^m National Institute for Criminalistics and Criminology, Brussels, Belgium

ⁿ Reparto Investigazioni Scientifiche di Roma, Rome, Italy

^o Institute of Legal Medicine, Innsbruck Medical University, Austria

^p LGC Forensics, Middlesex, UK

^q Department of Forensic Genetics, Norwegian Institute of Public Health, Oslo, Norway

^r Bundeskriminalamt, Wiesbaden, Germany

^s Orchid Cellmark Ltd., Abingdon, UK

^t Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain

^u Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, Scotland, UK

^v Academic Haematology, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK

^w Comenius University, Faculty of Natural Sciences, Department of Genetics, Bratislava, Slovakia

ARTICLE INFO

Article history:

Received 22 April 2012

Received in revised form 27 August 2012

Accepted 12 October 2012

Keywords:

Forensic science
Body fluid identification
Saliva
Semen
EDNAP exercise
mRNA profiling

ABSTRACT

A third collaborative exercise on RNA/DNA co-analysis for body fluid identification and STR profiling was organized by the European DNA Profiling Group (EDNAP). Twenty saliva and semen stains, four dilution series (10–0.01 µl saliva, 5–0.01 µl semen) and, optionally, *bona fide* or mock casework samples of human or non-human origin were analyzed by 20 participating laboratories using an RNA extraction or RNA/DNA co-extraction method. Two novel mRNA multiplexes were used: a saliva triplex (HTN3, STATH and MUC7) and a semen pentaplex (PRM1, PRM2, PSA, SEMG1 and TGM4). The laboratories used different chemistries and instrumentation and a majority (16/20) were able to successfully isolate and detect mRNA in dried stains. The simultaneous extraction of RNA and DNA from individual stains not only permitted a confirmation of the presence of saliva/semen (i.e. tissue/fluid source of origin), but allowed an STR profile of the stain donor to be obtained as well. The method proved to be reproducible and sensitive, with as little as 0.05 µl saliva or semen, using different analysis strategies. Additionally, we demonstrated the ability to positively identify the presence of saliva and semen, as well as obtain high quality DNA profiles, from old and compromised casework samples. The results of this collaborative exercise involving an RNA/DNA co-extraction strategy support the potential use of an mRNA based system for the identification of saliva and semen in forensic casework that is compatible with current DNA analysis methodologies.

© 2012 Elsevier Ireland Ltd. All rights reserved.

* Corresponding author at: Institute of Legal Medicine, Forensic Genetics, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

Tel.: +41 44 635 56 56; fax: +41 44 635 68 58.

E-mail address: cordula.haas@irm.uzh.ch (C. Haas).

1. Introduction

The analysis of cell-specific mRNA expression is a confirmative method for the identification of body fluids [1–24], as opposed to conventional immunological and enzymatic tests, most of which either lack specificity or cannot be used to identify all forensically relevant biological fluids. The suitability of mRNA profiling assays with forensic samples, such as old and environmentally compromised samples, has been demonstrated [25–28]. Previously, two collaborative exercises were performed by the European DNA Profiling Group (EDNAP – <http://www.isfg.org/EDNAP>) in order to evaluate the robustness and reproducibility of mRNA profiling for blood identification: (1) evaluation of three blood-specific markers (HBB, SPTB and PBGD) using singleplex reactions [29]; (2) evaluation of seven blood-specific markers using two multiplex systems, a ‘high sensitivity’ duplex (HBB, HBA) and a ‘moderate sensitivity’ pentaplex (ALAS2, CD3G, ANK1, PBGD and SPTB) [30]. Most laboratories, some of which had no prior experience with RNA, were able to successfully isolate and analyze RNA from the provided samples. While sensitivity varied between laboratories, the method proved to be reproducible and sensitive using different analysis strategies [29,30].

A third collaborative exercise was organized by the Institute of Legal Medicine, University of Zürich, Switzerland, on behalf of EDNAP in order to test forensically suitable saliva and semen markers. In a preliminary study performed by the Florida and Zürich laboratories, various saliva and semen mRNA markers described in the literature [3,6,8,9,13,15,18] as well as from unpublished data (Table 1) were evaluated in terms of sensitivity, specificity and performance with casework samples. In this study, 7 saliva and 11 semen markers were tested, of which 3 saliva and 5 semen markers were deemed most suitable for forensic use (Table 1). For this exercise, the following saliva and semen multiplexes were developed and provided to the participating laboratories: (1) a saliva triplex including the markers HTN3, STATH and MUC7 and (2) a semen pentaplex allowing the detection and differentiation of sperm (PRM1, PRM2) and seminal plasma (PSA, SEMG1 and TGM4), the latter of which is necessary for the identification of semen from azoospermic men. The exercise included two rounds of testing: In the original exercise (part 1), 5 laboratories encountered technical

problems (reduced sensitivity, no result at all, contamination of negative controls and spectral pull-ups). As a consequence, a second exercise (part 2) was arranged in order to identify the source of these problems, and potentially to obtain improved results. Each part of the exercise included the analysis of 10 mock casework samples, saliva and semen dilution series and, if available, optional extra casework samples from the participating laboratories. In addition to the provided samples analyzed by all laboratories, a set of human tissues and animal saliva samples were tested with the multiplexes by 4 laboratories to complement the specificity testing of the markers.

2. Materials and methods

2.1. Samples and materials provided

The organizing laboratory (Institute of Legal Medicine, University of Zürich, Switzerland) sent 2 sets of samples to the participating laboratories: part 1 included stains 1–10 (saliva, semen, non-saliva/semen and non-human stains) and dilution series A (5–0.01 µl semen) and B (5–0.01 µl saliva); part 2 included stains 11–20 (saliva, semen, non-saliva/semen and non-human stains) and dilution series C (10–0.05 µl saliva) and D (5–0.01 µl semen). Fresh saliva samples were collected from 10 individuals for 12 stains/dilution series (different donors were used for parts 1 and 2 except for one donor who was used in both parts) and were deposited onto different carrier materials including cotton swab/pads, recycling tissue, paper and glass slides. Additionally, quarter pieces of chewed chewing gums and licked plastic spoons were provided. Semen samples from 9 individuals were used to prepare 10 stains/dilution series (different donors were used for parts 1 and 2 except for one donor who was used in both parts). The semen, which had been frozen for up to 25 years, was deposited onto different carrier materials including cotton swab/pads, white textile, toilet paper and latex gloves. In addition, blood samples (2 donors), vaginal swabs (1 donor) and buccal swabs from a dog and a cat were used as non-saliva/semen and non-human stains. For the dilution series, saliva and semen samples were diluted in 0.9% NaCl to a final volume of 5 µl per sample and placed on swabs. The laboratories

Table 1

List of evaluated mRNA markers for the identification of saliva and semen.

RNA marker	Protein	Ref.	
<i>Saliva</i>			
HTN3	Histatin 3	6	} good candidates
STATH	Statherin	6	
MUC7	Mucin 7	*	
PRB1-3	Proline-rich proteins	*	} not sensitive
PRB4	Proline-rich protein	8	
SPRR2A	small proline-rich protein 2A	13	} not specific
KRT13	Keratin 13	13	
<i>Semen</i>			
PRM1	Protamine 1	3,6	} good candidates
PRM2	Protamine 2	3,6	
TGM4	Transglutaminase 4	18,*	
PSA/KLK3	Prostate Specific Antigen, Kallikrein 3	9,15	
SEMG1	Semenogelin 1	9,15	
SPANXB	SPANX family member B	*	} inconsintest results during singleplex testing
HSFY	Heat Shock Transcription Factor, Y-Linked	*	
SPAM1-v2	Sperm Adhesion Molecule	*	
ZBPB	Zona Pellucida Binding Protein	*	
ODF1	Outer Dense Fiber of Sperm Tails 1	*	
BPY2/VCY2	Basic Charged, Y-linked, 2	*	

* Ballantyne (unpublished)

were asked, but not required, to examine additional samples: (1) *bona fide* or mock casework material that could include human and/or non-human saliva and semen stains; or (2) other forensically relevant body fluids (semen, saliva, vaginal secretions or menstrual blood).

HPLC-purified primers were purchased from Microsynth (St. Gallen, Switzerland) or Applied Biosystems (AB, Rotkreuz, Switzerland). The organizing laboratory prepared primer mixes for the multiplexes and aliquots were provided to the participating laboratories for use in amplification reactions (see below). In the second part, singleplex primers for HTN3, MUC7, PRM2 and SEMG1 were provided to permit a comparison of results obtained using singleplex and multiplex analysis. The primers and samples were sent by normal post (part 1, arrival within 2–4 days, except 2 parcels that were 7 and 11 days in transit) or by Fedex (part 2, arrival within 1–3 days, except 1 parcel that was 7 days in transit).

2.2. Human tissue and animal samples

Total RNA from 20 human tissues (adipose, bladder, brain, cervix, colon, esophagus, heart, kidney, liver, lung, ovary, placenta, prostate, skeletal muscle, small intestine, spleen, testes, thymus, thyroid and trachea) included in the FirstChoice[®] Human Total RNA Survey Panel was obtained from Applied Biosystems/Ambion. All tissues included in the panel were 3-donor pooled samples.

Buccal swabs and semen samples of domestic and exotic animals were collected by veterinary practitioners, zookeepers, a volunteer at an animal shelter, and/or owners (laboratories 9, 10 and 19). Although some samples were obtained from the same species, they were collected from different animals.

2.3. RNA/(DNA co-)extraction and reverse transcription

The participating laboratories were asked to use the entire swab or stain for extraction (RNA only or RNA/DNA co-extraction). The organizing laboratory provided an example protocol for extraction and reverse transcription. The laboratories could, however, use methods of their own choice. The example protocol was described previously [29,30].

2.4. RNA- and DNA-quantification

Participating laboratories were asked to quantify RNA using one of the following methods: Quant-iT[™] RiboGreen[®] RNA Assay kit (Invitrogen) using a fluorescence microplate reader (high and low range protocol option) [6]; Quant-iT[™] RNA assay kit (Invitrogen) using the Qubit fluorometer [21]; or Bioanalyzer (Agilent) [15]. Laboratories were able to quantify DNA using a quantification system of their own choice (Table S1, supplementary materials).

2.5. Endpoint PCR

The primer sets and amplicon lengths are shown in Table S2. An additional larger amplicon (152 bp) detected in saliva samples using the HTN3 primers corresponds to another histatin isoform (HTN1), which shows 95% sequence identity to HTN3 [6,14]. For the animal sample testing, it was expected that many “negative” results would be obtained due to the deliberate design and use of putative human or primate specific primer sequences. In an attempt to ensure that a negative result was genuine and not due to an absence of high quality RNA in the extract itself, all animal samples were tested using a housekeeping gene (HKG) triplex (B2M, UBC and UCE).

In the first part of the exercise, the forward primers were 5'-labeled with the alternative dyes Yakima Yellow or ATTO550 (VIC[®] and NED[™] analogs, respectively). The alternative dyes caused

spectral issues such as pull-up in the other dyes due to the use of matrices generated with the commonly used dyes VIC[®] and NED[™] (AB). This is not a problem as long as all markers in the multiplex are labeled with one dye. For the second part of the exercise, the forward primers were labeled with the commonly used dyes (semen: VIC, saliva: NED) so that the correct spectral calibration matrices could be applied. The HKG forward primers were 5'-labeled with ATTO550.

The following amplification conditions were recommended: *saliva and HKG triplex*: the 25 µl reaction mix contained 2 µl cDNA, 2.5 µl primer mix (see below), 12.5 µl 2× Multiplex PCR Mastermix (Multiplex PCR kit, QIAGEN), 2.5 µl Q-Solution (Multiplex PCR kit; QIAGEN); *semen pentaplex*: the 25 µl reaction mix contained 2 µl cDNA, 2.5 µl primer mix (see below), 1 mM dNTPs (AB), 1× Buffer II (AB), 3.25 mM MgCl₂ (AB), and 1.5 U AmpliTaq[®] Gold DNA Polymerase (AB); *saliva and semen singleplexes*: the 25 µl reaction mix contained 2 µl cDNA, 2 µl primer mix (see below), 0.5 mM dNTPs (AB), 1× Buffer I (AB) and 1.25 U AmpliTaq Gold[®] DNA Polymerase (AB). Sterile water was used in place of cDNA for non-template controls.

The primer mixes were prepared using the following concentrations: *saliva triplex*: HTN3 0.2 µM, STATH 0.4 µM, MUC7 0.1 µM; *semen pentaplex*: PSA 0.72 µM, PRM2 0.048 µM, SEMG1 0.2 µM, TGM4 0.6 µM, PRM1 0.12 µM; *saliva and semen singleplexes*: 0.8 µM each; *HKG triplex*: B2M 0.2 µM, UBC 0.2 µM, UCE 1 µM.

The cycling conditions were as follows: *saliva triplex*: the initial denaturation was at 95 °C for 15 min, followed by 35 cycles of 94 °C 30 s, 57 °C (+0.2 °C per cycle) 90 s, 72 °C 60 s and the final elongation at 72 °C for 60 min; *semen pentaplex and saliva/semen singleplexes*: the initial denaturation was at 95 °C for 11 min, followed by 35 cycles of 94 °C 20 s, 58 °C 30 s, 72 °C 40 s and the final elongation at 72 °C for 60 min; *HKG triplex*: the initial denaturation was at 95 °C for 15 min, followed by 35 cycles of 94 °C 30 s, 55 °C (+0.2 °C per cycle) 90 s, 72 °C 40 s and the final elongation at 72 °C for 30 min.

Post PCR purification was recommended, particularly for the semen pentaplex, to eliminate dye blobs that could appear at marker specific positions (e.g. MinElute PCR purification kit, QIAGEN) [21,31].

2.6. Capillary electrophoresis (CE)

The participating laboratories used standard, multi-color fluorescent genetic analyzers and standard electrophoresis conditions for the detection of the saliva and semen specific amplicons (Table S1). Any dye set that included VIC[®] and NED[™], with associated internal lane standard, could be used. Raw data were analyzed with Genemapper[®] Software (AB). For data compilation, the threshold for a positive result was set to 100 RFUs (relative fluorescence units) for consistency between laboratories and to avoid inclusion of false positive results (due to the presence of dye blobs, baseline noise).

2.7. DNA-amplification and -detection

If DNA was co-extracted, the laboratories were able to use a standard STR typing kit, PCR and CE conditions of their choice (Table S1). A peak detection threshold of 50 RFUs was used.

3. Results

Laboratories 1–18 participated in part 1 of the exercise (stains 1–10 and dilution series A and B) and laboratories 1, 3–6, 8–13, 15–20 participated in part 2 (stains 11–20 and dilution series C and D). Each laboratory was asked to complete a questionnaire describing

the methods used (Table S1). The 20 laboratories used 5 different kits for RNA only or RNA/DNA co-extraction, 7 different reverse transcription kits, 7 different STR typing kits for DNA analysis, 5 different thermocyclers, 4 different genetic analyzers and 3 different polymers (Table S1).

The exercise was performed in 2 parts to allow for laboratories that encountered technical problems in part 1 to test additional samples once technical issues had been resolved (see Introduction). Laboratories that did not encounter technical challenges in part 1 were still encouraged to participate in part 2 of the exercise for comparison of the results from part 1 and for further evaluation of the included markers. From the experiences of part 1, various improvements were applied to part 2: (1) primers were labeled with common dyes to prevent pull-up into other spectra; (2) four RNA markers were provided additionally as singleplexes to permit comparisons of singleplex and multiplex analyses. Improved results could not be obtained in part 2 by laboratories 3, 4 and 6. These laboratories encountered non-identified technical problems, which were not attributable to the mRNA profiling method. Therefore, the results from these laboratories were excluded from

the final data compilation. The data of laboratories 5 and 18 (part 1) and laboratory 19 (part 2) were excluded from the RNA results due to the use of a non-functional reverse transcriptase (lab 5) and the presence of RT minus contamination (labs 18 and 19). With the above mentioned laboratories excluded from data compilation, there was no significant difference in the results between parts 1 and 2 in terms of detectability and sensitivity of RNA markers and recoverability of STR profiles. Therefore, the results of parts 1 and 2 were combined in the tables and figures presented here.

3.1. RNA results

The mRNA profiling results are summarized in Tables 2 and 3. For exercise parts 1 and 2, thirteen laboratories (although not the same 13 laboratories in each part) were able to perform the method successfully as demonstrated by detection of the RNA saliva and semen markers. The 8 saliva and 6 semen stains were identified correctly by most laboratories (Table 2). Analysis of stain 11 (5 µl saliva on paper) appeared to be challenging possibly due to the presence of inhibiting substances in the paper. Stains 4 and 12

Table 2

mRNA profiling results (saliva triplex and semen pentaplex) from 13 laboratories for the 20 stains. Dark gray squares represent stains/markers that were detected by more than half of the laboratories, light gray squares are stains/markers that were detected by 25–50% of the laboratories and white squares represent stains/markers that were detected by less than 25% of the laboratories. The number of laboratories that typed at least two markers of the respective body fluid correctly is shown in the last 2 columns.

Sample No.	Saliva triplex			Semen pentaplex					>2 markers positive	
	HTN3	STATH	MUC7	PSA	PRM2	SEMG1	TGM4	PRM1	Saliva	Semen
3 ul saliva on recycling tissue	3	7/13	4/13	11/13					8/13	
1/4 chewed chewing gum	6	6/13	6/13	9/13					7/13	
1/4 buccal swab	8	12/13	12/13	12/13					12/13	
5 ul saliva on pad	10	11/13	10/13	12/13					12/13	
5 ul saliva on paper	11	3/13	4/13	3/13					3/13	
5 ul saliva on glass slide	14	9/13	4/13	6/13					6/13	
2 ul saliva on swab / 1 ul blood	18	10/13	9/13	11/13				1/13	11/13	
licked plastic spoon	19	9/13	5/13	7/13					7/13	
5 ul saliva / 1 ul semen mix on swab	5	13/13	13/13	13/13	6/13	13/13	13/13	10/13	13/13	13/13
5 ul saliva / 1 ul semen mix on swab	16	12/13	12/13	13/13	8/13	10/13	12/13	6/13	10/13	12/13
2 ul semen on swab	1			1/13	8/13	13/13	13/13	10/13	13/13	13/13
3 ul semen on white textile	9				3/13	10/13	10/13	4/13	9/13	11/13
2 ul semen on toilet paper	15				4/13	9/13	8/13	4/13	9/13	9/13
3 ul semen inside latex glove	20				1/13	9/13	6/13	3/13	11/13	10/13
5 ul azoospermic semen on swab	4		1/13		10/13	3/13	12/13	11/13	1/13	12/13
3 ul azoospermic semen on pad	12				10/13	2/13	11/13	11/13		11/13
3 ul human blood on swab	2								13/13	13/13
1/4 vaginal swab	13		1/13	4/13			1/13		8/13	12/13
1/4 buccal swab (dog)	7		1/13						12/13	13/13
1/4 buccal swab (cat)	17	1/13			1/13				12/13	12/13

Table 3

mRNA profiling results (saliva triplex and semen pentaplex) from 13 laboratories for the 2 saliva and 2 semen dilution series (A, B from part 1; C, D from part 2). Dark gray squares represent stains/markers that were detected by more than half of the laboratories, light gray squares are stains/markers that were detected by 25–50% of the laboratories and white squares represent stains/markers that were detected by less than 25% of the laboratories.

	saliva dilution series B			saliva dilution series C			semen dilution series A					semen dilution series D				
	HTN3	STATH	MUC7	HTN3	STATH	MUC7	PSA	PRM2	SEMG1	TGM4	PRM1	PSA	PRM2	SEMG1	TGM4	PRM1
10 ul	n.d.	n.d.	n.d.	13/13	13/13	13/13	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
5 ul	12/13	11/13	12/13	13/13	13/13	13/13	11/13	12/13	12/13	10/13	12/13	12/13	13/13	13/13	13/13	13/13
1 ul	11/13	10/13	10/13	12/13	7/13	12/13	8/13	11/13	12/13	6/13	12/13	12/13	13/13	12/13	12/13	13/13
0.5 ul	10/13	5/13	10/13	9/13	6/13	11/13	6/13	10/13	11/13	3/13	10/13	8/13	11/13	11/13	12/13	11/13
0.1 ul	4/13		5/13	1/13	2/13	4/13	3/13	6/13	8/13	1/13	7/13	1/13	5/13	2/13	1/13	8/13
0.05 ul	2/13	1/13	4/13	1/13			5/13	6/13	2/13	3/13	3/13	1/13	6/13	3/13	2/13	7/13
0.01 ul	1/13			n.d.	n.d.	n.d.	2/13	3/13	2/13		4/13			1/13		

n.d., not determined.

were semen samples from azoospermic men and, therefore, no detection of PRM1 and PRM2 was expected. However, up to 3 laboratories detected one of these markers, which may be due to detection of residual sperm despite vasectomy. Two stains were saliva/semen mixtures (stains 5 and 16), both of which were identified as mixtures by all laboratories (Table 2). Four

non-saliva/non-semen/non-human samples were included, and most laboratories (Table 2) obtained the expected result, which was the lack of any mRNA signal from the putative semen and saliva markers. However, sporadic single marker signals of unknown origin were obtained from the multiplex analysis by a small number of laboratories, mainly from the vaginal swab

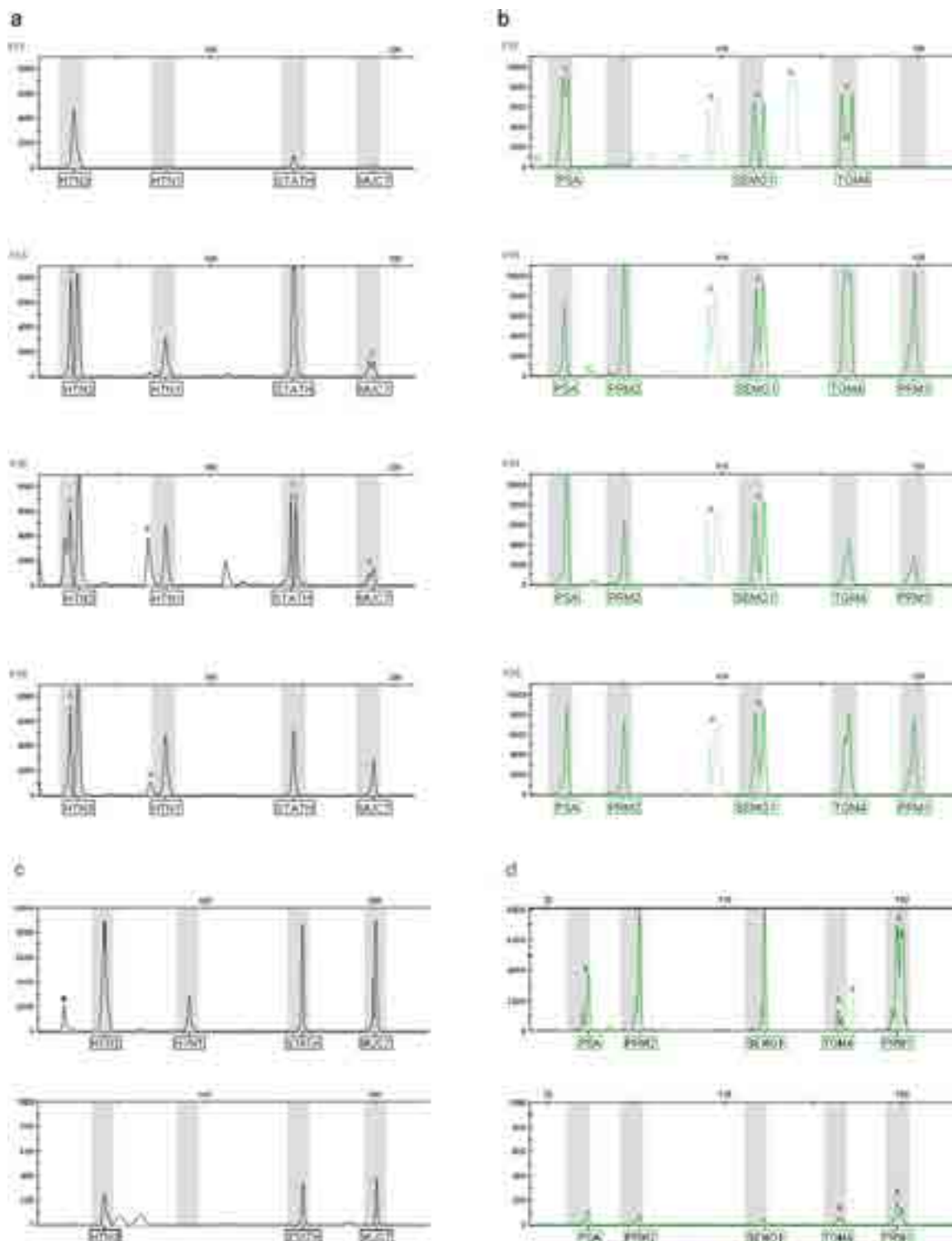


Fig. 1. Representative electropherograms of saliva (a) and semen (b) stains, a saliva dilution series (c: 5, 0.05 µl saliva) and a semen dilution series (d: 5, 0.01 µl semen) analyzed with the saliva triplex and the semen pentaplex. The saliva stains were 5 µl saliva on paper (#11), 5 µl saliva on glass slide (#14), mixture of 5 µl saliva and 1 µl semen on a swab (#16), licked plastic spoon (#19); the semen stains were 3 µl azoospermic semen on a pad (#12), 2 µl semen on toilet paper (#15), mixture of 5 µl saliva and 1 µl semen on a swab (#16), 3 µl semen inside latex glove (#20) (results from laboratory 12). The dilution series results were post PCR purified (results from laboratory 11). Some samples were saturated resulting in split peaks (s) and typical overflow peaks (*, dimmed) around 10 bp in front of the main peaks [21].

(Table 2). Representative electropherograms of saliva and semen stains are shown in Fig. 1a and b. HTN3, STATH and MUC7 were detected down to 0.5 µl saliva by more than half of the laboratories. A small number of laboratories were able to detect the saliva markers in as little as 0.05 µl saliva (Table 3 and Fig. 1c). The semen markers PSA, PRM2, SEMG1, TGM4 and PRM1 were detectable down to 0.5 µl semen by more than half of the laboratories. A small number of laboratories detected the semen markers in as little as 0.01 µl semen (Table 3 and Fig. 1d). In part 2, some markers were also tested as singleplexes on the stains and the dilution series. Since the differences between multiplex and singleplex testing were marginal, singleplex results are not shown nor discussed further.

No expression was detected for any of the markers in the RT minus (no reverse transcriptase added) or PCR negative controls by a majority of the laboratories. Laboratories 18 (only in part 1, not in part 2) and 19 (participated only in part 2) observed expression in RT minus controls and were therefore not included in data compilation.

Only four laboratories quantified the RNA in part 1: two used the Quant-iTTM RNA assay kit, one used the Quant-iTTM Ribo-Green[®] RNA Assay kit and one used the Bioanalyzer. Overall, the RNA quantification results were highly variable, probably because of the small stain sizes and the different quantification methods. Therefore, we did not attempt to determine the correlation (if any) between the quantification result and the RNA profiling success.

Post-PCR purification resulted in increased peak heights and reduced baseline signal noise but did not allow for detection of additional markers that were not observed prior to post-PCR purification. Post-PCR purification could be critical for the interpretation of the semen pentaplex results with dye blobs at marker-specific positions. Fifteen laboratories in part 1 and 11 laboratories in part 2 performed post-PCR purification.

3.2. DNA results

Fourteen and eleven laboratories (parts 1 and 2, respectively) performed RNA/DNA co-extraction. No specification of DNA analysis was provided. Various STR typing kits and cycle numbers were utilized by the participating laboratories (Table S1). Thus, as a

result of the varying conditions used, direct sensitivity and success rate comparisons could not be made. However, the results demonstrated that DNA of sufficient quantity and quality for STR analysis can be simultaneously extracted with RNA from small amounts of dried saliva and semen (Table 4). Full STR profiles were obtained from all human donor stains by most laboratories (data not shown). The obtained genotypes were confirmed by comparison with the reference profiles. Full profiles were obtained from the saliva dilution series C from 0.5 µl saliva stains, whereas partial profiles were obtained down to 0.1 µl saliva (Table 4). The quality of the STR typing results of saliva dilution series B was slightly lower, possibly resulting from donor-specific differences. From the semen dilution series A and D, full STR profiles were obtained from as little as 0.05 µl semen stains, and partial STR profiles were obtained from 0.01 µl semen (Table 4).

3.3. RNA and DNA results of the optional stain samples

The laboratories were invited to analyze additional samples including *bona fide* and mock casework samples, non-saliva, non-semen and non-human body fluid samples. Eighteen saliva stains and 18 semen stains were analyzed by 7 different laboratories. At least one of the specific markers was detectable in almost all saliva and semen stains, including various sizes, storage periods and storage conditions (Table 5). Representative RNA and DNA electropherograms are shown in Figs. S1 and S2. With saliva stains, only an outdoor saliva sample (covered, 7d) showed no result at all, probably due to degradation or absence of genetic material from exposure to outdoor conditions. Two of the semen stains were negative, probably because of low input amount, difficult starting material or questionable semen contribution.

Twenty-one non-saliva samples (4 blood, 6 semen, 8 vaginal swabs and 3 others) were analyzed with the saliva triplex by 8 laboratories. Five 'cross reactive' peaks (all <1000 RFUs) were observed, again mainly in vaginal swabs (Table 5). Twenty-one non-semen samples (4 blood, 8 saliva, 7 vaginal swabs and 2 others) were analyzed with the semen pentaplex by 8 laboratories. Only 1 'cross reactive' peak was observed (SEMG1 in a blood sample). Full autosomal STR profiles were obtained for most samples where an RNA/DNA co-extraction was performed (data not shown).

Table 4

DNA profiling results from 14 laboratories (part 1) and 11 laboratories (part 2) for 2 saliva and 2 semen dilution series (A, B from part 1; C, D from part 2) amplified with various STR typing kits. Only the 10 SGM Plus loci were taken into account. Dark gray squares represent stains/loci that were detected by more than half of the laboratories, light gray squares are stains/loci that were detected by 25–50% of the laboratories and white squares represent stains/loci that were detected by less than 25% of the laboratories.

saliva dilution series B

	vWA	D16	D2	AML	D8	D21	D18	D19	TH01	FGA	D3
10 ul	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
5 ul	13/14	12/14	12/14	13/14	13/14	13/14	12/14	13/14	13/14	13/14	13/14
1 ul	10/14	7/14	6/14	11/14	9/14	9/14	6/14	9/14	11/14	7/14	8/14
0.5 ul	8/14	6/14	2/14	10/14	6/14	5/14	5/14	9/14	10/14	5/14	6/14
0.1 ul	1/14			2/14	1/14	2/14	1/14				2/14
0.05 ul	1/14	1/14		2/14			1/14	1/14			1/14
0.01 ul				1/14							

saliva dilution series C

	vWA	D16	D2	AML	D8	D21	D18	D19	TH01	FGA	D3
	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11
	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11
	10/11	9/11	8/11	10/11	10/11	9/11	10/11	9/11	9/11	9/11	10/11
	9/11	9/11	8/11	10/11	10/11	9/11	9/11	9/11	10/11	9/11	9/11
	3/11	2/11	1/11	5/11	3/11	1/11	2/11	2/11	2/11	1/11	5/11
				2/11							
	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

semen dilution series A

	vWA	D16	D2	AML	D8	D21	D18	D19	TH01	FGA	D3
5 ul	14/14	14/14	14/14	14/14	14/14	14/14	13/14	14/14	14/14	13/14	14/14
1 ul	14/14	12/14	13/14	13/14	14/14	12/14	12/14	12/14	12/14	12/14	12/14
0.5 ul	14/14	13/14	13/14	14/14	14/14	13/14	13/14	13/14	13/14	13/14	13/14
0.1 ul	12/14	11/14	11/14	11/14	11/14	12/14	10/14	12/14	12/14	12/14	12/14
0.05 ul	12/14	8/14	8/14	11/14	10/14	8/14	7/14	9/14	10/14	9/14	8/14
0.01 ul	9/14	5/14	8/14	8/14	8/14	7/14	7/14	6/14	8/14	7/14	8/14

semen dilution series D

	vWA	D16	D2	AML	D8	D21	D18	D19	TH01	FGA	D3
	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11
	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11
	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11	11/11
	9/11	6/11	4/11	9/11	10/11	9/11	7/11	9/11	5/11	7/11	7/11
	8/11	4/11	3/11	6/11	8/11	6/11	5/11	5/11	4/11	5/11	3/11
	3/11				2/11	1/11			1/11		

n.d., not determined.

Table 5

mRNA profiling results of the optional stains that were analyzed by the participating laboratories (casework or mock samples, non-saliva/non-semen samples, old stains and environmentally exposed stains). Numbers are peak heights in RFUs, light gray squares represent RNA peaks <1000 RFUs, dark gray squares >1000 RFUs.

additional saliva stains or non-saliva samples

lab	saliva samples	HTN3	STATH	MUC7
5	wet swab of coffee mug (area of mouth contact)	8280	0	0
5	wet swab of water bottle (area of mouth contact)	7837	0	89
5	wet swab of mouth speaker office phone	0	1117	0
7	defrost sperm + saliva	9566	9469	9985
7	saliva E	9302	9309	9221
7	saliva A	9667	9675	9538
7	saliva N	9556	9494	9300
7	saliva S	9616	9626	9256
7	saliva E 1 µl	2080	429	928
7	saliva E 5 µl	9710	4275	7269
7	saliva E 10 µl	7610	4261	9722
11	saliva, outside covered, 1 day	8644	9349	3072
11	saliva, outside covered, 7 days	0	0	0
11	saliva, stored at 37°C for 3 months	9336	675	8965
11	saliva, stored at 56°C for 1 month	9064	9458	8732
11	saliva on denim, stored at RT for 6 months	0	0	5255
16	cigarette butt (only filter, 2 mm wide)	1169	258	0
17	mixed stain (menstrual blood, saliva) on swab	6214	3273	9540

additional semen stains or non-semen samples

lab	semen samples	PSA	PRM2	SEMG1	TGM4	PRM1
5	0.9cm2 area from panties, sexual assault (PSA pos.)	1835	8899	6610	0	8231
7	sperm on paper	0	0	2248	0	0
7	defrost sperm	0	0	46	0	0
7	defrost sperm + saliva	0	0	200	0	0
9	semen from human 3170	139	2221	3479	2677	8092
9	semen from human 3280	0	859	1079	523	3570
9	semen from human 5302	108	4846	1803	1487	7941
11	semen, outside covered, 1 day	4689	8583	7493	5683	8236
11	semen, outside covered, 7 days	0	0	0	0	0
11	semen stored at 37°C for 3 months	7889	7837	8501	7486	7821
11	semen stored at 56°C for 1 month	4589	8375	8266	6009	8289
11	semen on denim, stored at RT for 6 months	0	8483	7412	5454	8502
13	mixed stain on fabric, stored at RT for 4 years	9791	32199	20902	24514	31647
16	1 µl prostate secretion	6809	0	1369	7307	0
16	vaginal swab + sperm (1 µl)	0	0	0	0	0
16	sperm in condom (1 µl 1:10 dil.)	1655	7885	8132	3616	7577
17	mixed stain (menstrual blood, semen) on swab	551	7437	6532	1051	7377
17	5 µl azoospermic semen stain, 2 years old	1660	0	6617	1337	0

lab	non-saliva samples	HTN3	STATH	MUC7
2	3 µl blood on FTA	0	0	0
11	50 µl blood on cotton	0	0	0
13	10 µl blood on swab, stored at RT for 1 year	0	0	0
14	blood	0	0	0
5	0.9cm2 area from panties, sexual assault (PSA pos.)	0	0	0
7	sperm on paper	0	0	0
7	defrost sperm	0	0	0
16	sperm in condom (1 µl 1:10 dil.)	0	0	0
16	1 µl prostate secretion	0	0	292
17	5 µl azoospermic semen stain, 2 years old	0	0	0
11	1/2 vaginal swab	0	0	0
16	vaginal swab	0	0	0
16	vaginal swab + Sperm (1 µl)	0	0	0
17	vaginal swab TSI 4 days	221	0	0
17	vaginal swab TSI ≥7 days	0	0	0
17	vaginal swab TSI ≥7 days	0	0	0
17	vaginal swab TSI 2 days	0	0	550
17	vaginal swab TSI 2 days	0	0	272
11	1/2 menstrual swab	0	0	0
13	mixed stain on fabric, stored at RT for 4 years	0	0	0
17	mixed stain (menstrual blood, semen) on swab	0	624	0

lab	non-semen samples	PSA	PRM2	SEMG1	TGM4	PRM1
2	3 µl blood on FTA	0	0	781	0	0
11	50 µl blood on cotton	0	0	0	0	0
13	10 µl blood on swab, stored at RT for 1 year	0	0	0	0	0
14	blood	0	0	0	0	0
5	wet swab of coffee mug (area of mouth contact)	0	0	0	0	0
5	wet swab of water bottle (area of mouth contact)	0	0	0	0	0
5	wet swab of mouth speaker office phone	0	0	0	0	0
7	saliva E	0	0	0	0	0
7	saliva A	0	0	0	0	0
7	saliva N	0	0	0	0	0
7	saliva S	0	0	0	0	0
16	cigarette butt (only filter, 2 mm wide)	0	0	0	0	0
11	1/2 vaginal swab	0	0	0	0	0
16	vaginal swab	0	0	0	0	0
17	vaginal swab TSI 4 days	0	0	0	0	0
17	vaginal swab TSI ≥7 days	0	0	0	0	0
17	vaginal swab TSI ≥7 days	0	0	0	0	0
17	vaginal swab TSI 2 days	0	0	0	0	0
17	vaginal swab TSI 2 days	0	0	0	0	0
11	1/2 menstrual swab	0	0	0	0	0
17	mixed stain (menstrual blood, saliva) on swab	0	0	0	0	0

3.4. Tissue- and species-specificity of the saliva and semen multiplexes

The organizing laboratory (Zürich) tested the saliva and semen multiplexes on total RNA from a commercially available human tissue panel, including 20 human tissues (Table S3). The saliva triplex only showed positive signals with thyroid and trachea. The semen pentaplex showed reactions only with prostate and testes (Table S3).

Three laboratories (Brussels, Glasgow, Copenhagen) tested the saliva and semen multiplexes on animal saliva and semen samples. The only animal samples where species cross-reactivity was observed were gorilla saliva (HTN3, STATH and MUC7 (Table 6)) and chimpanzee semen (TGM4 (Table 7)). Most animal samples showed positive reactions with at least one of the housekeeping genes, confirming the presence of RNA in these samples.

4. Discussion

The purpose of this exercise was to evaluate an RNA/DNA co-extraction strategy with challenging saliva and semen samples and two novel saliva- and semen-specific mRNA multiplexes (a saliva triplex with the saliva-specific markers HTN3, STATH and MUC7 as well as a semen pentaplex with the semen-specific markers PSA, PRM2, SEMG1, TGM4 and PRM1). The semen pentaplex permits the detection of both, sperm and seminal plasma, thereby allowing the

identification of both spermic and azoospermic semen. Co-extracted DNA was analyzed with various commercial STR typing kits.

Most participating laboratories successfully applied the method using their own laboratory equipment and different kits and chemicals for RNA only or RNA/DNA co-extraction, reverse transcription and PCR. The samples to be analyzed in this exercise were challenging because the laboratories did not know the tissue origin and all stains were low template samples. The differing sensitivities between the laboratories can be explained by the different extraction kits and reverse transcription biochemicals and the skill sets within the laboratories, since those laboratories with more experience in RNA analysis in general produced better results than the others. The data from a small number of laboratories was excluded due to technical problems not attributable to the mRNA profiling method.

The evaluated saliva and semen markers proved to be specific and sensitive and therefore suitable for forensic stains, even aged and low template samples. "Specificity" in the context of gene expression and in reference to a specific body fluid can indicate either that expression of a particular gene is only observed in the body fluid of interest or can be used to describe markers whose expression is significantly higher in the body fluid of interest than in other body fluids. While the latter is perhaps more frequently observed and presents a greater challenge in terms of interpretation, proper interpretation guidelines could be developed to still permit

Table 6

Saliva triplex results of animal saliva samples. In addition, the expression of 3 housekeeping genes (B2M, UBC, UCE) was analyzed to confirm successful analysis despite negative saliva results. Numbers are peak heights in RFUs, positive results (>100 RFUs) are highlighted by shading.

lab	Animal	HTN3	STATH	MUC7	B2M	UBC	UCE
10	Bearded dragon	0	0	0	0	0	0
10	Bull	0	0	0	0	0	0
10	Cat	0	0	0	0	0	1851
9	Cat	0	0	0	n.d.	n.d.	n.d.
10	Chicken	0	0	0	0	0	0
10	Chinese water dragon	0	0	0	0	0	522
9	Cow	0	0	0	n.d.	n.d.	n.d.
10	Degu	0	0	0	0	0	2014
10	Dog	0	0	0	0	0	0
9	Dog	0	0	0	n.d.	n.d.	n.d.
10	Duck	0	0	0	0	0	378
10	Gerbil	0	0	0	0	0	0
10	Goat	0	0	0	0	0	404
10	Gorilla	9259	8890	454	9481	366	651
10	Guinea pig	0	0	0	0	0	0
10	Horse	0	0	0	0	0	0
9	Horse	0	0	0	n.d.	n.d.	n.d.
10	Jackdaw	0	0	0	0	0	1061
10	Leguan	0	0	0	0	0	498
10	Miniature horse	0	0	0	63	641	0
10	Pig	0	0	0	0	0	4052
10	Pigeon	0	0	0	0	0	7733
10	Rabbit	0	0	0	0	0	0
10	Human	7878	6405	7748	9331	4002	156

n.d., not determined.

an accurate identification of the presence of a particular biological fluid. The purpose of this study was not to develop such guidelines (hopefully it will be the subject of future EDNAP studies once markers for all relevant body fluids have been evaluated), but to simply begin to critically evaluate the sensitivity and specificity of reported semen- and saliva-specific markers and to initially demonstrate their potential future implementation in forensic casework.

Table 7

Semen pentaplex results of animal semen samples. In addition, the expression of 3 housekeeping genes (B2M, UBC, UCE) was analyzed to confirm successful analysis despite negative semen results. Numbers are peak heights in RFUs, positive results (>100 RFUs) are highlighted by shading, asterisks indicate saturation.

lab	Animal	PSA	PRM2	SEMG1	TGM4	PRM1	B2M	UBC	UCE
9	Chimpanzee	0	0	0	2527	0	n.d.	n.d.	n.d.
19	Cow	0	0	0	0	0	0	0	0
19	Deer	0	0	0	0	0	0	0	1187
9	Dog	0	0	0	0	0	n.d.	n.d.	n.d.
19	Dog 1	0	0	0	0	0	0	0	4151
19	Dog 2	0	0	0	0	0	0	0	0
19	Dog 3	0	0	0	0	0	541	0	3355
19	Goat 1	0	0	0	0	0	1627	0	0
19	Goat 2	0	0	0	0	0	3374	2650	0
19	Horse 1	0	0	0	0	0	0	0	1055
19	Horse 2	0	0	0	0	0	0	0	946
19	Horse 3	0	0	0	0	0	10,492*	5149	2535
9	Pig	0	0	0	0	0	n.d.	n.d.	n.d.
19	Pig 1	0	0	0	0	0	0	0	0
19	Pig 2	0	0	0	0	0	106	0	0
19	Pig 3	0	0	0	0	0	0	0	0
19	Sheep 1	0	0	0	0	0	0	0	0
19	Sheep 2	0	0	0	0	0	0	0	0
19	Human	8945*	8556*	9330*	9119*	8851*	10,331*	10,164*	10,470*

n.d., not determined.

With the examined saliva and semen markers, single cross-reactive peaks were observed in non-target body fluids. Most of them appeared only once and may be characterized as outliers. The only reproducible finding of putative cross reactivity was a MUC7 transcript in some vaginal samples. Mucins are high molecular weight glycosylated proteins that form a major part of a protective biofilm on the surface of epithelial cells, where they can provide a barrier to particulate matter and bind microorganisms. MUC4 was originally described as vaginal marker, but has been found to cross-react with saliva [16,24]. Therefore, some cross-reactivity of MUC7 in vaginal samples is not surprising, particularly because vaginal swabs normally contain large quantities of epithelial cells. Additionally, in this study many laboratories did not perform an RNA quantification prior to reverse transcription resulting in the input of likely increased amounts of RNA. Since some markers will have lower level expression in non-target body fluids, it is critical that suitable input RNA levels are utilized in order to reduce the occurrence of such cross reactivity. The use of RNA quantification prior to the reverse transcription reaction would allow for a standardization of the input of RNA into the reaction thus possible resulting in a reduction in the occurrence of such cross reactivity. However, it may be premature at this stage to make a definitive determination of whether MUC7 will have a role in the identification of saliva. It may be possible, for example, to use MUC7 in combination with other markers in order to develop an expression profile for saliva.

The multiplex PCR systems need some further optimization prior to routine use in casework analysis because of multiplex design issues (e.g. split peaks, noisy baseline). However, the multiplexes used in these initial studies were simply designed to permit more efficient analysis compared to singleplex reactions (reducing time and cost for participating laboratories) and therefore did not undergo a full developmental validation before their use in the study. As a result, some of these design issues are expected. The use of post-PCR purification improved the quality of the multiplex results (reduced appearance of split peaks, dye blobs, etc.) and increased peak heights from low level samples. Therefore, it is possible that post-PCR purification could be included as part of the standard protocol in order to resolve quality issues without requiring significant labor-intensive multiplex optimization experiments.

RNA quantification was performed by 4 laboratories, but the results were not consistent, probably because of the low template samples and/or insufficient sensitivity of the quantification methods. None of the suggested RNA quantification systems are ideal as they are not human-specific. The Bioanalyzer normally attests poor RNA quality and low RNA Integrity Numbers (RIN) with forensic samples [15]. One of the 4 laboratories used an RNA-carrier before RNA quantification which could have interfered with the results.

The saliva and semen multiplexes confirmed body fluid- and species-specificity when tested on human tissues and animal samples. Positive reactions of the saliva triplex with thyroid and trachea and the semen pentaplex with prostate and testes is not surprising, since these tissues are associated with the respective body fluids. Both multiplexes showed positive reactions with primate samples. Cross-reactivity with primate samples is not entirely unexpected due to the close evolutionary relatedness to humans, but the presence of primate biological material is not likely to be found in forensic cases.

HKG expression could not be shown for all animal samples to confirm successful analysis despite a negative body fluid result. The 3 housekeeping gene markers were designed for human use, but the sequences are very similar in mammals. Failure to detect HKG expression could therefore result from species specificity of the HKG primers, low input amount or the lack of suitable quality

RNA. The results are consistent with the saliva and semen markers having a high degree of specificity for humans/primates.

The following negative controls were included in RNA analysis: (1) RT minus (no reverse transcriptase added) to identify possibly contaminating DNA (frequently a larger size than the expected RNA product) or the presence of pseudogenes (same size as RNA product); (2) amplification blank. Laboratories 18 and 19 encountered problems with negative controls and therefore their results were not used in this study. In the evaluated multiplexes, RT minus controls should be negative because of prudent primer-design and no known pseudogenes. Contamination could happen during PCR (e.g. contamination of the mastermix with template) or during CE (e.g. re-using sample septa or buffer septa, crosstalk or carryover [32]).

The possibility of co-extracting RNA and DNA from the same stain sample is an important advantage since the amount of sample is often limited in forensic casework. The quantity and quality of DNA from co-extracted samples seemed to be sufficient for both casework and environmentally exposed samples. From almost all stains, good quality DNA profiles and the positive identification of saliva and semen could be achieved; even stains exposed to un-controlled humidity and old stains could be identified correctly. The use of RNA/DNA co-extraction methods presents an interesting opportunity to evaluate possible quantitative relationships between DNA and RNA from the same sample. However, such an evaluation could not be performed in the present study since there was variation in the chemistries, instruments and protocols that were used by the various laboratories, particularly for DNA where any STR kit could be used (i.e. no standard conditions between laboratories). Additionally, many laboratories did not perform RNA quantification and the current RNA quantification systems that are available may be influenced by contaminating residual DNA or endogenous bacteria present in some fluids. Therefore, this study, as was originally intended, simply affirms that a co-extraction strategy can be successfully utilized to obtain RNA and STR profiling results from individual samples. Future evaluation and validation of such an approach could indeed determine if quantitative relationships between DNA and RNA can be made in order to assist in interpretation of results.

In summary, the results of this study support an RNA/DNA co-extraction strategy allowing for positive identification of the tissue/fluid source of origin (saliva and semen) by mRNA profiling as well as a simultaneous identification of the body fluid donor by STR profiling. The evaluated 3 saliva and 5 semen markers proved to be robust, reproducible and sensitive. Co-extracted DNA from the same stain provided good-quality STR profiles. For future body fluid identification systems, multiplexes for the simultaneous identification of several body fluids and possibly tissues would be preferable. A subsequent EDNAP mRNA exercise will include an evaluation of mRNA markers for the identification of menstrual blood and housekeeping genes. The collective knowledge gained from this series of collaborative exercises is likely to facilitate the formulation of recommended practices and procedures for mRNA profiling for body fluid identification. A major task will be the interpretation of RNA results, especially of partial and mixed RNA profiles. This will be the subject of future EDNAP exercises, once all markers and possibly multiplexes have been evaluated.

Acknowledgements

Special thanks for support and technical assistance to C. Bouakaze^c, G. Förster^g, S. Graziano (Department of Therapeutic Research and Medicines Evaluation, Istituto Superiore di Sanità, Rome, Italy), D. Niederwieser^o, S.S. Tobe^u and M. Vennemann^u. We are grateful to all volunteers who provided samples for this study. We also would like to thank I. Jennes (from “Vlaams Opvang

Centrum voor vogels en wilde dieren”) and F. Vercammen (from “Zoo Antwerpen”) for kindly providing saliva from exotic animals.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2012.10.011>.

References

- [1] M. Bauer, A. Kraus, D. Patzelt, Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction, *J. Forensic Sci.* 44 (1999) 1232–1236.
- [2] M. Bauer, D. Patzelt, Evaluation of mRNA markers for the identification of menstrual blood, *J. Forensic Sci.* 47 (2002) 1278–1282.
- [3] M. Bauer, D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains, *Int. J. Legal Med.* 117 (2003) 175–179.
- [4] J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, *Forensic Sci. Int.* 135 (2003) 85–96.
- [5] G. Ferri, C. Bini, S. Ceccardi, S. Pelotti, Successful identification of two years old menstrual blood stain by using MMP-11 shorter amplicons, *J. Forensic Sci.* 49 (2004) 1387.
- [6] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, *Forensic Sci. Int.* 152 (2005) 1–12.
- [7] C. Nussbaumer, E. Gharehbaghi-Schnell, I. Korschneck, Messenger RNA profiling: a novel method for body fluid identification by real-time PCR, *Forensic Sci. Int.* 157 (2006) 181–186.
- [8] R. Fang, C.F. Manohar, C. Shulze, M. Brevnov, A. Wong, O.V. Petruskane, P. Brzoska, M.R. Furtado, Real-time PCR assays for the detection of tissue and body fluid specific mRNAs, *Int. Congr. Ser.* 1288 (2006) 685–687.
- [9] T.L. Noreault-Conti, E. Buel, The use of real-time PCR for forensic stain identification, *Promega, Profiles in DNA* 10 (1) (2007) 3–5.
- [10] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, *J. Forensic Sci.* 52 (2007) 1252–1262.
- [11] M. Bauer, RNA in forensic science, *Forensic Sci. Int. Genet.* 1 (2007) 69–74.
- [12] M. Bauer, D. Patzelt, Identification of menstrual blood by real time RT-PCR: technical improvements and the practical value of negative results, *Forensic Sci. Int.* 174 (2008) 54–58.
- [13] D. Zubakov, E. Hanekamp, M. Kokshoorn, W. van Ijcken, M. Kayser, Stable RNA markers for identification of blood and saliva stains from whole genome expression analysis of time-wise degraded samples, *Int. J. Legal Med.* 122 (2008) 135–142.
- [14] C. Haas, B. Klessner, C. Maake, W. Bär, A. Kratzer, mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, *Forensic Sci. Int. Genet.* 3 (2009) 80–88.
- [15] C. Haas, C. Muheim, A. Kratzer, W. Bär, C. Maake, mRNA profiling for the identification of sperm and seminal plasma, *Forensic Sci. Int. Genet. Suppl. Ser.* 2 (2009) 534–535.
- [16] C. Cossu, U. Germann, A. Kratzer, W. Bär, C. Haas, How specific are the vaginal secretion markers HBD1 and MUC4? *Forensic Sci. Int. Genet. Suppl. Ser.* 2 (2009) 536–537.
- [17] H. Nakanishi, A. Kido, T. Ohmori, A. Takada, M. Hara, N. Adachi, K. Saito, A novel method for the identification of saliva by detecting oral streptococci using PCR, *Forensic Sci. Int.* 183 (2009) 20–23.
- [18] R.I. Fleming, S. Harbison, The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids, *Forensic Sci. Int. Genet.* 4 (2010) 244–256.
- [19] R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, *Forensic Sci. Int. Genet.* 4 (2010) 311–315.
- [20] H. Nakanishi, T. Ohmori, M. Hara, A. Takada, H. Shoji, N. Adachi, K. Saito, A simple identification method of saliva by detecting *Streptococcus salivarius* using loop-mediated isothermal amplification, *J. Forensic Sci.* 56 (2011) S158–S161.
- [21] C. Haas, E. Hanson, A. Kratzer, W. Bär, J. Ballantyne, Selection of highly specific and sensitive mRNA biomarkers for the identification of blood, *Forensic Sci. Int. Genet.* 5 (2011) 449–458.
- [22] J. Wobst, R. Banemann, I. Bastisch, RNA can do better – an improved strategy for RNA-based characterization of different body fluids and skin, *Forensic Sci. Int. Genet. Suppl. Ser.* 3 (2011) e421–e422.
- [23] M.L. Lennard Richard, K.A. Harper, R.L. Craig, A.J. Onorato, J.M. Robertson, J. Donfack, Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis, *Forensic Sci. Int. Genet.* (2011), <http://dx.doi.org/10.1016/j.fsigen.2011.09.007>.
- [24] G. Hadžić, A. Lukan, K. Drobnič, Practical value of the marker MUC4 for identification of vaginal secretion in penile swabs, *Forensic Sci. Int. Genet. Suppl. Ser.* 3 (2011) e222–e223.
- [25] M. Bauer, S. Polzin, D. Patzelt, Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of blood stains? *Forensic Sci. Int.* 138 (2003) 94–103.
- [26] M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen and saliva stains, *J. Forensic Sci.* 53 (2008) 296–305.

- [27] D. Zubakov, M. Kokshoorn, A. Kloosterman, M. Kayser, New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains, *Int. J. Legal Med.* 123 (2009) 71–74.
- [28] F. Kohlmeier, P.M. Schneider, Successful mRNA profiling of 23 years old blood stains, *Forensic Sci. Int. Genet.* 6 (2012) 274–276.
- [29] C. Haas, E. Hanson, W. Bär, R. Banemann, A.M. Bento, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, A. Choma, M. Dötsch, M. Durianciková, P. Hoff-Olsen, C. Hohoff, P. Johansen, P.A. Lindenbergh, B. Lodenkötter, B. Ludes, O. Maroñas, N. Morling, H. Niederstätter, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviežená, L. Zatkalíková, J. Ballantyne, mRNA profiling for the identification of blood – results of a collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* 5 (2011) 21–26.
- [30] C. Haas, E. Hanson, M.J. Anjos, W. Bär, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, V. Castella, A. Choma, G. De Cock, M. Dötsch, P. Hoff-Olsen, P. Johansen, F. Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maroñas, D. Moore, M.L. Møller, N. Morling, H. Niederstätter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviežená, M. Turanská, L. Zatkalíková, J. Ballantyne, RNA/DNA co-analysis from blood stains – results of a second collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* 6 (2012) 70–80.
- [31] P.J. Smith, J. Ballantyne, Simplified low-copy-number DNA analysis by post-PCR purification, *J. Forensic Sci.* 52 (2007) 820–829.
- [32] Maximizing the Performance of Capillary Electrophoresis Systems, Applied Biosystems, Forensic News, 2010 July.

RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: results of a fourth and fifth collaborative EDNAP exercise

Haas C, Hanson E, Anjos MJ, Ballantyne KN, Banemann R, Bhoelai B, Borges E, Carvalho M, Courts C, De Cock G, Drobic K, Dötsch M, Fleming R, Franchi C, Gomes I, Hadzic G, Harbison SA, Harteveld J, Hjort B, **Hollard C**, Hoff-Olsen P, Hüls C, Keyser C, Maroñas O, McCallum N, Moore D, Morling N, Niederstätter H, Noël F, Parson W, Phillips C, Popielarz C, Roeder AD, Salvaderi L, Sauer E, Schneider PM, Shanthan G, Court DS, Turanská M, van Oorschot RA, Vennemann M, Vidaki A, Zatkalíková L, Ballantyne J.

Forensic Sci Int Genet. 2014 Jan;8(1):203-12



RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: Results of a fourth and fifth collaborative EDNAP exercise



C. Haas^{a,*}, E. Hanson^b, M.J. Anjos^k, K.N. Ballantyne^w, R. Banemann^r, B. Bhoelai^e, E. Borges^j, M. Carvalho^k, C. Courts^h, G. De Cock^d, K. Drobnić^p, M. Dötsch^r, R. Fleming^f, C. Franchi^x, I. Gomesⁿ, G. Hadzic^p, S.A. Harbison^f, J. Hartevelde^e, B. Hjort^v, C. Hollardⁱ, P. Hoff-Olsen^g, C. Hüls^l, C. Keyserⁱ, O. Maroñas^t, N. McCallum^m, D. Moore^q, N. Morling^v, H. Niederstätter^o, F. Noël^d, W. Parson^o, C. Phillips^t, C. Popielarz^j, A.D. Roeder^s, L. Salvaderi^y, E. Sauer^h, P.M. Schneiderⁿ, G. Shanthan^g, D. Syndercombe Court^c, M. Turanská^u, R.A.H. van Oorschot^w, M. Vennemann^l, A. Vidaki^c, L. Zatkalíková^u, J. Ballantyne^b

^a Institute of Legal Medicine, University of Zurich, Switzerland

^b National Center for Forensic Science, University of Central Florida, Orlando, USA

^c Department of Forensic and Analytical Science, King's College London, UK

^d National Institute for Criminalistics and Criminology, Brussels, Belgium

^e Netherlands Forensic Institute, The Hague, The Netherlands

^f Institute of Environmental Science and Research Ltd., Auckland, New Zealand

^g Department of Forensic Genetics, Norwegian Institute of Public Health, Oslo, Norway

^h Institute of Legal Medicine, University of Bonn, Germany

ⁱ Institute of Legal Medicine, University of Strasbourg, France

^j Institut National de Police Scientifique, Laboratoire de Police Scientifique de Lyon, France

^k Forensic Genetic and Biology Service, National Institute of Legal Medicine and Forensic Sciences, Coimbra, Portugal

^l Institute of Legal Medicine, Medical School Hannover, Germany

^m Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, Scotland, UK

ⁿ Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Germany

^o Institute of Legal Medicine, Innsbruck Medical University, Austria

^p National Forensic Laboratory, Ljubljana, Slovenia

^q Innovation and Development Team, LGC Forensics, Teddington, UK

^r Bundeskriminalamt, Wiesbaden, Germany

^s Orchid Cellmark Ltd., Abingdon, UK

^t Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain

^u Institute of Forensic Science, Ministry of the Interior, Department of Biology and DNA Analysis, Slovenska Lupca, Slovakia

^v Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

^w Victoria Police Forensic Services Department, Melbourne, Victoria, Australia

^x RIS Carabinieri, Sezione di Biologia di Roma, Italy

^y NGB Genetics di Bologna, Italy

ARTICLE INFO

Article history:

Received 2 September 2013

Accepted 28 September 2013

Keywords:

Forensic science

Body fluid identification

Vaginal secretion

Menstrual blood

ABSTRACT

The European DNA Profiling Group (EDNAP) organized a fourth and fifth collaborative exercise on RNA/DNA co-analysis for body fluid identification and STR profiling. The task was to identify dried menstrual blood and vaginal secretion stains using specific RNA biomarkers, and additionally test 3 housekeeping genes for their suitability as reference genes. Six menstrual blood and six vaginal secretion stains, two dilution series (1/4–1/64 pieces of a menstrual blood/vaginal swab) and, optionally, *bona fide* or mock casework samples of human or non-human origin were analyzed by 24 participating laboratories, using RNA extraction or RNA/DNA co-extraction methods. Two novel menstrual blood mRNA multiplexes were used: MMP triplex (MMP7, MMP10, MMP11) and MB triplex (MSX1, LEFTY2, SFRP4) in conjunction with a housekeeping gene triplex (B2M, UBC, UCE). Two novel mRNA multiplexes and a HBD1 singleplex were

* Corresponding author at: Institute of Legal Medicine, Forensic Genetics, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

Tel.: +41 44 635 56 56; fax: +41 44 635 68 58.

E-mail address: cordula.haas@irm.uzh.ch (C. Haas).

Housekeeping genes mRNA profiling

used for the identification of vaginal secretion: Vag triplex (MYOZ1, CYP2B7P1 and MUC4) and a Lactobacillus-specific Lacto triplex (Ljen, Lcris, Lgas). The laboratories used different chemistries and instrumentation and all were able to successfully isolate and detect mRNA in dried stains. The simultaneous extraction of RNA and DNA allowed for positive identification of the tissue/fluid source of origin by mRNA profiling as well as a simultaneous identification of the body fluid donor by STR profiling, also from old and compromised casework samples. The results of this and the previous collaborative RNA exercises support RNA profiling as a reliable body fluid identification method that can easily be combined with current STR typing technology.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Conventional tests for the identification of blood, saliva and semen are based on enzymatic and immunological reactions of cell specific enzymes and antigens, but many of these tests lack specificity [1]. Recently two screening tests for the identification of menstrual blood were described, that utilize antibodies for fibrin–fibrinogen degradation products (latex agglutination test, D-dimer assay) [2,3]. Neither test has been confirmed and validated for forensic use. Up to now no reliable presumptive test for the identification of vaginal secretion exists [1]. The analysis of cell-specific mRNA expression is a confirmative method for the identification of body fluids [1,4–37]. The suitability of mRNA profiling assays has also been demonstrated for old and environmentally compromised samples [17,18,35]. Recently, three collaborative exercises were performed by the European DNA Profiling Group (EDNAP – <http://www.isfg.org/EDNAP>) in order to evaluate the robustness and reproducibility of mRNA profiling for blood, saliva and semen identification: (1) evaluation of three blood-specific markers (HBB, SPTB and PBGD) using singleplex reactions [38]; (2) evaluation of seven blood-specific markers using two multiplex systems, a ‘high sensitivity’ duplex (HBB, HBA) and a ‘moderate sensitivity’ pentaplex (ALAS2, CD3G, ANK1, PBGD and SPTB) [39] and (3) a saliva triplex including the markers HTN3, STATH and MUC7 and a semen pentaplex allowing the detection and differentiation of sperm (PRM1, PRM2) and seminal plasma (PSA, SEMG1 and TGM4), the latter of which is necessary for the identification of semen from azoospermic men [40]. Most laboratories, some of which had no prior experience with RNA, were able to successfully isolate and analyze RNA from the

provided samples. While sensitivity varied between laboratories, the method proved to be reproducible and sensitive using different analytical strategies.

A fourth and fifth collaborative exercise was organized by the Institute of Legal Medicine, University of Zürich, Switzerland, on behalf of EDNAP in order to test forensically suitable menstrual blood and vaginal secretion markers, respectively. In addition 3 housekeeping genes were evaluated. In a preliminary study performed by the Florida and Zürich laboratories, various menstrual blood and vaginal secretion mRNA markers described in the literature [11,19,23,34,36] were evaluated in terms of sensitivity, specificity and performance with casework samples. For this study, 7 menstrual blood, 7 vaginal secretion and 3 housekeeping gene markers were tested, 16 of which were deemed most suitable for forensic use (Table 1). Hs202072 was seldom detectable during pre-testing and therefore excluded from further analysis. In these exercises, the following multiplexes were developed and provided to the participating laboratories: (1) two menstrual blood multiplexes, MMP triplex (MMP7, MMP10, MMP11) and MB triplex (MSX1, LEFTY2, SFRP4) in conjunction with a housekeeping gene triplex (B2M, UBC, UCE); (2) two vaginal secretion triplexes, including human Vag triplex (MYOZ1, CYP2B7P1 and MUC4) and bacterial Lacto triplex (Ljen, Lcris, Lgas) and a HBD1 singleplex. These multiplexes were developed for ease of use in the exercises and have not been validated for routine use.

The exercises included the analysis of mock casework samples (2 × 8), dilution series of menstrual blood and vaginal secretions and optional additional *bona fide* or mock casework samples from the participating laboratories. The housekeeping gene markers were included in order to evaluate their use as positive controls for

Table 1
List of evaluated mRNA markers for the identification of vaginal secretion, menstrual blood and housekeeping genes.

Gene		
<i>Menstrual Blood</i>		
MMP7	matrix metalloproteinase 7	} MMP triplex
MMP10	matrix metalloproteinase 10	
MMP11	matrix metalloproteinase 11	
MSX1	msh homeobox 1	} MB triplex
LEFTY2	left-right determination factor 2	
SFRP4	secreted frizzled-related protein 4	
Hs202072*	uncharacterized LOC100505776	
<i>Vaginal secretion</i>		
MYOZ1	myozenin-1	} Vag triplex
CYP2B7P1	cytochromeP450, family 2, subfamily B, polypeptide 7 pseudogene 1	
MUC4	mucin 4	
HBD1	human beta-defensin 1	} HBD1 singleplex
Ljen	Ljen [16s ribosomal RNA]	} Lacto triplex
Lcris	Lcris [16S-23S intergenic spacer region]	
Lgas	Lgas [16S-23S intergenic spacer region]	
<i>Housekeeping genes</i>		
B2M	Beta-2 microglobulin	} HKG triplex
UBC	Ubiquitin C	
UCE	Ubiquitin conjugating enzyme	

*Not sensitive.

successful RNA extraction/reverse transcription for future work. The expression of these housekeeping genes was additionally tested in blood, saliva and semen dilution series. Several housekeeping gene markers have already been tested on body fluids [19,24,32,37], but up to now, no marker has been described that is universally suitable for all body fluids. Saliva and semen normally show reduced expression of housekeeping genes, probably because spermatozooids contain little cytoplasm and few ribosomes and the desquamated cells of the buccal mucosa have almost no cell metabolism [19].

Here we present data from 20 and 23 laboratories participating in the fourth and fifth collaborative exercise on the identification of menstrual blood and vaginal secretion respectively.

2. Materials and methods

2.1. Samples and materials provided

The organizing laboratory (Institute of Legal Medicine, University of Zürich, Switzerland) sent one set of samples for each exercise to the participating laboratories.

Exercise 4 included stains 1–8 (menstrual blood, non-menstrual blood, Table 2), a menstrual blood dilution series (1/4–1/64 of a menstrual blood swab) and 3 dilution series of blood, semen and saliva (1–0.001 µl blood, 5–0.04 µl semen, 25–0.2 µl saliva). Menstrual blood from 6 different donors (for 6 stains and 1 dilution series) was collected on swabs or on sanitary towels (4 fresh samples and 3 that had been stored for up to 5 years at room temperature in the dark). Additional fresh samples not containing menstrual blood (blood, saliva, vaginal secretion) were provided by 3 female donors. Semen (stored frozen for 5 years, thawed and placed on swabs) was provided by 1 male donor. For stain 2, a quarter of a swab containing 40 µl EDTA-blood (corresponds to about 10 µl blood per stain) was provided. Stain 5 was a quarter of a vaginal secretion swab. For the blood, semen and saliva dilution series, samples were diluted in 0.9% NaCl to a final volume of 5 µl per sample and placed on swabs.

Exercise 5 included stains 9–16 (vaginal secretion, non-vaginal secretion, Table 2) and a vaginal secretion dilution series (1/4–1/64 of a vaginal secretion swab). Vaginal secretion from 6 different

donors (for 6 stains and 1 dilution series) was collected on swabs, on a sanitary towel or in underpants, 5 of them were fresh and 2 had been stored for up to 5 years at room temperature in the dark. Fresh non-vaginal secretion samples (urine, saliva) were provided by 2 female donors. Stain 11 was half of a swab that was immersed into liquid urine and stain 15 was half of a buccal swab.

In both exercises, the laboratories were asked, but not required, to examine additional samples: (1) *bona fide* or mock casework material that could include human and/or non-human menstrual blood and vaginal secretion stains or (2) other forensically relevant body fluids (e.g. blood, semen, saliva).

HPLC-purified primers were purchased from Microsynth (St. Gallen, Switzerland). The organizing laboratory prepared primer mixes for the multiplexes and aliquots were provided to the participating laboratories for use in amplification reactions (see below). The primers and samples were sent non-refrigerated by Fedex or DHL (exercise 4, arrival within 1–2 days, except 1 parcel that was 20 days in transit; exercise 5, arrival within 1–3 days, except 2 parcels that were 8 and 10 days in transit).

2.1.1. RNA/(DNA co-)extraction and reverse transcription

The participating laboratories were asked to use the entire swab or stain for extraction (RNA only or RNA/DNA co-extraction). The organizing laboratory provided an example protocol for extraction and reverse transcription. The laboratories could, however, use methods of their own choice. The example protocols were described previously [38,39].

2.1.2. RNA- and DNA-quantification

Participating laboratories were asked to quantify RNA using one of the following methods: Quant-iTTM RiboGreen[®] RNA Assay kit (Invitrogen) using a fluorescence microplate reader (high and low range protocol option) [11], Quant-iTTM RNA assay kit (Invitrogen) using the Qubit fluorometer [26] or Bioanalyzer (Agilent) [20]. Laboratories were able to quantify DNA using a quantification system of their own choice.

2.2. Endpoint PCR

The primer sets and amplicon lengths are shown in Table S1. The forward primers of the menstrual blood and vaginal secretion

Table 2

mRNA profiling results (MMP, MB, HKG, Vag, Lacto triplexes and HBD1 singleplex) from 24 laboratories for the 16 stains. Dark grey squares represent stains/markers that were detected by more than half of the laboratories, light grey squares are stains/markers that were detected by 25–50% of the laboratories and white squares represent stains/markers that were detected by <25% of the laboratories.

Stain	Sample	MMP triplex			MB triplex			HKG triplex		
		MMP11	MMP7	MMP10	MSX1	LEFTY2	SFRP4	B2M	UBC	UCE
1	Menstrual blood on sanitary towel (fresh)	19/20	18/20	20/20	10/20	20/20	3/20	20/20	20/20	19/20
3	1/4 menstrual blood swab (fresh)	20/20	19/20	18/20	18/20	16/20	15/20	20/20	20/20	20/20
4	Menstrual blood on sanitary towel (5 year old)	19/20	12/20	20/20	8/20	19/20	3/20	20/20	19/20	17/20
6	1/4 menstrual blood swab (5 year old, d1–4)	17/20	16/20	18/20	11/20	7/20	8/20	20/20	20/20	18/20
7	1/12 menstrual blood swab (5 year old, d1/d4)	9/20	7/20	8/20	11/20	6/20	2/20	18/20	18/20	13/20
8	1/4 menstrual blood swab (fresh)	4/20	9/20	11/20	2/20	0/20	0/20	20/20	20/20	20/20
5	1/4 vaginal swab (fresh)	1/20	10/20	4/20	12/20	1/20	0/20	20/20	20/20	20/20
2	1/4 swab with EDTA-blood (fresh)	5/20	1/20	1/20	0/20	0/20	0/20	20/20	20/20	19/20
Stain	Sample	Vag triplex			Lacto triplex			Singleplex		
		MYOZ1	CYP	MUC4	Ljen	Lcris	Lgas	HBD1		
9	1/4 vaginal swab (2 year old)	22/22	22/22	22/22	10/22	19/22	1/22	21/22		
10	5 × 5 mm from white worn underpant (fresh)	9/22	18/22	15/22	22/22	17/22	7/22	6/22		
12	1/2 vaginal swab, pregnant (fresh)	21/22	21/22	22/22	8/22	3/22	22/22	21/22		
13	1/4 vaginal swab (5 year old)	20/22	19/22	20/22	19/22	19/22	1/22	20/22		
14	1/2 vaginal swab (fresh)	22/23	22/23	22/23	23/23	2/23	21/23	20/23		
16	5 × 5 mm from sanitary towel (fresh)	1/23	5/23	2/23	13/23	0/23	1/23	7/23		
15	1/2 buccal swab (fresh)	6/23	2/23	7/23	17/23	3/23	2/23	7/23		
11	1/2 swab urine (fresh)	0/22	2/22	8/22	21/22	1/22	6/22	9/22		

markers were 5'-labeled with FAM. The HKG forward primers were 5'-labeled with the NED-analogue ATTO550.

The following amplification conditions were recommended: *MMP*, *MB*, *HKG*, *Vag*, *Lacto triplexes*: The 25 µl reaction mix contained 2 µl cDNA, 2.5 µl 10× primer mix (see below), 12.5 µl 2× Multiplex PCR Mastermix (Multiplex PCR kit, QIAGEN), 2.5 µl Q-Solution (Multiplex PCR kit; QIAGEN), 5.5 µl H₂O. *HBD1 singleplex*: The 25 µl reaction mix contained 5 µl cDNA, 2.5 µl 10× PCR Buffer II, 2.5 µl dNTPs (2.5 mM each), 3 µl MgCl₂ (25 mM), 2 µl 12.5× primer mix, 0.4 µl AmpliTaq Gold® (5 U/µl) and 9.6 µl H₂O (all singleplex reagents from Life Technologies). Sterile water was used in place of cDNA for no-template controls.

The 10× primer mixes were prepared using the following concentrations: *MMP triplex*: 2 µM each; *MB triplex*: MSX1 5 µM, SFRP4 5 µM, LEFTY2 2 µM; *Vag triplex*: 2 µM each; *Lacto triplex*: 2 µM each; *HKG triplex*: B2M 2 µM, UBC 2 µM, UCE 5 µM. The 12.5× primer mix for the *HBD1 singleplex* was 20 µM.

The cycling conditions were as follows: *MMP*, *MB*, *HKG*, *Vag*, *Lacto triplexes*: the initial denaturation was at 95 °C for 15 min, followed by 35 cycles of 94 °C 30 s, 55 °C (+0.2 °C per cycle) 90 s, 72 °C 40 s and final elongation at 72 °C for 30 min; *HBD1 singleplex*: The initial denaturation was at 95 °C for 11 min, followed by 35 cycles of 94 °C 20 s, 55 °C 60 s, 72 °C 45 s and final elongation at 72 °C for 30 min.

Post PCR purification was optional, but recommended, to eliminate dye blobs and to concentrate/dilute the samples by variation of the elution volume (e.g. MinElute PCR purification kit, QIAGEN) [26,41].

2.2.1. Capillary electrophoresis (CE)

The participating laboratories used standard, multi-color fluorescent genetic analyzers and standard electrophoresis conditions for the detection of the menstrual blood, vaginal secretion and housekeeping gene specific amplicons (Table S2). Any dye set that included FAM and NED, with associated internal lane standard, could be used. Raw data were analyzed with GeneScan® or Genemapper® Software (Life Technologies). For data compilation, the threshold for a positive result was set to 100 RFUs (relative fluorescence units) or 200 RFUs (Genetic Analyzer 3500 series) respectively, for consistency between laboratories and to avoid inclusion of false positive results (due to the presence of dye blobs and baseline noise).

2.2.2. DNA-amplification and -detection

If DNA was co-extracted, the laboratories were asked to use a standard STR typing kit, PCR and CE conditions of their choice (Table S2). A peak detection threshold of 50 RFUs was used.

3. Results

Laboratories 1–20 participated in exercise 4 and laboratories 1–9, 11–24 participated in exercise 5. Each laboratory was asked to complete a questionnaire describing the methods used (Table S2). The 24 laboratories used 6 different kits for RNA only or RNA/DNA co-extraction, 6 different reverse transcription kits, 9 different STR typing kits for DNA analysis, 7 different thermocyclers, 4 different genetic analyzers and 3 different polymers (Table S2). Since 3 packages experienced longer shipping times at outside/room temperature, primer mixes may have been impaired. Slightly reduced sensitivities were reported with the vaginal markers (8 and 10 days in transit), but no impairment was detected with the menstrual blood multiplexes (20 days in transit). Since no significant adverse effects were observed (slight reduction in sensitivity could be due to other contributing factors as well), results from these laboratories were included in the final analysis.

While unfortunate, the delays have provided some useful information regarding the robustness of the consumables and mRNA.

3.1. RNA results

The mRNA profiling results are summarized in Table 2. All 24 laboratories were able to perform the method successfully as demonstrated by detection of the RNA menstrual blood, vaginal secretion and housekeeping gene markers. No results for stains 9–13 were available for laboratory 7 due to incorrect manipulation of the stains, but results were obtained for all other samples in the exercises.

The 6 menstrual blood and 6 vaginal secretion stains were identified correctly by all laboratories (Table 2). The MMPs, MSX1 and LEFTY2 were detected in most menstrual blood samples (stains 1, 3–4, 6–8) by more than half of the laboratories, but SFRP4 was only detected in the high input menstrual blood samples 3 and 6. MYOZ1, CYP2B7P1, MUC4, HBD1 and at least one of the lactobacilli markers were reliably detected in the vaginal secretion samples (stains 9–10, 12–14, 16). Stains 8 (1/4 fresh menstrual blood swab) and 16 (5 × 5 mm from fresh sanitary towel) appeared to be challenging possibly due to donor specific light menstrual blood/vaginal secretion flow.

For stain 6, menstrual blood swabs were collected from the same donor on days 1–4 of menstruation, for stain 7 from another donor on days 1 and 4, all stored 5 years at room temperature in the dark. Several laboratories were given part of the same menstrual blood swab for analysis. We then evaluated the variation in marker detection amongst the samples from days 1–4 (Table S3). By day 3 and day 4 of menstruation some of the menstrual blood specific markers (MSX1, LEFTY2, SFRP4) tended to disappear, whereas the metalloproteinase markers (MMP7, MMP10, MMP11) were detected for most of the days of menstruation, with some (possible intra-individual) variation. Variation in gene expression is expected over the reported days of menstruation, especially towards the end of menstruation, as menstrual blood flow can vary significantly [42]. For stain 13, vaginal swabs were collected from the same donor on days 5, 7, 13, 20, 24 and 28 of the menstrual cycle (stored 5 years at room temperature in the dark). Some laboratories were given a quarter of the same vaginal swab for analysis (in one instance, 3 labs using one of the samples and 4 labs using another). The results were similar for all time points and markers and are therefore collated into one result (Table 2, stain 13). Representative electropherograms of menstrual blood and vaginal secretion stains are shown in Figs. 1 and 2.

In exercise 4, two non-menstrual blood samples (stain 2: blood; stain 5: vaginal secretion) and in exercise 5, two non-vaginal secretion samples (stain 11: urine; stain 15: saliva) were included in order to evaluate the specificity of the included menstrual blood and vaginal secretion markers. MMP7 and MSX1 were detected by more than half of the laboratories in stain 5 (1/4 vaginal swab), and MUC4, HBD1, MYOZ1, Ljen and Lgas were detected in stains 11 (1/2 urine swab) and 15 (1/2 buccal swab). Minor occurrences of cross reactivity were observed for the other markers by a small number of laboratories (Table 2). In general, the cross reactive peaks showed reduced signal intensities compared to those of menstrual blood and vaginal secretion samples (Table S4).

For the dilution series, the HKG 3plex markers were detected by most laboratories down to the smallest sample size (1/64 menstrual swab) (Table 3). The MMP markers, whereof MMP10 was the most sensitive, were detected by more than half of the laboratories down to the smallest sample size. MSX1 was detected by more than half of the laboratories in 1/8-sized

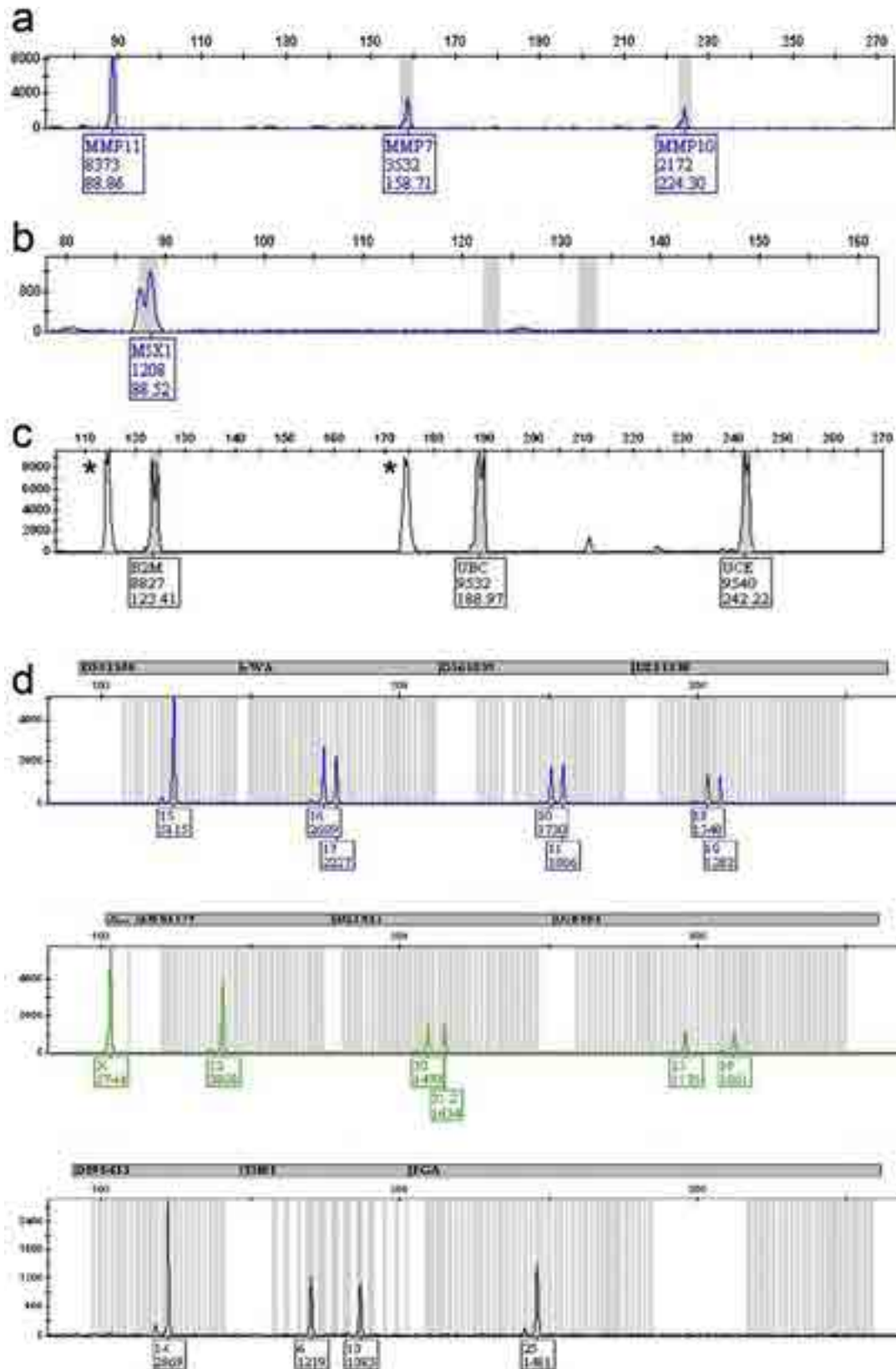


Fig. 1. Representative electropherograms from a menstrual blood sample (stain 3, 1/4 fresh menstrual blood swab, laboratory 15). RNA was analyzed with the MMP triplex (a), the MB triplex (b) and the HKG triplex (c). DNA was analyzed with the SGMplus kit (d). Overflow peaks are marked with an asterisk (*).

menstrual swabs, but only sporadically in smaller sample sizes. LEFTY2 and SFRP4 were the least sensitive markers and were detected only by a few laboratories (even in high input samples). MYOZ1, CYP2B7P1, MUC4, HBD1, Ljen and Lcris were also

detected by most laboratories down to the smallest sample size (1/64 vaginal swab) (Table 3). Lgas was only sporadically detected throughout the range of sample sizes tested, which might be due to the limited co-presence of different Lactobacilli

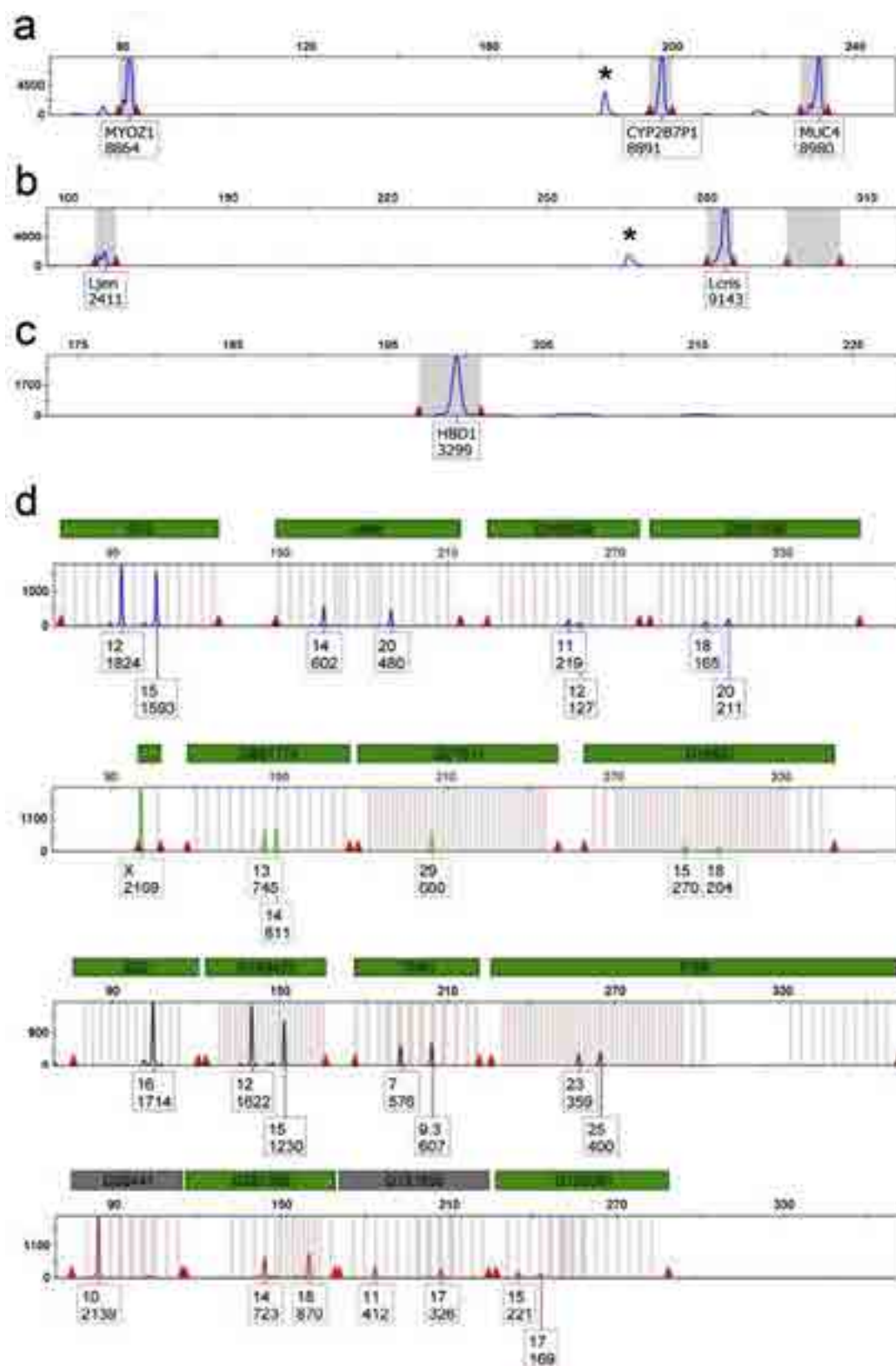


Fig. 2. Representative electropherograms from a vaginal secretion sample (stain 13, 1/4 vaginal swab (5 year old), laboratory 17). RNA was analyzed with the Vag triplex (a), the Lacto triplex (b) and the HBD1 singleplex (c). DNA was analyzed with the NGM kit (d). Overflow peaks are marked with an asterisk (*).

within one individual (see Section 4) rather than the detection sensitivity per se of the marker.

Only sporadic peaks were detected in the RT minus (no reverse transcriptase added) or PCR negative controls (data not

shown). In exercise 5, eight laboratories detected 1, 2 or all 3 Lacto markers in RT minus controls. In most, but not all, cases the peaks in the RT minus controls were considerably smaller than the respective peaks in the RT plus samples. Additionally,

Table 3

mRNA profiling results (MMP, MB, HKG, Vag, Lacto triplexes and HBD1 singleplex) from 24 laboratories for the menstrual blood and vaginal secretion dilution series. Dark grey squares represent stains/markers that were detected by more than half of the laboratories, light grey squares are stains/markers that were detected by 25–50% of the laboratories and white squares represent stains/markers that were detected by <25% of the laboratories.

Sample	MMP triplex			MB triplex			HKG triplex		
	MMP11	MMP7	MMP10	MSX1	LEFTY2	SFRP4	B2M	UBC	UCE
1/4	15/20	16/20	19/20	11/20	7/20	3/20	20/20	20/20	20/20
1/8	14/20	13/20	18/20	12/20	4/20	6/20	20/20	20/20	20/20
1/16	13/20	15/20	17/20	10/20	4/20	2/20	20/20	19/20	20/20
1/32	12/20	9/20	17/20	3/20	3/20	3/20	20/20	20/20	19/20
1/64	7/20	9/20	14/20	3/20	4/20	3/20	19/20	19/20	15/20

Sample	Vag triplex			Lacto triplex			Singleplex
	MYOZ1	CYP	MUC4	Ljen	Lcris	Lgas	HBD1
1/4	23/23	23/23	23/23	22/23	21/23	2/23	22/23
1/8	22/23	22/23	21/23	22/23	21/23	2/23	19/23
1/16	21/23	21/23	20/23	21/23	21/23	2/23	16/23
1/32	19/23	21/23	20/23	21/23	19/23	2/23	14/23
1/64	15/23	19/23	19/23	20/23	18/23	4/23	12/23

Ljen was detected in extraction or PCR negative controls by two laboratories.

Only two laboratories in exercise 4 and four laboratories in exercise 5 quantified the RNA: One used the Quant-iTTM RNA assay kit as well as the Bioanalyzer; one used the Quant-iTTM Ribo-Green[®] RNA Assay kit and two used a Nanodrop microvolume spectrophotometer (Table S5). Despite the small stain sizes and the different quantification methods, the RNA quantification results correlated quite well, in that the body fluid dilution series showed decreasing RNA amounts and high/low level stains could be differentiated (Table S5). Since only a small number of laboratories performed RNA quantitation, there was insufficient data in order to determine the correlation (if any) between the quantification result and RNA profiling success.

Post-PCR purification resulted in increased peak heights and reduced baseline signal noise but did not allow for detection of additional markers that were not observed prior to post-PCR purification. Eight laboratories in exercise 4 and 7 laboratories in exercise 5 performed post-PCR purification.

3.1.1. Comparison to other multiplexes

Two laboratories additionally tested the extracted RNA/cDNA with their own body fluid specific multiplexes, a 19plex from NFI [32] in EDNAP RNA exercise 4 and the CellTyper 11plex from ESR [23,24] in EDNAP RNA exercise 5. The 19plex includes the 2 menstrual secretion markers MMP7 and MMP11 with different primer sets than used in the MMP 3plex. The 19plex and the MMP 3plex results were difficult to compare, because different cDNA dilutions, PCR cycle numbers and CE injection voltage/time were applied (19plex data not shown). However, the menstrual blood stains and menstrual blood dilution series were identified correctly with both multiplexes. Additionally, the 19plex provided indications for stains 2 and 5 to be blood and vaginal secretion and also identified the body fluid dilution series as blood, saliva and semen. The CellTyper 11plex includes the 2 vaginal markers Lcris and Lgas with different primer sets than used in the Lacto 3plex. The results of the CellTyper 11plex (data not shown) and Lacto 3plex were in good agreement for stains 9–16 and the vaginal secretion dilution series. Additionally, the CellTyper 11plex identified stain 15 successfully as saliva.

3.2. DNA results

A total of 19 and 22 laboratories (exercise 4 and 5, respectively) performed RNA/DNA co-extraction. No specification for DNA analysis was provided. Various STR typing kits and

cycle numbers were utilized by the participating laboratories (Table S2). Full STR profiles were obtained from all stains and the menstrual blood/vaginal secretion dilution series (down to the 1/64 swab) by most laboratories (data not shown). The obtained genotypes were confirmed by comparison with reference profiles. A total of 15 and 19 laboratories (exercise 4 and 5, respectively) quantified the DNA using different quantification systems. Due to the small stain sizes and the different extraction and quantification methods used, the DNA quantification results were difficult to compare. However, the results were in the same order of magnitude for each particular stain (Table S6). Overall, the results demonstrated that DNA of sufficient quantity and quality for STR analysis could be simultaneously extracted with RNA from small amounts of dried stains.

3.2.1. RNA and DNA results of the blood/semen/saliva dilution series

Dilution series of blood (1–0.001 µl), semen (5–0.04 µl) and saliva (25–0.2 µl) were tested with the HKG triplex to test these housekeeping genes for their suitability as positive control for the presence of biological material and for successful RNA extraction/reverse transcription. These body fluid dilution series were prepared in such a way that loss of the housekeeping gene markers by dilution could be observed. Since saliva and semen are known to exhibit reduced housekeeping gene activity, higher input amounts of saliva and semen were used. Most laboratories performed RNA/DNA co-extraction on these samples and could provide RNA and DNA results. Some of these housekeeping genes could be detected down to 0.01 µl blood, 0.2 µl semen and 1 µl saliva (Table S7). Based on these dilution series, B2M and UBC seemed to be more sensitive than UCE for all tested body fluids. STR profiles could be recovered down to about the same amounts, except for semen where partial profiles down to 0.04 µl were reported (Table S7). On the other hand good RNA results (B2M and UBC) could be obtained from 0.01 µl blood, whereas poor STR results were obtained from this same amount of blood.

3.2.2. RNA and DNA results of the optional stain samples

The laboratories were invited to analyze additional samples including *bona fide* and mock casework samples, non-menstrual blood, non-vaginal secretion and non-human samples. Forty-nine menstrual blood samples, 28 non-menstrual blood samples (blood, saliva, semen, urine, vaginal secretion, skin) and 5 non-human samples (cat, dog, chimpanzee) were tested by 12 laboratories with the menstrual blood multiplexes. Fifty-five vaginal samples, 12 menstrual blood, 52 non-vaginal samples (blood, saliva, semen, skin, anal) and 2 non-human samples (cat, dog) were tested by 16

laboratories with the vaginal and housekeeping gene markers. At least one of the specific markers was detectable in almost all menstrual blood and vaginal secretion stains (Tables S7 and S8), including various stain sizes, storage periods and storage conditions. Vaginal markers were also detected in menstrual blood stains as is expected since menstrual blood is a complex mixture of menstrual blood, peripheral blood and vaginal secretions. For the menstrual blood stains, only one sample showed no RNA result at all (body fluid specific markers and housekeeping genes), perhaps due to limited cell amount, but a DNA profile could be generated. For two other menstrual blood stains housekeeping genes were detected, but no body fluid specific marker. This may be attributed to the presence of a limited amount of actual menstrual blood in the sample but sufficient biological material (e.g. vaginal secretion) present to permit detection of the housekeeping genes. Two vaginal secretion stains were negative for vaginal specific markers: One endovaginal swab from a 13 year old girl and another specified as underwear from a man in a rape case. In the latter case the presence of the stain being from this source might be uncertain or not enough female cell material was present.

Sporadic cross reactive peaks (most of them <1000 RFUs) were observed with the menstrual blood markers in blood, saliva, semen and more prominently in a skin sample (described as 'human skin wound') and several vaginal swabs (not specified at which day of cycle) (Table S8). The 3 housekeeping genes performed well in all human stains tested, only UCE appeared to perform less well in saliva samples (Table S8). Only sporadic peaks were observed with the vaginal markers in blood and skin, but several cross reactive peaks in saliva and semen were observed (mainly from Ljen) (Table S9). Full autosomal STR profiles were obtained for most samples where an RNA/DNA co-extraction was performed (data not shown).

4. Discussion

The purpose of this exercise was to evaluate an RNA/DNA co-extraction strategy with challenging menstrual blood and vaginal secretion samples and novel menstrual blood- and vaginal secretion-specific mRNA multiplexes (2 menstrual blood triplexes including the markers MMP7, MMP10, MMP11, MSX1, LEFTY2, SFRP4 as well as 2 vaginal secretion triplexes including the markers MYOZ1, CYP2B7P1, MUC4, Ljen, Lcris, Lgas) and a HBD1 singleplex. In addition, a housekeeping gene triplex (B2M, UBC, UCE) was tested to confirm the presence of RNA of sufficient quantity and quality for analysis. Co-extracted DNA was analyzed with various commercial STR typing kits.

All participating laboratories successfully applied the method using their own laboratory equipment and different kits and chemicals for RNA only or RNA/DNA co-extraction, reverse transcription and PCR. The samples to be analyzed in this exercise were challenging because the laboratories did not know the tissue origin and some stains were low template samples. The differing sensitivities between the laboratories can be explained by the different extraction kits and reverse transcription chemistries used. In addition, it is not possible to prepare identical stains for so many laboratories, different swabs and even different parts from the same swab can contain different cell amounts. And some packages experienced longer transit times at outside/room temperature, which could have impaired the primer mixes.

The evaluated menstrual blood markers proved to be specific, apart from single positive reactions with non-menstrual blood stains (mainly vaginal secretion). Towards the end of menstruation, as menstrual blood flow can be significantly lighter, some of the menstrual blood specific markers (MSX1, LEFTY2, SFRP4) tended to disappear, whereas the metalloproteinase markers were detected for most days of menstruation. We could not confirm the

reported reduced expression of MMP11 on day 1 [5], which might be due to the time point of sample taking and inter-individual differences. Amongst the menstrual blood markers, the metalloproteinase markers exhibited the highest sensitivity with more than half of the laboratories detecting all three markers with as little as 1/16 of a swab and MMP10 detected by more than half of the laboratories with as little as 1/64 swab. The sensitivity for MSX1, LEFTY2 and SFRP4, was considerably lower than those observed for the metalloproteinase markers. MSX1 was detected by more than half of the laboratories down to 1/16 of a swab, while LEFTY2 and SFRP4 were not detected by more than half of the laboratories for any of the sample sizes.

All of the evaluated vaginal secretion markers were able to detect vaginal secretion. However, MUC4 and HBD1 cross-reacted with other body fluids (mainly saliva) as has been reported previously [12,21,37,43]. It was reported that predominance of one *Lactobacillus* species is common in vaginal microbial populations [44,45]. In particular, 4 species of this genus (*L. iners*, *L. crispatus*, *L. gasseri*, *L. jensenii*) are mainly represented in women from different regions and ethnic groups with very limited evidence of their co-presence in the same host [44,45]. The results from this study appear to support that finding in that most vaginal samples exhibited only one of the 2 species Lcris or Lgas. Ljen was detected in the majority of the samples, but also in some negative controls, suggesting the possibility of general Ljen contamination in the environment as a possible reason for these results. A potential drawback of vaginal specific *Lactobacillus* markers is, that they are also detected on body sites close to (female groin) or in contact with (penis) the vagina, albeit this was tested only in small sample sizes [46]. The sensitivity was good for all evaluated vaginal secretion markers. MUC4, HBD1 and Ljen results should be interpreted carefully because of their cross-reactive potential. After the commencement of this study, a new vaginal marker (Mesothelin, MSLN) was described [47]. However, a large sample to sample variation and some cross reactivity with saliva was evident. This marker should be subjected to a validation exercise similar to the one described here to determine whether it offers superior sensitivity and specificity in comparison to any of the tested markers. In summary, most of the evaluated menstrual blood and vaginal secretion markers proved to be specific and sensitive and therefore suitable for forensic stains, even for aged and low template samples.

Expression analysis of housekeeping genes was evaluated as a positive control for the presence of biological material and a successful extraction and reverse transcription of RNA of sufficient quality and quantity for analysis. The sensitivity of housekeeping gene detection was in the same range as the body-fluid-specific markers [26,39,40] and STR analysis (Table S7), though slightly reduced in saliva and semen compared to the saliva and semen specific markers [40]. Two of the housekeeping genes tested in this study, B2M and UBC, proved to be quite sensitive and were detected in most body fluid stains (Table S8). Moreno et al. tested the expression levels of a selection of housekeeping genes and showed B2M to be expressed most consistent among body fluids [48]. Based on our data, these housekeeping genes do not appear to be particularly good positive controls in the sense that they should reliably show positive results when biological material is present. At least for saliva and semen, they are less sensitive than the body fluid specific genes. Anyhow, they can be useful for the interpretation of results together with RNA and DNA data. If for example, housekeeping gene expression analysis and STR profiling is successful but no body fluid specific genes are detected, the biological source material might be another, not tested body fluid or tissue. If on the other hand STRs can be detected, but no RNA profiling of either housekeeping or body fluid specific genes were obtained, sensitivity might be a problem. There is no consensus yet

as to whether housekeeping gene detection should be a requirement for RNA based body fluid identification assays. More studies are warranted in this area.

Another issue is the development of suitable interpretation guidelines, such as those suggested by Lindenberg et al. [32] or Roeder et al. [34]. What criteria should one use when considering the likely presence or absence of a particular fluid, given the absence/presence and/or relative peak heights of specific markers? How many and which markers should at least be present to identify a body fluid? In general, any false positive reduces the reliability of a test, therefore further investigations on specificity would be useful.

RNA and DNA quantification results were reasonably consistent among laboratories in that high/low input samples could be differentiated and the dilution series showed a monotonic gradual decrease in analytical signal. When comparing absolute numbers, the variation was of course quite high and standard deviations were in the same range as the mean values. This can easily be explained by the different extraction and quantification methods used by the laboratories. None of the suggested RNA quantification systems are ideal as they are not human-specific.

Two laboratories additionally used their own multiplexes (a 19plex from NFI and CellTyper 11plex from ESR) to identify the stains and dilution series. The results regarding menstrual blood and vaginal secretion were concordant. No detailed comparisons with multiplexes were performed since that was not the primary object of this study.

The following negative controls were included in RNA analysis: (1) RT minus (no reverse transcriptase added) to identify possible contaminating DNA (frequently a larger size than the expected RNA product) or the presence of pseudogenes (same size as RNA product) and (2) amplification blank. In exercise 4, only sporadic peaks appeared in the RT minus controls and are regarded as outliers. In exercise 5, eight laboratories found 1, 2 or 3 Lacto peaks in RT minus controls. This was most probably due to genomic DNA contamination since 4 of these laboratories did not use the recommended TURBO DNA-free DNase treatment. In addition, 2 laboratories found Ljen peaks in extraction or PCR negative controls, which could be due to crosstalk/carryover [49] or a general Ljen environmental contamination in the laboratory or contamination of the primer mix.

The possibility of co-extracting RNA and DNA from the same stain sample is an important operational consideration since the amount of sample is often limited in forensic casework. From almost all stains, good quality DNA profiles and the positive identification of menstrual blood and vaginal secretion could be achieved, even for aged stains and those exposed to un-controlled humidity.

In summary, the results of this study support an RNA/DNA co-extraction strategy allowing for positive identification of the tissue/fluid source of origin by mRNA profiling as well as a simultaneous identification of the body fluid donor by STR profiling. Most of the evaluated menstrual blood and vaginal secretion markers proved to be robust, reproducible and sensitive. Care should be taken with the vaginal markers MUC4, HBD1 and Ljen because of the potential for cross reactivity with other body fluids as well as presence in negative controls in the case of Ljen. Co-extracted DNA from the same stain provided good-quality STR profiles. Together with the precedent RNA exercises [38–40], a set of reliable RNA markers is now available for the identification of the most common forensic body fluids, namely blood, saliva, semen, menstrual blood and vaginal secretion. A subsequent EDNAP mRNA exercise will include an evaluation of mRNA markers for the identification of skin. Another task for future exercises would be the above mentioned development of suitable interpretation guidelines.

Acknowledgements

Special thanks for support and technical assistance to A.M. Bento (Forensic Genetic and Biology Service, National Institute of Legal Medicine and Forensic Sciences, Coimbra, Portugal), A. Carracedo (Institute of Legal Medicine, University of Santiago de Compostela, Spain), D. Niederwieser (Institute of Legal Medicine, Innsbruck Medical University, Austria), T. Sijen (Netherlands Forensic Institute, The Hague, The Netherlands), J. Stacey (Institute of Environmental Science and Research Ltd., Auckland, New Zealand) and M. Walt (Institute of Legal Medicine, University of Zurich, Switzerland).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.09.009.

References

- [1] K. Virkler, I.K. Lednev, Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene, *Forensic Sci. Int.* 188 (2009) 1–17.
- [2] T. Akutsu, K. Watanabe, H. Motani, H. Iwase, K. Sakurada, Evaluation of latex agglutination tests for fibrin–fibrinogen degradation products in the forensic identification of menstrual blood, *Leg. Med. (Tokyo)* 14 (2012) 51–54.
- [3] D.J. Baker, E.A. Grimes, A.J. Hopwood, D-dimer assays for the identification of menstrual blood, *Forensic Sci. Int.* 212 (2011) 210–214.
- [4] M. Bauer, A. Kraus, D. Patzelt, Detection of epithelial cells in dried blood stains by reverse transcriptase–polymerase chain reaction, *J. Forensic Sci.* 44 (1999) 1232–1236.
- [5] M. Bauer, D. Patzelt, Evaluation of mRNA markers for the identification of menstrual blood, *J. Forensic Sci.* 47 (2002) 1278–1282.
- [6] M. Bauer, D. Patzelt, A method for simultaneous RNA and DNA isolation from dried blood and semen stains, *Forensic Sci. Int.* 136 (2003) 76–78.
- [7] M. Bauer, D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains, *Int. J. Leg. Med.* 117 (2003) 175–179.
- [8] M. Bauer, S. Polzin, D. Patzelt, Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Sci. Int.* 138 (2003) 94–103.
- [9] J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, *Forensic Sci. Int.* 135 (2003) 85–96.
- [10] M. Alvarez, J. Juusola, J. Ballantyne, An mRNA and DNA co-isolation method for forensic casework samples, *Anal. Biochem.* 335 (2004) 289–298.
- [11] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, *Forensic Sci. Int.* 152 (2005) 1–12.
- [12] C. Nussbaumer, E. Gharehbaghi-Schnell, I. Korschneck, Messenger RNA profiling: a novel method for body fluid identification by real-time PCR, *Forensic Sci. Int.* 157 (2006) 181–186.
- [13] J. Ballantyne, Validity of messenger RNA expression analyses of human saliva, *Clin. Cancer Res.* 13 (2007) 1350, author reply 1351.
- [14] M. Bauer, RNA in forensic science, *Forensic Sci. Int. Genet.* 1 (2007) 69–74.
- [15] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, *J. Forensic Sci.* 52 (2007) 1252–1262.
- [16] M. Bauer, D. Patzelt, Identification of menstrual blood by real time RT-PCR: technical improvements and the practical value of negative test results, *Forensic Sci. Int.* 174 (2008) 55–59.
- [17] M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains, *J. Forensic Sci.* 53 (2008) 296–305.
- [18] D. Zubakov, E. Hanekamp, M. Kokshorn, W. van Ijcken, M. Kayser, Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples, *Int. J. Leg. Med.* 122 (2008) 135–142.
- [19] C. Haas, B. Klessner, C. Maake, W. Bar, A. Kratzer, mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, *Forensic Sci. Int. Genet.* 3 (2009) 80–88.
- [20] C. Haas, C. Muheim, A. Kratzer, W. Bar, C. Maake, mRNA profiling for the identification of sperm and seminal plasma, *Forensic Sci. Int. Genet. Suppl. Series 2* (2009) 534–535.
- [21] C. Cossu, U. Germann, A. Kratzer, W. Bar, C. Haas, How specific are the vaginal secretion markers HBD1 and MUC4? *Forensic Sci. Int. Genet. Suppl. Series 2* (2009) 536–537.
- [22] K. Sakurada, H. Ikegaya, H. Fukushima, T. Akutsu, K. Watanabe, M. Yoshino, Evaluation of mRNA-based approach for identification of saliva and semen, *Leg. Med. (Tokyo)* 11 (2009) 125–128.
- [23] R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, *Forensic Sci. Int. Genet.* 4 (2010) 311–315.

- [24] R.I. Fleming, S. Harbison, The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids, *Forensic Sci. Int. Genet.* 4 (2010) 244–256.
- [25] K. Sakurada, T. Akutsu, H. Fukushima, K. Watanabe, M. Yoshino, Detection of dermcidin for sweat identification by real-time RT-PCR and ELISA, *Forensic Sci. Int.* 194 (2010) 80–84.
- [26] C. Haas, E. Hanson, A. Kratzer, W. Bar, J. Ballantyne, Selection of highly specific and sensitive mRNA biomarkers for the identification of blood, *Forensic Sci. Int. Genet.* 5 (2011) 449–458.
- [27] K. Sakurada, T. Akutsu, K. Watanabe, Y. Fujinami, M. Yoshino, Expression of statherin mRNA and protein in nasal and vaginal secretions, *Leg. Med. (Tokyo)* 13 (2011) 309–313.
- [28] T. Takasaka, K. Sakurada, T. Akutsu, K. Nishigaki, H. Ikegaya, Trials of the detection of semen and vaginal fluid RNA using the genome profiling method, *Leg. Med. (Tokyo)* 13 (2011) 265–267.
- [29] T. Akutsu, H. Motani, K. Watanabe, H. Iwase, K. Sakurada, Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid, *Leg. Med. (Tokyo)* 14 (2012) 160–162.
- [30] J.H. An, K.J. Shin, W.I. Yang, H.Y. Lee, Body fluid identification in forensics, *BMB Rep.* 45 (2012) 545–553.
- [31] C. Haas, E. Hanson, J. Ballantyne, Capillary electrophoresis of a multiplex reverse transcription-polymerase chain reaction to target messenger RNA markers for body fluid identification, *Methods Mol. Biol.* 830 (2012) 169–183.
- [32] A. Lindenbergh, M. de Pagter, G. Ramdayal, M. Visser, D. Zubakov, M. Kayser, T. Sijen, A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces, *Forensic Sci. Int. Genet.* 6 (2012) 565–577.
- [33] M.L. Richard, K.A. Harper, R.L. Craig, A.J. Onorato, J.M. Robertson, J. Donfack, Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis, *Forensic Sci. Int. Genet.* 6 (2012) 452–460.
- [34] A.D. Roeder, C. Haas, mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification, *Int. J. Leg. Med.* 127 (2012) 707–721.
- [35] F. Kohlmeier, P.M. Schneider, Successful mRNA profiling of 23 years old blood stains, *Forensic Sci. Int. Genet.* 6 (2012) 274–276.
- [36] E.K. Hanson, J. Ballantyne, Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, *Sci. Justice* 53 (2013) 14–22.
- [37] J. Jakubowska, A. Maciejewska, R. Pawlowski, K.P. Bielawski, mRNA profiling for vaginal fluid and menstrual blood identification, *Forensic Sci. Int. Genet.* 7 (2013) 272–278.
- [38] C. Haas, E. Hanson, W. Bar, R. Banemann, A.M. Bento, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, A. Choma, M. Dotsch, M. Duriancikova, P. Hoff-Olsen, C. Hohoff, P. Johansen, P.A. Lindenbergh, B. Loddenkotter, B. Ludes, O. Maronas, N. Morling, H. Niederstatter, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Svieszena, L. Zatkalikova, J. Ballantyne, mRNA profiling for the identification of blood – results of a collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* 5 (2011) 21–26.
- [39] C. Haas, E. Hanson, M.J. Anjos, W. Bar, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, V. Castella, A. Choma, G. De Cock, M. Dotsch, P. Hoff-Olsen, P. Johansen, F. Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maronas, D. Moore, M.L. Morerod, N. Morling, H. Niederstatter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Svieszena, M. Turanska, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from blood stains – results of a second collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* 6 (2012) 70–80.
- [40] C. Haas, E. Hanson, M.J. Anjos, R. Banemann, A. Berti, E. Borges, A. Carracedo, M. Carvalho, C. Courts, G. De Cock, M. Dotsch, S. Flynn, I. Gomes, C. Hollard, B. Hjort, P. Hoff-Olsen, K. Hribikova, A. Lindenbergh, B. Ludes, O. Maronas, N. McCallum, D. Moore, N. Morling, H. Niederstatter, F. Noel, W. Parson, C. Popielarz, C. Rapone, A.D. Roeder, Y. Ruiz, E. Sauer, P.M. Schneider, T. Sijen, D.S. Court, B. Svieszena, M. Turanska, A. Vidaki, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from human saliva and semen stains – results of a third collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* 7 (2013) 230–239.
- [41] P.J. Smith, J. Ballantyne, Simplified low-copy-number DNA analysis by post-PCR purification, *J. Forensic Sci.* 52 (2007) 820–829.
- [42] I.S. Fraser, G. McCarron, R. Markham, T. Resta, Blood and total fluid content of menstrual discharge, *Obstetr. Gynecol.* 65 (1985) 194–198.
- [43] G. Hadzic, A. Lukan, K. Drobnic, Practical value of the marker MUC4 for identification of vaginal secretion in penile swabs, *FSI Genet. Suppl. Series 3* (2011) e222–e223.
- [44] S. Giampaoli, A. Berti, F. Valeriani, G. Gianfranceschi, A. Piccolella, L. Buggiotti, C. Rapone, A. Valentini, L. Ripani, V. Romano Spica, Molecular identification of vaginal fluid by microbial signature, *Forensic Sci. Int. Genet.* 6 (2012) 559–564.
- [45] J. Ravel, P. Gajer, Z. Abdo, G.M. Schneider, S.S. Koenig, S.L. McCulle, S. Karlebach, R. Gorle, J. Russell, C.O. Tacket, R.M. Brotman, C.C. Davis, K. Ault, L. Peralta, L.J. Forney, Vaginal microbiome of reproductive-age women, *Proc. Natl. Acad. Sci. U.S.A.* 108 (Suppl 1) (2011) 4680–4687.
- [46] C.C.G. Benschop, F.C.A. Quak, M.E. Boon, T. Sijen, I. Kuiper, Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? *Int. J. Legal Med.* 126 (2012) 303–310.
- [47] S.M. Park, S.Y. Park, J.H. Kim, T.W. Kang, J.L. Park, K.M. Woo, J.S. Kim, H.C. Lee, S.Y. Kim, S.H. Lee, Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification, *Forensic Sci. Int. Genet.* 7 (2013) 143–150.
- [48] L.I. Moreno, C.M. Tate, E.L. Knott, J.E. McDaniel, S.S. Rogers, B.W. Koons, M.F. Kavlick, R.L. Craig, J.M. Robertson, Determination of an effective housekeeping gene for the quantification of mRNA for forensic applications, *J. Forensic Sci.* 57 (2012) 1051–1058.
- [49] Maximizing the Performance of Capillary Electrophoresis Systems, *Applied Biosystems, Forensic News*, July (2010).

Clémence Hollard

Peuplement du sud de la Sibérie et de l'Altai à l'âge du Bronze : apport de la paléogénétique

Les steppes eurasiennes connectent l'Europe, l'Asie Centrale, la Chine, l'Asie du Sud et le Moyen-Orient. Au cours du temps, de nombreuses populations, issues d'Europe ou d'Asie, s'y sont succédées. L'accès direct au génome des populations anciennes semble être le moyen le plus fiable d'étudier la diversité génétique présente à un moment donné en s'affranchissant des événements postérieurs qui ont pu considérablement la modifier. Ce travail s'est plus particulièrement intéressé à la dynamique de peuplement du sud de la Sibérie et de l'Altai à l'âge du Bronze (III^{ème} - I^{er} millénaire avant J.-C), période pendant laquelle les steppes eurasiennes auraient connues de nombreux mouvements d'Ouest en Est. Les analyses moléculaires ont porté sur 69 spécimens anciens. La stratégie adoptée a consisté, en plus de l'étude de l'ADN mitochondrial, à étudier les lignées paternelles de ces individus, ainsi que des marqueurs autosomaux, informatifs de leur origine biogéographique (AIM) et des caractères pigmentaires tels que la couleur des yeux et des cheveux. L'ensemble des données moléculaires obtenues nous a permis de mettre en évidence une évolution du pool génique au sein de la région étudiée. Tout d'abord, il semble qu'en plus de partager des caractères culturels, les porteurs de la culture d'Afanasievo bien que mis au jour dans différentes régions (début de l'âge du Bronze) avaient des caractères génétiques communs laissant penser qu'ils proviennent d'un même groupe ancestral. Ces individus semblent génétiquement proches des populations européennes actuelles. Cela semble donc en accord avec l'hypothèse proposée par de nombreux archéo- et anthropologues, consistant à présenter la culture d'Afanasievo comme le résultat d'une diffusion vers l'Est des porteurs de cultures de l'ouest des steppes. L'analyse du pool génique des populations ayant suivies a révélé un apport de lignées est-eurasiennes, notamment au niveau des lignées paternelles. Des populations asiatiques autochtones se sont donc probablement mélangées avec ces peuples venus de l'Ouest. Enfin, durant le II^{ème} millénaire avant J.-C, le pool génique de ces populations semblent avoir été de nouveau modifié, avec l'apparition, sur une large zone s'étendant du sud de la Sibérie au bassin du Tarim en passant par l'Altai, de l'haplogroupe R1a1a. Cet haplogroupe pourrait ainsi être associé à l'expansion de la culture d'Andronovo et aux locuteurs de langues indo-iraniennes. Le faible effectif observé par groupe culturel ainsi que le processus de recrutement funéraire peut bien évidemment avoir biaisé une partie des analyses. Néanmoins, ces données moléculaires ont amené de nouveaux éléments pour la compréhension du peuplement de cette région qui apparaît comme un processus complexe, qu'il sera nécessaire d'approfondir lors de nouvelles études paléogénétiques voire paléogénomiques.

Settlement of South Siberia and Altai during the Bronze Age: contribution of paleogenetic

The Eurasian steppes connect Europe, Central Asia, China, South Asia and the Middle East. Over time, there was a succession of populations from both Europe and Asia in this area. Direct access to ancient human genomes seems to be the most reliable way to study the genetic diversity found in the past and escape from later events that could have modified this diversity. The present work has focused on the settlement of the South Siberia and the Altai mountains during the Bronze Age (III-I millennium BC), period during which the Eurasian Steppes knew eastward population movements. The molecular analyses were performed on 69 ancient samples. The used strategy consisted in the study of maternal and paternal lineages and autosomal markers informative of biogeographical ancestry (AIM) and physical appearance. Taken together, these results show an evolution of the genetic pool in this area during the Bronze Age. First, in addition to shared cultural features, the bearers of Afanasievo culture (early Bronze Age) from different regions shared common genetic features. These ancient specimens seemed close to current European populations in accordance to archaeo- and anthropological hypothesis which present Afanasievo culture as an eastward movement of population from western steppes. Then, the analysis of the later cultures revealed an input of east Eurasian lineages, particularly in paternal lineages. Native east Eurasian populations would have mix with these western populations. Finally, during the II millennium BC, this genetic pool seems to have evolved again, especially with the emergence of R1a1a haplogroup on a wide area from South Siberia through Altai to the Tarim Basin. This haplogroup can be associated to the Andronovo diffusion and the speaker of Indo-Iranian language. The low effective observed in each cultural group and the funeral recruitment could of course have induced a bias in some analyses. Nevertheless, these molecular data gave new elements to understand the settlement of this region which seems to be a complex process that it will be necessary to deepen with new paleogenetic even paleogenomic studies.