



Thèse présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur
Strasbourg I

Discipline :
Sciences du vivant - Aspects moléculaires et cellulaires de la biologie

par Elizabet Petkovski

*Polymorphismes ponctuels de séquence
et identification génétique
Etude par spectrométrie de masse MALDI-TOF*

Soutenue publiquement le 10 février 2006

Membres du jury

Directeur de Thèse :

M. Bertrand Ludes, Professeur, Université Louis Pasteur, E.A. 3428

Rapporteur Interne :

M. Jean-Marc Jeltsch, Professeur, Université Louis Pasteur, UMR 7175

Rapporteurs Externes :

M. Peter M. Schneider, Professeur, Institut de Médecine Légale, Université de Cologne

M. Daniel Rougé, Professeur, Institut de Médecine Légale, Université Paul Sabatier

Examineurs :

M. Eric Crubézy, Professeur, Université Paul Sabatier, UMR 8555, CNRS

M. Alain Van Dorsselaer, HDR, Université Louis Pasteur, UMR 7512

Notre vie est toujours emportée par le temps, qui ne cesse de nous échapper.

Jacques-Bénigne Bossuet

1627 - 1704

A la mémoire de mon père

Il a été et restera un modèle pour moi. Il m'a tout appris, tout donné. Mes actions, mes pensées, mes gestes sont "imprégnés" de lui.

A ma mère

Son soutien, sa patience, sa confiance, sa générosité et sa tendresse immense sont le moteur de mon avancée.

A ma sœur

Pour son encouragement et son affection.

Ce travail n'aurait vu le jour sans la confiance et la patience de mon directeur de recherche Monsieur le Professeur Bertrand LUDES qui m'a accueillie au sein de son équipe pendant ces années de thèse. Je tiens à lui exprimer toute ma reconnaissance d'avoir su m'encadrer, me conseiller et me soutenir tout en me laissant travailler très librement.

Mes plus sincères remerciements vont à Monsieur le Professeur Jean Marc JELTSCH qui m'a fait l'honneur d'accepter de juger mon travail et dont l'enseignement et la passion de ce dernier resteront des exemples pour moi.

Je remercie vivement Monsieur le Professeur Peter SCHNEIDER de l'honneur qu'il me fait en acceptant la charge de rapporteur de ma thèse. La grande estime et admiration que je lui porte m'ont naturellement conduit à solliciter son jugement.

Je remercie Messieurs les Professeurs Daniel ROUGÉ et Eric CRUBEZY qui ont eu la gentillesse d'accepter de juger mon travail.

Ce travail a bénéficié de la patience et des compétences de Monsieur le Docteur Alain VAN DORSSELAER et de son équipe, qui m'a accueillie de nombreuses fois (tous les vendredis de toutes les semaines de tous les mois ...). Sa présence dans ce jury, pour laquelle je le remercie, est un grand honneur.

Je remercie le Docteur Rémi HIENNE pour m'avoir accueillie au sein du laboratoire Codgene et pour m'avoir accordé sa confiance en m'offrant la possibilité de poursuivre ce travail.

Je tiens à exprimer ma reconnaissance au Docteur Christine KEYSER qui m'a encouragé et soutenue dans mes moments d'enthousiasme étincelant comme dans les moments de doute inconsolables.

Merci à ceux qui étaient mes collègues mais sont aujourd'hui des amis précieux. Ces personnes exceptionnelles se reconnaîtront car ils m'ont inlassablement encouragé, supporté et soutenu en participant aux moments joyeux et aux épreuves dures de ma vie.

Je dois ajouter que ce travail n'a pu être possible que grâce à la contribution de toutes les personnes travaillant à l'Institut de Médecine Légale et au laboratoire Codgène. Ils m'ont chacun à sa manière, à un moment ou à un autre, aidé, guidé, soutenu ou tout simplement ont pris de leur temps pour converser avec moi.

Merci à l'Association Nationale pour la Recherche Technique (ANRT) et au Ministère de l'Enseignement Supérieur et de la Recherche pour le soutien financier qui a rendu possible la réalisation de ce travail.

Je ne puis citer tous ceux qui de près ou de loin ont contribué à mon travail mais je ne les oublie pas. Je ne peux pas ignorer que la réalisation de cette thèse de doctorat n'aurait jamais été possible sans leur contribution. Merci à toutes et à tous!

Table des matières

Table des matières	1
Liste des figures	7
Liste des tableaux	9
Liste des abréviations	11
Introduction générale.....	12
I. Données de la littérature.....	14
I.1. Les empreintes génétiques	15
I.1.1. Polymorphisme de longueur des fragments de restriction	16
I.1.2. Polymorphisme de répétition	16
I.1.2.1. Minisatellites	16
I.1.2.1.1. Les sondes multiloculaires	17
I.1.2.1.2. Les sondes uniloculaires	18
I.1.2.1.3. Développement de la PCR.....	18
I.1.2.2. Microsatellites	19
I.1.3. Les marqueurs haplotypiques.....	20
I.1.3.1. Marqueurs du chromosome Y	21
I.1.3.2. L'ADN mitochondrial.....	22
I.1.4. Détermination du sexe.....	24
I.1.5. Domaines d'applications.....	24
I.1.5.1. Transplantations	25
I.1.5.2. Détermination du sexe.....	26
I.1.5.3. Identification d'un individu et des liens de parenté	26
I.1.5.4. L'ADN extrait de tissus anciens	27
I.2. Qualité et quantité des échantillons d'ADN	29
I.2.1. Dégradation de la molécule d'ADN.....	30
I.2.1.1. Ruptures de brin	30
I.2.1.2. Lésions hydrolytiques	31
I.2.1.3. Lésions oxydatives	31
I.2.2. Facteurs environnementaux.....	31
I.2.2.1. La quantité d'oxygène	32

I.2.2.2.	La température.....	32
I.2.2.3.	L'humidité	32
I.2.2.4.	Le pH.....	32
I.2.2.5.	Les UV	33
I.3.	Avantages et inconvénients des techniques actuelles	34
I.4.	Polymorphisme ponctuel de séquence	35
I.4.1.	Intérêt des marqueurs SNP autosomaux.....	37
I.4.2.	Technologies : Génotypage de SNP.....	38
I.4.2.1.	Technologies de discrimination allélique.....	38
I.4.2.1.1.	Hybridation	38
I.4.2.1.1.1.	Phares moléculaires	38
I.4.2.1.1.2.	Technologie 5'Nuclease	39
I.4.2.1.1.3.	Technologie Light Cycler	40
I.4.2.1.2.	Clivage invasif	41
I.4.2.1.3.	Ligature	42
I.4.2.1.4.	RFLP	43
I.4.2.1.5.	Extension d'amorce	44
I.4.2.1.5.1.	Extension de sonde allèle spécifique	44
I.4.2.1.5.2.	Extension allèle spécifique d'amorce	45
I.4.2.2.	Méthodes de détection des produits de la discrimination allélique.....	46
I.4.2.2.1.	Electrophorèse et détection directe de fluorescence	46
I.4.2.2.2.	Puces à ADN et détection directe de fluorescence	46
I.4.2.2.3.	Transfert résonant d'énergie de fluorescence	47
I.4.2.2.4.	Polarisation de fluorescence	48
I.4.2.2.5.	Luminescence	49
I.4.2.2.6.	Spectrométrie de masse.....	49
I.4.2.3.	Possibilités d'enchaînement des technologies	50
I.5.	Spectrométrie de masse MALDI-TOF.....	51
I.5.1.	Historique	51
I.5.2.	La source MALDI	53
I.5.2.1.	Ionisation.....	54
I.5.2.2.	Extraction retardée	55
I.5.2.3.	Matrices et modes de dépôt.....	56
I.5.2.3.1.	Les matrices	57

I.5.2.3.2.	Les modes de dépôt.....	58
I.5.3.	L'analyseur TOF.....	59
I.5.3.1.	Principe.....	59
I.5.3.2.	Quelques formules fondamentales	59
I.5.3.3.	Mode linéaire.....	60
I.5.3.4.	Mode réflectron	61
I.5.3.5.	Filtrage des ions de basse masse	61
I.5.4.	Avantages et inconvénients de la spectrometrie de masse MALDI-TOF	62
I.6.	Problématique.....	64
II.	Matériel et Méthodes.....	65
II.1.	Choix des SNP	66
II.1.1.	SNP autosomaux	66
II.1.1.1.	Critères de sélection.....	66
II.1.1.2.	Sources.....	66
II.1.1.3.	Aspects techniques	66
II.1.2.	SNP du chromosome X.....	68
II.1.3.	SNP du chromosome Y	68
II.2.	Echantillons.....	69
II.2.1.	Consentement éclairé	69
II.2.2.	Echantillons d'optimisation et validation technologique.....	70
II.2.2.1.	SNP autosomaux et du chromosome X	70
II.2.2.2.	SNP du chromosome Y	70
II.2.3.	Echantillons de population française.....	70
II.2.4.	Echantillons anciens : étude des SNP-Y	71
II.3.	De l'ADN dans l'échantillon à l'ADN dans le tube	71
II.3.1.	Extraction d'ADN.....	71
II.3.2.	Purification de l'ADN.....	71
II.3.2.1.	Précipitation à l'éthanol	71
II.3.2.2.	Purification à l'aide du kit Clean-Mix™ (Talent).....	72
II.3.3.	Quantifications de l'ADN par PCR en temps réel	72
II.3.4.	Concentration et purification par ultrafiltration	73
II.3.5.	Utilisation de la technologie FTA® (Whatman, Inc).....	73
II.4.	Amplification par PCR.....	74
II.4.1.	Dessin et construction d'amorces PCR.....	74

II.4.2.	Optimisation des réactions de PCR.....	75
II.4.2.1.	Plan d'expérience	76
II.4.2.2.	Les autres paramètres de réaction PCR	76
II.4.2.3.	Détection des amplicons	77
II.5.	Séquençage.....	77
II.6.	Extension d'amorce.....	78
II.6.1.	Dessin d'amorces PEX	78
II.6.2.	Optimisation PEX	78
II.6.2.1.	Plan d'expérience	79
II.6.2.2.	Les autres paramètres de réaction PEX	79
II.6.3.	Purification des produits PEX	80
II.7.	Les réactions simultanées.....	80
II.8.	Spectrométrie de masse MALDI-TOF.....	81
II.8.1.	Préparation de la matrice.....	81
II.8.2.	Préparation du dépôt.....	82
II.8.3.	Analyse par spectrométrie de masse MALDI-TOF	82
II.8.4.	Détermination de la masse	83
II.9.	Analyse des SNP du chromosome Y.....	83
II.10.	Analyse des SNP du chromosome X.....	84
II.11.	Analyses statistiques	85
II.11.1.	Paramètres spécifiques des loci.....	85
II.11.2.	Utilisation du logiciel Genepop.....	86
II.11.3.	Probabilités de paternité	87
III.	Résultats	88
III.1.	Les marqueurs autosomaux.....	89
III.1.1.	Sélection théorique	89
III.1.2.	Tri "expérimental"	89
III.1.2.1.	La PCR simple	89
III.1.2.2.	Le séquençage	90
III.1.2.3.	La PEX simple	91
III.1.2.4.	Analyse en spectrométrie de masse MALDI-TOF.....	91
III.1.2.4.1.	Préparation de la matrice et du dépôt.....	91
III.1.2.4.2.	Analyse des produits d'extension d'amorce.....	92
III.1.2.4.3.	Détermination de la masse	93

III.1.2.5.	Les réactions "multiplexes"	94
III.1.2.6.	Les hauteurs de pics	101
III.1.3.	Résultats de l'évaluation statistique	102
III.1.3.1.	Les marqueurs	102
III.1.3.2.	Les génotypes	103
III.1.3.3.	Evaluations statistiques	106
III.1.3.3.1.	Les marqueurs	106
III.1.3.3.1.1.	Fréquences et degré d'hétérozygotie	106
III.1.3.3.1.2.	Equilibre de Hardy-Weinberg	108
III.1.3.3.1.3.	Pouvoir discriminatoire	109
III.1.3.3.2.	L'ensemble des marqueurs	109
III.1.3.3.2.1.	Déséquilibre de liaison	110
III.1.3.3.2.2.	Qualité de l'ensemble des SNP	110
III.1.3.3.2.3.	Fréquence du profil	111
III.1.4.	Résultats sur les paternités	111
III.2.	Marqueur pour la détermination du sexe	111
III.3.	Les SNP du chromosome X	113
III.4.	Les SNP du chromosome Y	113
III.4.1.	La détection	113
III.4.2.	TAT, M242 et RPS4Y dans le contexte des échantillons	114
IV.	Discussion	115
IV.1.	Etude de séquences de plus en plus courtes	116
IV.1.1.	Utilisation des SNP	117
IV.1.1.1.	Caractéristiques intéressantes des SNP en général	117
IV.1.1.2.	Avantages des SNP par rapport aux STR	118
IV.1.1.3.	Inconvénients des SNP par rapport aux STR	118
IV.1.2.	Détection par spectrométrie de masse MALDI-TOF	119
IV.2.	La problématique au niveau de chaque étape technique	120
IV.2.1.	La PCR	120
IV.2.2.	PEX & MALDI-TOF	122
IV.3.	SNP : "to be or not to be"	126
IV.3.1.	Identification génétique	127
IV.3.1.1.	Evaluation des loci	127
IV.3.1.2.	Etude de mélanges	129

IV.3.2.	Etude des paternités.....	130
IV.3.3.	Détermination du sexe.....	132
IV.3.4.	Les SNP du chromosome X	133
IV.4.	SNP versus STR.....	134
IV.5.	Les SNP dans les laboratoires d'identification génétique	137
IV.6.	Etude des échantillons anciens.....	138
V.	Conclusions.....	141
VI.	Références bibliographiques	146
VII.	Annexes.....	182
	<i>Annexe 1 : Exemple de consentement éclairé.</i>	183
	<i>Annexe 2: Tous les SNP considérés.</i>	184
	<i>Annexe 3: Amorces pour la réaction de PCR.</i>	188
	<i>Annexe 4 : Amorces pour la réaction de PEX.</i>	192
VIII.	Publications.....	196
	SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology.....	197
	Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise.	205
	A set of 51 biallelic markers for human identification and filiation in French people.	218
	Validation of SNPs as markers for individual identification.	237
	SNPs: tools for individual identification.....	242
	Detection of genetic variation by MALDI-TOF mass spectrometry: rapid SNP genotyping using the GENOLINK system.....	246
	Does the Tat polymorphism originate in northern Mongolia?.....	250
	MALDI-TOF MS analysis of Y-SNPs in ancient samples.	254

Liste des figures

Figure 1. Polymorphismes ponctuels de séquence.	35
Figure 2. Principe de la discrimination allélique par hybridation allèle spécifique de phares moléculaires.	39
Figure 3. Principe de la discrimination allélique par hybridation allèle spécifique de sondes TaqMan.	39
Figure 4. Principe de la discrimination allélique par hybridation allèle spécifique et la technologie Light Cycler.	40
Figure 5. Principe de la discrimination allélique par clivage invasif.	41
Figure 6. Principe de la discrimination allélique par réaction de ligature.	42
Figure 7. Principe de la discrimination allélique par digestion spécifique de la séquence par une enzyme de restriction.	43
Figure 8. Principe de la discrimination allélique par extension de sonde allèle spécifique. ...	44
Figure 9. Principe de la discrimination allélique par extension allèle spécifique d'amorce sur une seule base, selon le SNP uniquement.	45
Figure 10. Principe de la discrimination allélique par extension allèle spécifique d'amorce sur plusieurs base, selon le SNP et la séquence en 5' de ce dernier.	45
Figure 11. Principe de la spectrométrie de masse MALDI-TOF.	49
Figure 12. Représentation schématique des différentes technologies de discrimination allélique et de détection de ses produits.	50
Figure 13. Source et tube de vol de l'appareil Ultraflex TOF-TOF (Bruker Dlatonics).	53
Figure 14. Schématisation du processus de désorption / ionisation MALDI.	54
Figure 15. Schéma du procédé de l'extraction retardée.	56
Figure 16. Structure des matrices les plus utilisées pour l'analyse d'acides nucléiques.	58

Figure 17. Présentation schématique des étapes de sélection et de validation des marqueurs SNP.	67
Figure 18. Arbre phylogéographique montrant les SNP du chromosome Y étudiés ainsi que les haplogroupes qu'ils définissent.	68
Figure 19. Vue en microscopie électronique d'une portion d'un dépôt de sang sur un support FTA® (Whatman, Inc).	74
Figure 20. Alignement de séquences du fragment contenant le marqueur rs1321158.	90
Figure 21. Aspect du dépôt de la matrice et de l'échantillon par la méthode de la goutte séchée après séchage à l'étuve et dans une cloche à vide.	91
Figure 22. Aspects de deux dépôts d'un même échantillon réalisés consécutivement dans les conditions identiques.	92
Figure 23. Exemples de spectres obtenus lors de l'analyse du polymorphisme rs820129.	93
Figure 25. Déviation standard des distributions de masses observées.	94
Figure 26. Exemples de spectres d'analyses simultanées de plusieurs SNP.	100
Figure 27. Spectres obtenus à partir de deux dépôts consécutifs d'un même échantillon	101
Figure 28. Représentation graphique de la distribution des fréquences alléliques observées sur l'échantillon de population étudié.	106
Figure 29. Alignement des séquences de fragments du gène de l'amélogénine sur les chromosomes X et Y contenant le marqueur AmelSD.	112
Figure 30. Exemples de spectres obtenus lors de l'analyse du marqueur AmelSD.	112
Figure 31. Carte de la nécropole de la vallée d'Egyin Gol.	139

Liste des tableaux

Tableau 1. Méthodes de discrimination allélique dont les produits peuvent être détecté par la spectrométrie de masse MALDI-TOF.	62
Tableau 2. Facteurs et leurs niveaux évalués dans le plan d'expérience pour l'optimisation des réactions PEX.	79
Tableau 3. Composition en nucléotidique des différents milieux PEX.	80
Tableau 4. Facteurs et leurs niveaux évalués dans le plan d'expérience pour l'optimisation de la cristallisation des dépôts.	81
Tableau 5. Valeurs de paramètres de la méthode d'analyse appliquées à la détection des oligonucléotides par l'instrument Ultraflex TM (Bruker Daltonics).	82
Tableau 6. Séquence des amorces utilisées pour l'amplification simultanée des fragments contenant les 3 SNP-Y sélectionnés.	83
Tableau 7. Séquence et masse des produits attendus lors de l'analyse des marqueurs TAT, M242 et RPS4Y.	84
Tableau 8. Les SNP du chromosome X étudiés.	85
Tableau 9. Rapports de vraisemblance dans les tests de paternité.	87
Tableau 10. Les groupes d'amplification des fragments contenant les SNP sélectionnés.	94
Tableaux 11a à 11g. Composition nucléotidique des réactions PEX pour l'étude des différents marqueurs.	95 - 99
Tableau 12. Numérotation des SNP selon leur position génomique.	102
Tableau 13. Représentation sous forme de code ternaire des génotypes obtenus.	104
Tableau 14. Fréquences alléliques et degré d'hétérozygotie observés des SNP.	107
Tableau 15. Résultats des tests exacts de divergence de l'équilibre de Hardy-Weinberg pour les marqueurs étudiés.	108

Tableau 16. Valeurs de pouvoir discriminatoire des marqueurs calculé à partir des fréquences des génotypes observés.	109
Tableau 17. Résultats des tests exacts de déséquilibre de liaison pour les marqueurs situés sur le même chromosome.	110
Tableau 18. Résultats des tests exacts de déséquilibre de liaison pour les marqueurs situés sur le chromosome.	113
Tableau 19. Profils obtenus sur les marqueurs M242, TAT et RPS4Y à partir de 3 spécimens anciens de la Nécropole de la vallée d'Egyin Gol et les haplogroupes auxquels ils appartiennent.	114

Liste des abréviations

ADNmt	Acide D ésoxyribo N ucléique Mi Tochondrial
ARMS	A mplification R efractory M utation S ystem
ASO	A llele S pecific O ligonucleotide
BET	B romure d' E Thidium
DAHC	D i A mmonium H ydrogène C itrate
DHPLC	D enaturing H igh P erformance L iquid C hromatography
ESI	E lectro S pray I onisation
FP	F luorescence P olarisation
FRET	F luorescence R esonance E nergy T ransfert
HPA	H ydroxy P icolinic A cid
HV	H yper V ariable (se réfère aux régions hypervariables de ADNmt)
LR	L ikelihood R atio
MALDI-TOF	M atrix A ssisted L aser D esorption/ I onisation T ime O f F light
MS	M ass S pectrometry
MWCO	M olecular W eight C ut O ff
NCBI	N ational C enter for B io T echnology I nformation
NR _Y	N on R ecombining region of the Y chromosome
OLA	O ligonucleotide L igation A ssay
PCR	P olymerase C hain R eaction
PEX	P rimer E Xtension
RFLP	R estriction F ragment L ength P olymorphism
SLP	S ingle L ocus P robe
SNP	S ingle N ucleotide P olymorphism
SR _Y	S ex D etermining R egion of the Y chromosome
SSO	S equence S pecific O ligonucleotide
STR	S hort T andem R epeat
THAP	T ri H ydroxy A ceto P henone
VNTR	V ariable N umber of T andem R epeat

Introduction générale

Les techniques de profils génétiques ou génotypage mettent en évidence les régions polymorphes du génome des individus. Elles permettent d'une part d'identifier les individus avec certitude et d'autre part d'établir les relations de parenté intrafamiliales. Les méthodes appliquées aujourd'hui en routine sont complémentaires. L'analyse de microsatellites (*Short Tandem Repeats*, STR) présents au niveau des autosomes, permet d'obtenir des résultats fiables, reproductibles, applicables à de nombreux types d'échantillons. Elle permet l'étude de mélanges d'ADN de plusieurs individus et possède un pouvoir de discrimination très élevé. Cependant, certains échantillons très dégradés ne peuvent pas être identifiés par le biais des microsatellites. En effet, étant des polymorphismes de longueur de séquence ils nécessitent l'analyse de fragments de taille relativement importante. Le séquençage de l'ADN mitochondrial, du fait de l'existence d'un très grand nombre de copies par cellule, peut être réalisé sur des échantillons très dégradés. Néanmoins, cette méthode est beaucoup moins discriminante que l'analyse des marqueurs autosomaux et permet d'identifier une lignée maternelle et non un individu. De façon à associer les caractéristiques complémentaires des méthodes d'analyse appliquées aujourd'hui en routine, il s'avère nécessaire de développer de nouvelles techniques d'analyse basées sur l'étude de marqueurs nucléaires de petite taille.

Le travail présenté dans ce mémoire a été réalisé dans le cadre d'une convention CIFRE entre le laboratoire d'identification génétique Codgene et l'Institut de Médecine Légale de Strasbourg. L'objectif de cette collaboration a été de sélectionner et de valider un ensemble de polymorphismes ponctuels de séquence (*Single Nucleotide Polymorphism*, SNP) pour l'identification génétique chez l'homme, et de développer leur analyse basée sur la PCR et applicable à une grande variété de types d'échantillons. Nous avons opté pour la détection par spectrométrie de masse MALDI-TOF (*Matrix Assisted Laser Desorption / Ionisation Time Of Flight*) car cette méthode très sensible et rapide permet l'analyse simultanée d'un grand nombre d'analytes. L'application de cette technologie aux domaines de l'identification génétique est le fruit d'une collaboration entre l'Institut de Médecine Légale et le Laboratoire de Spectrométrie de Masse Bioorganique (LSMBO, Strasbourg).

Ce mémoire débute par une synthèse des données de la littérature relatives à l'identification et la filiation génétiques permettant d'appréhender le thème de recherche. Après un énoncé des méthodes employées dans la réalisation du travail, l'ensemble des résultats est exposé de façon séquentielle par étape de validation des marqueurs. L'ensemble des observations est ensuite discuté au regard des connaissances actuelles afin de dégager les avantages et inconvénients de la méthode développée. Enfin, les conclusions mettront en évidence les perspectives de recherche.

I. Données de la littérature

I.1. Les empreintes génétiques

La génétique criminalistique a débuté il y a plus d'un siècle lorsque Karl Landsteiner a appliqué sa découverte des polymorphismes des groupes sanguins ABO chez l'homme à la résolution de crimes [Landsteiner et al., 1900]. Jusqu'aux années 1980, des méthodes sérologiques [Mourant et al., 1976] et d'électrophorèse de protéines [Harris et al., 1976] étaient utilisées pour accéder aux polymorphismes des groupes sanguins et des groupes sécréteurs. Le principal désavantage de ces marqueurs était leur dégradation rapide, ainsi que leur sensibilité aux enzymes et à la contamination bactérienne menant à des faux positifs. Une faible variabilité est détectée lors de l'analyse de huit systèmes de groupes sécréteurs à partir d'une tache de sang. Ces "profils d'expression" permettent d'obtenir une probabilité d'identité (*probability of match*, pM), c'est à dire une probabilité que deux personnes sans lien de parenté partagent la même combinaison, variant entre 10^{-3} et 10^{-2} . Cette probabilité augmente lorsque l'analyse porte sur différents tissus humains puisque les groupes sécréteurs ne permettent l'analyse que de cellules dans lesquelles ils sont exprimés. Cependant, ces techniques permettaient des exclusions rapides.

En 1985, l'analyse d'ADN est introduite en criminalistique sous la forme dite "d'empreintes génétiques" suite à la découverte de séquences répétitives hautement polymorphes [Jeffreys et al., 1985a].

Le génome humain, constitué de 6 milliards de paires de bases (pb) comporte environ 35000 gènes [International Human Genome Mapping Consortium, 2001 ; Venter et al., 2001]. Cependant, plus de 90% du génome est non-codant, c'est à dire ne comporte pas d'information en rapport direct avec la synthèse de protéines. Contrairement à l'ADN codant, la partie non-codante du génome n'est pas soumise à la forte pression de sélection ce qui permet aux mutations qui y surviennent d'être conservées et transmises à la descendance. Ces régions où l'on rencontre une forte variabilité sont très informatives pour la génétique criminalistique tout en étant phénotypiquement neutres. Plusieurs types de polymorphismes de l'ADN ont été utilisés comme marqueurs du génétique, chacun révélant un niveau différent de variabilité.

I.1.1. POLYMORPHISME DE LONGUEUR DES FRAGMENTS DE RESTRICTION

Le polymorphisme de longueur des fragments de restriction ou RFLP (*Restriction Fragment Length Polymorphism*) résulte de variations individuelles de la localisation de sites de restriction pour une enzyme donnée. Il peut être dû soit à une création ou une suppression d'un site de restriction en relation avec une mutation soit à une variation de distance entre deux sites suite à une insertion ou une délétion d'ADN.

Ces polymorphismes sont révélés par la méthode du Southern blot [Southern, 1975] après digestion enzymatique (généralement par *HinfI* en Europe et par *HaeIII* aux Etats Unis) de l'ADN extrait afin d'observer, grâce à leur reconnaissance par une sonde marquée, des fragments différant par leur longueur [Botstein et al., 1980 ; Wyman et White, 1980].

I.1.2. POLYMORPHISME DE REPETITION

Environ 30% de la partie non codante du génome humain est constitué de séquences répétées. Le polymorphisme de répétition est créé par des séquences courtes (dites séquences noyau ou motif de base), non codantes, répétées de manière juxtaposée. Ces séquences montrent une très grande variabilité dans le nombre de répétitions dites en tandem de la séquence noyau et sont transmises de façon stable selon les lois mendéliennes. Selon la structure et le nombre de répétitions du motif de base formant le polymorphisme, on distingue les minisatellites et les microsatellites.

I.1.2.1. Minisatellites

La découverte des minisatellites par Alec Jeffreys en 1985 [Jeffreys et al., 1985a] a révolutionné l'utilisation de l'ADN en matière d'identification et de filiation. Le motif de base de ces marqueurs, également appelés VNTR (*Variable Number of Tandem Repeats*) [Nakamura et al., 1987], compte entre 9 et 100 pb. Il est réitéré entre deux et quelques centaines de fois à chaque locus [Tautz, 1989 ; 1993] générant ainsi des fragments de 500 pb à 20 kpb en taille. Les minisatellites se trouvent plus fréquemment dans les régions subtélomériques des chromosomes [Royle et al., 1988 ; Amarger et al., 1998] et leur

variabilité semble être liée à des recombinaisons méiotiques ("*crossing-over*") inégales et des conversions génétiques [Jeffreys et al., 1994 ; Jeffreys et al., 1998].

I.1.2.1.1. Les sondes multiloculaires

Les VNTR étaient en premier lieu détectés par hybridation de sondes constituées de répétitions de la séquence noyau à l'ADN génomique transféré sur membrane après restriction enzymatique (technique Southern blot).

Les séquences noyau partagées par différents loci minisatellites permettaient la détection simultanée de nombreux minisatellites, produisant ainsi des combinaisons hypervariables de multiples bandes appelées empreintes ADN. Les deux sondes multiloculaires 33.6 et 33.15 ont été décrites en premier et sont les plus utilisées [Jeffreys et al., 1985b ; 1985c ; Gill et al., 1985]. Chacune de ces sondes permet de détecter typiquement 17 fragments par personne variant de 3,5 à plus de 20 kpb. Les deux sondes permettent la détection de différents fragments d'ADN se chevauchant sur 1% seulement [Jeffreys et al., 1986]. Par la suite, d'autres sondes multiloculaires ont été décrites en raison de la très grande variabilité de structure du motif de base. En effet, il a été montré que quasiment toute sonde à séquence répétée en tandem, permet, dans une certaine mesure, de détecter de multiples fragments variables d'ADN [Vergnaud, 1989]. Les sondes multiloculaires les plus efficaces comportent un fort pourcentage de guanine et sont similaires à la séquence noyau des sondes de Jeffreys. L'utilisation de deux sondes, complémentaires de deux types de séquences noyau, permet d'obtenir une empreinte ADN partagée uniquement par les jumeaux monozygotes [Jeffreys et al., 1991].

Les analyses par sondes multiloculaires nécessitent 500 ng d'ADN pour obtenir un profil génétique aisément interprétable [Ludes et Mangin, 1992]. Leur efficacité se trouve limitée lors de l'interprétation d'empreintes incomplètes résultant d'une dégradation partielle ou d'une faible quantité d'ADN et lors de comparaisons de différentes autoradiographies montrant de subtiles variations de stringence d'hybridation de la sonde. L'interprétation est impossible dans le cadre d'analyse d'échantillons contenant l'ADN de plus d'un contributeur. De plus, cette méthode est très sensible aux contaminations par des microorganismes puisque les motifs de base des sondes multiloculaires permettent de détecter des séquences de génomes bactériens, notamment celui de *Escherichia coli*.

Les empreintes génétiques fournissent un "phénotype de l'ADN" et non le génotype, ce qui rend les informations sur les loci polymorphiques et les allèles inaccessibles. Ainsi, dans le

cas de tests de paternité, la survenue de mutations dans les lignées germinales induit l'apparition de nouveaux fragments. Celles-ci ne pouvant pas être attribuées aux parents véritables, la comparaison des empreintes révèle une exclusion apparente.

I.1.2.1.2. Les sondes uniloculaires

Alors que l'utilisation de sondes multiloculaires a persisté quelques années dans les tests de paternité, l'analyse génétique en criminalistique a rapidement opté pour l'utilisation de minisatellites spécifiques, les sondes uniloculaires (*single locus probes*, SLP) [Giusti et al., 1986 ; Kanter et al., 1986]. Chaque SLP révèle un seul RFLP hautement polymorphe ce qui permet de déduire les informations génotypiques et simplifie grandement l'interprétation des empreintes obtenues. De multiples SLP ont été isolés et clonés [Nakamura et al., 1987 ; 1988 ; Wong et al., 1986 ; 1987].

Toutefois, le pouvoir d'incrimination pour chacune des sondes uniloculaires est plus faible que celui des sondes multiloculaires et il est nécessaire de réhybrider à plusieurs reprises les membranes de transfert. Les sondes permettant d'explorer des régions polymorphes recommandées pour l'identification des individus, ont été isolées et décrites en 1987 [Nakamura et al., 1987]. Le locus le plus variable et le plus informatif est MS1 dont les allèles varient en taille entre 1 et 23 kpb. Par hybridation Southern blot il est possible de différencier deux allèles chez plus de 99% des individus. Les répétitions en tandem du motif noyau de 9 pb engendrent 2400 allèles tous représentés dans la population humaine [Wong et al., 1987 ; Smith et al., 1990].

Le principal avantage de l'analyse SLP est l'énorme variabilité de certains minisatellites et la connaissance de leur taux de mutation. De même, l'analyse de mélanges d'ADN de plusieurs individus est possible. Cette méthode nécessite beaucoup de temps et d'ADN non dégradé (20 à 50 ng d'ADN génomique) [Ludes et Mangin, 1992].

I.1.2.1.3. Développement de la PCR

Les limites des technologies décrites ci-dessus sont essentiellement dues à la quantité minimale d'ADN "cible" nécessaire et l'absence de dégradation par des facteurs physiques et de contamination fongique ou bactérienne. Cependant, elles peuvent être contournées par l'amplification spécifique des régions informatives de l'ADN. La technique dite de *polymerase chain reaction* (PCR) décrite en 1986 [Mullis et al., 1986 ; Mullis et Faloona, 1987] autorise

l'analyse de quantités minimales d'ADN même dégradé. En effet, cette technique permet à chaque cycle thermique, comportant la dénaturation, l'hybridation et l'élongation, la synthèse spécifique *de novo* du fragment flanqué par les amorces, par leur extension simultanée sur les deux brins complémentaires *in vitro*. Ainsi, à partir d'une molécule d'ADN, en 20 à 25 cycles de PCR, sont théoriquement formées 2^{20} à 2^{25} copies, si l'efficacité de la réaction atteint 100%. La PCR est aujourd'hui à la base des profils génétiques.

Les premiers systèmes d'identification génétique basés sur la PCR étaient couplés aux sondes SSO (*Sequence specific oligonucleotide*) [Conner et al., 1983] et avaient pour cible un faible nombre de polymorphismes ponctuels dans le gène HLA DQ alpha [Saiki et al., 1986 ; Helmuth et al., 1990]. Ces systèmes étaient utiles lorsque la technologie SLP ne permettait pas de conclure mais étaient de pouvoir discriminant très faibles et difficilement interprétables dans les cas de mélanges. Ainsi pendant une période, les tests PCR et SLP étaient faits en parallèle et ce jusqu'au développement de l'amplification des minisatellites [Jeffreys et al., 1988 ; Boerwinkle et al., 1989]. C'est la découverte des microsatellites couplée au développement du séquençage qui a conduit aux puissants systèmes d'identification dont nous disposons aujourd'hui.

I.1.2.2. Microsatellites

Le motif de base des microsatellites ou STR, compte entre 2 et 6 pb et est réitéré entre deux et environ cinquante fois à chaque locus [Litt et Luty, 1989 ; Tautz, 1989] générant des allèles variant en taille entre 50 et 500 pb.

Contrairement aux minisatellites, les STR ont une distribution plutôt uniforme sur le génome avec la présence d'un locus toutes les 6 à 10 kpb [Beckman et Weber, 1992]. Leur variabilité semble principalement liée à des dérapages de réplication [Di Rienzo et al., 1994].

La structure des STR peut être simple ou complexe [Urquhart et al., 1994 ; Lazaruk et al., 2001]. Les STR complexes montrent une grande variabilité et sont de ce fait plus informatifs, alors que les STR simples permettent une standardisation facile et montrent un taux de mutation plus faible. Les taux de mutation des STR sont très variables [Ascioglu et al., 2004 ; Leopoldino et Pena, 2003] et dépendent principalement de la structure et la longueur du marqueur [Brinkmann et al., 1998].

Les microsatellites dont la séquence noyau est un dinucléotide, sont les plus fréquents dans le génome humain mais leur analyse basée sur la PCR est la moins adaptée à l'identification génétique. En effet, des dérapages, analogues à ceux survenant lors de la réplication *in vivo*,

surviennent lors de l'amplification *in vitro* de dinucléotides répétés en tandem provoquant l'apparition de sous-produits artéfactuels appelés *stutters* ou pics de "bégaiement" de la polymérase [Hauge et Litt, 1993 ; Litt et al., 1993]. Ces produits correspondent en taille à des allèles réduits le plus souvent d'une ou de deux répétitions par rapport à l'allèle réel. Les *stutters* apparaissent moins fréquemment lors de l'analyse de tétra- ou pentanucléotides, ce qui les rend plus appropriés aux applications médico-légales. La réduction de la proportion des *stutters* est cruciale pour l'analyse de mélanges d'ADN de plusieurs individus.

D'autres facteurs de validation entrent dans la sélection de STR pour l'identification génétique parmi lesquels l'absence d'autres produits artéfactuels, la robustesse de l'analyse du marqueur et en particulier la taille et la possibilité d'analyse simultanée avec d'autres marqueurs. En effet, ces derniers critères sont cruciaux lors de l'analyse d'ADN dégradé ou en très faible quantité.

Aujourd'hui, les microsatellites sont détectés par amplification PCR grâce à des amorces marquées par des fluorophores et analysés par électrophorèse capillaire automatique. Les kits commercialement disponibles permettent l'analyse entièrement automatisée et simultanée de 15 marqueurs ainsi que celle du gène de l'amélogénine afin de révéler le sexe de la personne à l'origine de l'échantillon [Akane et al., 1992 ; Sullivan et al., 1993]. Le pouvoir discriminant de ces profils génétiques est très fort avec la probabilité pM que deux personnes prises au hasard partagent les mêmes allèles inférieure à de 10^{-15} [Pascal, 1998].

I.1.3. LES MARQUEURS HAPLOTYPIQUES

Les marqueurs autosomaux subissent un brassage génétique à chaque génération puisqu'ils sont transmis de manière biparentale. Ainsi, la moitié de l'information génétique d'un individu lui vient de son père et l'autre moitié de sa mère. Les marqueurs uniparentaux, c'est à dire ceux situés sur le chromosome Y et sur l'ADN mitochondrial, sont transmis d'une génération à l'autre sans changement sauf dans le cas de mutations. Les marqueurs de l'ADN mitochondrial, transmis de mère à enfant, permettent de retracer les lignées maternelles [Oota et al., 1995] et ceux du chromosome Y, transmis de père en fils, les lignées paternelles [Schultes et al., 1999]. Cette caractéristique rend les marqueurs moins informatifs pour l'identification individuelle. L'information génétique de chaque marqueur uniparental est appelée haplotype au lieu de génotype puisqu'un seul allèle est détecté par individu.

I.1.3.1. Marqueurs du chromosome Y

Le premier microsatellite polymorphique du chromosome Y, aujourd'hui nommé DYS19, a été décrit en 1992 [Roewer et Epplen, 1992]. Depuis, des centaines de marqueurs polymorphiques ont été décrits comme résultat direct de la disponibilité des informations de séquence du Projet du Génome Humain (*Human Genome Project*) et des avancements des outils bioinformatiques pour l'exploitation des banques de données de séquences [Ayub et al., 2000].

En 1997, une communauté scientifique européenne a établi un ensemble de 9 microsatellites du chromosome Y définissant l'haplotype "minimal" et des marqueurs supplémentaires créant l'haplotype "étendu" [de Knijff et al., 1997 ; Kayser et al., 1997 ; Roewer et al., 2001]. En 2003, la communauté scientifique américaine (*Scientific Working Group on DNA Analysis Methods*, SWGDAM) a défini l'haplotype minimal incluant les 9 loci "européens" et 2 autres marqueurs, qui font partie des kits pour application en criminalistique commercialement disponibles aujourd'hui.

Les marqueurs bialléliques tels que les polymorphismes d'un seul nucléotide (*Single Nucleotide Polymorphisms*, SNP) ou les petites insertions ou délétions (indels) représentent une autre classe importante des marqueurs du chromosome Y. On se réfère parfois à ces marqueurs en les nommant les polymorphismes d'événement unique car leur taux de mutation est beaucoup plus faible que celui des STR ($\sim 10^{-8}$ contre $\sim 10^{-3}$ mutations par génération) [de Knijff, 2000 ; Kayser et al., 2000 ; Kayser et Sajantila, 2001]. Le premier marqueur biallélique du chromosome Y a été décrit en 1994 [Hammer, 1994]. L'utilisation de la chromatographie liquide à haute pression en condition dénaturante (DHPLC) par le groupe de Underhill a permis la disponibilité d'information sur plusieurs centaines de SNP du chromosome Y [Underhill et al., 1997 ; Underhill et al., 2000 ; Y-chromosome Consortium, 2002].

L'étude du chromosome Y présente deux avantages principaux : la spécificité de l'ADN masculin lors de l'analyse de mélanges d'ADN et la possibilité de suivre les lignées paternelles. Outre dans les recherches médicales comme celle de l'origine de l'infertilité masculine [Carvalho et al., 2003], les marqueurs du chromosome Y peuvent être utilisés pour de nombreuses applications telles que :

- l'analyse criminalistique d'indices dans le cadre d'agressions sexuelles [Sibille et al., 2002 ; Parson et al., 2003] en permettant l'amplification spécifique de l'ADN masculin. Ceci peut éviter les extractions différentielles pour séparer les

spermatozoïdes des cellules épithéliales ainsi que le masquage du profil masculin par le profil féminin ;

- la recherche de personnes disparues pour laquelle tout parent en lignée paternelle peut être utilisé comme échantillon de référence [Dettlaff-Kakol et Pawlowski 2002 ; Koyama et al., 2002] ;
- les tests de paternité déficiente pour relier les enfants masculins à une lignée paternelle [Santos et al., 1993 ; Jobling et al., 1997 ; Rolf et al., 2001] ;
- les études d'évolution et de migrations humaines, car l'absence de recombinaisons (région non recombinante du chromosome Y, *Non recombining region of the Y chromosome*, NRY) permet la comparaison d'individus masculins séparés par de longues périodes de temps [Underhill et al., 2001 ; Ke et al., 2001 ; Keyser-Tracqui et al., 2005] ;
- les recherches historiques ou généalogiques puisque dans la majorité des sociétés le nom de famille se transmet de père en fils [Foster et al., 1998 ; Jobling, 2001, Sykes et Irven, 2000 ; Trumme et al., 2004].

Pour tous les marqueurs uniparentaux, l'interprétation statistique est plus compliquée et des corrections appropriées tenant compte des sous-populations et des erreurs d'échantillonnage sont nécessaires [Roewer et al., 2000 ; Roewer et al., 2001].

I.1.3.2. L'ADN mitochondrial

La majorité du génome humain est contenue dans le noyau de la cellule. Cependant, les mitochondries, organelles cytoplasmiques qui fournissent l'énergie à la cellule, contiennent un petit génome circulaire ou chondrome de 16569 pb, portant 37 gènes codant pour des protéines impliquées dans le processus de la phosphorylation oxydative [Anderson et al., 1981 ; Andrews et al., 1999]. Chaque cellule contient plusieurs mitochondries contenant chacune plusieurs copies du chondrome. De ce fait, chaque cellule somatique comporte entre environ 200 et 1700 copies d'ADN mitochondrial (ADNmt) selon le type de tissus [Holland et Parsons, 1999]. Ainsi, suite à la mort cellulaire, il est plus probable de mettre en évidence une information issue de l'ADNmt, que celle contenue dans l'ADN nucléaire. L'analyse de l'ADNmt est une méthode efficace pour l'étude d'ossements, d'échantillons anciens contenant de l'ADN dégradé et particulièrement des tiges des cheveux ne contenant pas de bulbe [Holland et Parsons, 1999 ; Budowle et al., 2003].

Le chondrome humain a été séquencé en 1981 [Anderson et al., 1981]. Cette séquence (Genbank accession M63933) est aujourd'hui utilisée comme référence à laquelle est comparée toute nouvelle séquence d'ADNmt [Andrews et al., 1999].

La région contrôle, connue sous le nom de D-loop (*displacement loop*), présente un degré de variation inter-individus relativement important, rendant son analyse utile pour l'identification individuelle. De plus, le taux de mutation de l'ADNmt est de 5 à 10 fois supérieur à celui de l'ADN génomique. La méthode la plus répandue consiste en l'amplification par PCR puis le séquençage d'environ 610 pb sur deux segments qui sont particulièrement polymorphes, les régions hypervariables I et II (HVI et HVII). Il a été estimé qu'environ 1 à 3% des nucléotides disséminés sur HVI et HVII varient entre deux individus sans liens de parenté [Butler et Levin, 1998 ; Kogelnik et al., 1996 ; Brandon et al., 2005]. L'analyse de SNP situés en dehors de ces régions augmente le pouvoir discriminant de l'ADNmt [Vallone et al., 2004 ; Coble et al., 2004 ; Niederstatter et al., 2005 ; Divne et Allen, 2005].

Dans le cas de l'ADNmt, la pM moyenne étant élevée (de environ 0,005 à 0,025) [Budowle et al., 1999], il est préférable d'évaluer la signifiante d'une identité par la méthode de comptage [Parson et al., 2004] en examinant le nombre de fois où une séquence est observée dans une base de données spécifique d'une population.

L'ADNmt présente les caractéristiques suivantes :

- absence de recombinaison : les marqueurs (quasiment tous SNP mais aussi variants de longueur tels les stretch de C) n'ont pas une ségrégation indépendante ce qui réduit la diversité [Comas et al., 2004] ;
- transmission uniparentale (par la mère) : tous les membres d'une lignée maternelle partagent le même haplotype ;
- fréquences élevées de certains haplogroupes dans certaines populations [Richards et Macaulay, 2001 ; Comas et al., 2004] ;
- présence d'hétéroplasmies, en proportion plus ou moins importante : différentes séquences peuvent être trouvées dans les cheveux ou tissus d'une personne voire même dans des parties différentes d'un même cheveu ce qui rend l'interprétation des résultats compliquée [Tully et al., 2004 ; Grzybowski, 2000 ; Grzybowski et al., 2003]. Les mutations qui distinguent les types hétéroplasmiques sont particulièrement fréquentes à des sites particuliers appelés "*hot spots*" ce qui peut être incorporé à l'interprétation [Tully et al., 2001]. Une hétéroplasmie partagée par deux échantillons peut être un avantage et augmenter le pouvoir discriminant de l'ADNmt [Ivanov et al., 1996].

I.1.4. DETERMINATION DU SEXE

Le détermination génétique du sexe repose sur l'amplification d'une partie du gène de l'amélogénine présent sur les deux chromosomes sexuels X et Y avec une homologie de 90% entre les deux séquences homologues. Le gène de l'amélogénine a été séquencé sur le chromosome X et Y en 1991 [Nakahori et al., 1991a]. Une différence entre les séquences présentes sur chacun des deux chromosomes, due à de nombreuses délétions dans les régions répétées du gène, permet l'utilisation du gène de l'amélogénine pour la détermination du sexe [Nakahori et al., 1991b].

La méthode de sexage utilisée aujourd'hui est basée sur l'amplification d'une région du 1^{er} intron, située en dehors des régions recombinantes, présentant une délétion de 6 pb sur le chromosome X [Sullivan et al., 1993]. La taille des fragments amplifiés, 106 pb pour le chromosome X et 112 pb pour le chromosome Y, est parfaitement adaptée à l'étude d'échantillons tels qu'ils peuvent être rencontrés lors de l'identification médico-légale ou criminalistique et paléanthropologie, c'est-à-dire dégradés ou en très faible quantité. Actuellement, les kits commerciaux d'amplification des STR autosomaux incluent systématiquement l'amplification de cette portion du gène de l'amélogénine.

I.1.5. DOMAINES D'APPLICATIONS

Les profils génétiques, et *a fortiori* les marqueurs génétiques de manière générale, trouvent leurs applications dans différents domaines.

Dans le domaine médical, que ce soit lors du diagnostic prénatal [Camire et al., 2003 ; Shawky et al., 2002], la détection de maladies génétiques, la détection de sensibilité ou de résistance à des médicaments [Suzuki et al., 2005], ou la pharmacogénétique. Certains marqueurs, quelque soit leur nature (polymorphisme de longueur, de répétition ou de séquence), sont responsables de maladies génétiques ou peuvent y être liés. La dystrophie musculaire, la maladie de Huntington, le syndrome de l'X fragile, la fibrose kystique ou encore des maladies cancéreuses sont des exemples de maladies génétiques liées à des STR du fait de leur expansion [Panzer et al., 1995 ; Sinden et al., 2002]. Des marqueurs SNP sont aussi liés à des maladies génétiques ou leur prédisposition telles le diabète, la maladie de Creutzfeldt-Jakob, ou encore différents cancers [Silander et al., 2004 ; Croes et al., 2004 ; Kidd et al., 2005 ; Kammerer et al., 2005 ; Hemminki et al., 2005 etc.] Ces marqueurs peuvent alors être utilisés pour le diagnostic prédictif.

Les marqueurs génétiques non liés aux maladies permettent aussi la détection d'aneuploïdies chromosomiques [Quaife et al., 2004] ou l'existence d'un mosaïsme somatique dans le cadre du diagnostic prénatal.

Dans le domaine vétérinaire les empreintes génétiques permettent l'établissement de pedigrees, l'identification des espèces et les études phylogénétiques [Parson et al., 2000 ; Hebert et al., 2003]. Ceci est particulièrement important dans le cadre de la lutte contre le trafic d'individus appartenant à des espèces protégées par la convention de Washington et de la protection d'espèces en voie d'extinction [Martinez-Cruz et al., 2004]. L'étude génétique de fossiles permet de déterminer leur aire de répartition avant extinction de l'espèce et facilite ainsi sa réintroduction. L'histoire des populations animales, mais aussi végétales et humaines, et leur répartition peuvent ainsi être étudiées [Leonard et al., 2000]. L'identification de l'espèce ou d'un individu animal trouve aussi son application en médecine légale.

De même, dans le cadre d'élevage d'oiseaux la détermination génétique du sexe d'un individu est très utile chez certaines espèces qui ne présentent pas de dimorphisme sexuel apparent [Griffiths et al., 1998].

Dans le domaine agroalimentaire l'analyse génétique peut être utilisée dans le cadre de la certification des produits.

Enfin, les applications spécifiques des empreintes génétiques aux fins d'identification et de filiation regroupent les études des populations du passé et la médecine légale, englobant la criminalistique, les tests de filiation, les recherches de personnes disparues et l'identification des victimes de catastrophes de masse. Les réponses apportées par les empreintes génétiques telles qu'elles sont utilisées en médecine légale sont listées ci-dessous.

I.1.5.1. Transplantations

Les marqueurs génétiques pourraient permettre de déterminer l'origine d'un échantillon dans le cas de transplantations, domaine sujet au trafic d'organes.

De même, ils permettent le suivi de greffes de la moelle osseuse ou le profil génétique obtenu à partir du sang de l'individu transplanté (receveur), qui passe par 3 stades en cas de greffe réussie [Schraml et al., 2003]. Tout d'abord on détecte le profil du receveur seul, puis un mélange des profils du receveur et du donneur et enfin uniquement le profil du donneur.

I.1.5.2. Détermination du sexe

En médecine légale et dans le cadre de spécimens anciens, la détermination du sexe de restes squelettiques par l'analyse morphométrique peut être difficile voire impossible. En effet, cette approche basée sur des différences ostéologiques entre les deux sexes, outils de l'anthropologie physique dite classique, dépend du type d'ossements retrouvés et n'est donc pas toujours possible (par exemple dans le cas de squelettes incomplets) ni parfaitement fiable (sexage de squelettes juvéniles). La méthode la plus fiable, dite méthode de Bruzek, est basée sur l'étude des os coxaux. Cependant, il est souvent difficile de l'appliquer car ces os sont positionnés de manière verticale par rapport à l'axe du squelette. Ainsi, lorsqu'un corps n'est pas protégé par un cercueil ou un coffre, l'érosion endommage ces ossements. L'étude de l'ADN a rendu possible la détermination du sexe d'un individu de manière relativement fiable à partir d'éléments osseux ne permettant pas un sexage morphométrique. L'analyse de l'ADN ancien a permis d'étudier l'organisation des zones d'inhumations et des pratiques funéraires notamment en fonction de l'organisation suivant le sexe des individus.

I.1.5.3. Identification d'un individu et des liens de parenté

Un marqueur génétique est utilisable en médecine légale à condition de répondre à un certain nombre de critères. Tout d'abord, il doit suivre un mode héréditaire bien établi. Ainsi on distingue les marqueurs autosomaux, à hérédité mendélienne, et les marqueurs haplotypiques, à hérédité monoparentale et donc liée. Ensuite, il doit présenter un faible taux de mutation et être polymorphe, et par conséquent posséder un fort degré d'hétérozygotie, pour être informatif. Il ne doit pas présenter de déséquilibre de liaison avec les autres marqueurs sélectionnés ou avec des marqueurs de maladies génétiques. Il doit être phénotypiquement neutre c'est-à-dire non acteur dans la synthèse protéique. Enfin, il est nécessaire de connaître ses fréquences alléliques dans la population concernée. La technique d'analyse doit être simple, rapide, reproductible, peu onéreuse. Elle doit permettre la détection fiable des allèles et nécessiter peu d'ADN.

L'identification est l'étape préalable à toute étude de parenté ou de génétique des populations. L'application de l'identification individuelle (d'agresseurs ou de victimes) en criminalistique devient pratique courante car elle apporte un complément de preuves aux investigations policières. Cependant, il est important d'intégrer la détection de l'empreinte d'un individu sur

une scène de crime dans le contexte de ce dernier. En effet, il existe un risque non négligeable que celle-ci soit rapportée sur les lieux antérieurement ou postérieurement au crime.

Les empreintes génétiques sont utilisées pour les tests de parenté que ce soit dans un cadre civil, pénal ou de recherche en paléoanthropologie. Dans ce dernier domaine, l'établissement de relations de parenté est abordé classiquement par l'anthropologie physique grâce à l'analyse des caractères discrets considérés comme des variations phénotypiques non pathologiques. Une fois la relation de parenté déterminée, ces caractères ne permettent cependant pas la précision du lien familial existant. Certains de ces caractères ont une forte héritabilité, mais leur cause multifactorielle peut conduire à des résultats biaisés.

Les marqueurs de l'ADN mitochondrial permettent de retracer les lignées maternelles [Oota et al., 1995] et ceux du chromosome Y les lignées paternelles [Schultes et al., 1999]. L'analyse des marqueurs de l'ADN nucléaire (STR autosomaux) permet de déduire des liens de proche parenté (parents/enfants, grand parents/petits enfants, oncle/neveu, frère/frère, etc.) [Keyser-Tracqui et al., 2003]. Le mode de transmission mendélien de l'ADN nucléaire limite l'efficacité à des parentés proches (grand-parents et petits enfants, oncles neveux, cousins etc.). Lors de l'étude de parents plus éloignés, et dans le cas idéal, il convient d'étudier les 3 types de marqueurs génétiques (ADN nucléaire - autosomes et chromosome Y - et ADNmt).

Le recours à l'analyse génétique dans le cadre de recherches de personnes disparues et de l'identification de victimes de catastrophes de masse (guerres [Primorac et al., 1996 ; Huffine et al., 2001 ; Ladika, 2001 ; Birus et al., 2003], actes terroristes [Holland et al., 2003 ; Budimlija et al., 2003 ; Brenner et Weir, 2003 ; Budowle et al., 2005], crash d'avions [Ludes et al., 1994 ; Olaisen et al., 1996 ; Hsu et al., 1999 ; Leclair et al., 2004] ou encore catastrophes naturelles [Ladika, 2005]) permet des investigations et des conclusions qui seraient impossibles en leur absence. De plus, sur de tels sites, à l'image de sites anthropologiques où les corps sont souvent superposés et les restes mélangés, il est difficile d'attribuer correctement les restes à un individu [Komar, 2003].

I.1.5.4. L'ADN extrait de tissus anciens

L'étude de l'ADN "ancien" s'est développée relativement récemment. Durant les premières années, la grande majorité des investigations était restreinte à l'analyse de l'ADN issu de tissus mous provenant de momies. Or, les tissus anciens mous sont relativement rares et les modifications physico-chimiques de l'ADN présentent un frein au clonage. Ainsi, il fallut attendre la découverte de la technique d'amplification de l'ADN par PCR [Saiki et al., 1985],

son optimisation [Müllis et Faloona 1987 ; Saïki et al., 1988] et le développement de l'extraction d'ADN à partir de tissus durs [Hagelberg et al., 1989 ; Horai et al., 1989 ; Hanni et al., 1990] pour que les études en ADN ancien prennent leur essor. Contrairement au clonage, la PCR est peu sensible à de légères modifications chimiques de la molécule d'ADN et permet de travailler avec une quantité d'ADN beaucoup plus faible. Il est cependant démontré qu'en dessous de 1000 copies intactes les risques d'erreurs d'amplifications ne permettent pas d'authentifier les résultats [Handt et al., 1996]. De même le pouvoir d'amplification par PCR rend toute contamination par de l'ADN exogène fatale pour les éventuels fragments d'ADN ancien susceptibles d'être amplifiés. Dans ces cas, le clonage des produits d'amplification est utile pour révéler les différentes copies d'ADN présentes à l'origine [Cooper et Poinar, 2000]. Parmi les applications de l'ADN ancien, on peut souligner la génétique des populations, l'écologie, la climatologie, la médecine, l'archéologie et l'anthropologie moléculaires, l'évolution culturelle et comportementale [Brown et Brown, 1994 ; Cooper et al., 2001 ; Hofreiter et al., 2001 ; Kaestle et Horsburgh, 2002 ; Paabo et al., 2004 ; etc.] ou encore la paléopathologie, la paléoépidémiologie [Filon et al., 1995 ; Crubézy et al., 1998 ; Zink et al., 2001 ; 2005 ; Raoult et al., 2000] et l'étude de la domestication d'espèces végétales et animales [Watanobe et al., 2001 ; Jaenicke-Despres et al., 2003 ; Manen et al., 2003 ; Vouillamoz et al., 2003].

I.2. *Qualité et quantité des échantillons d'ADN*

Dans les cellules vivantes, l'intégrité de l'ADN est continuellement maintenue par des processus de réparation enzymatique [Lindahl, 1993]. Après la mort de l'organisme, les compartiments cellulaires qui normalement séquestrent les enzymes cataboliques éclatent, ce qui engendre la dégradation rapide de l'ADN par des enzymes telles les nucléases. De plus, la molécule d'ADN est alors exposée aux attaques bactériennes, fongiques et à celle des animaux saprophytes qui se nourrissent en dégradant les molécules [Eglinton et Logan, 1991 ; Mitchell et al., 2005]. Dans de rares circonstances, telle l'adsorption dans une matrice minérale, l'ADN peut échapper à la dégradation microbienne et enzymatique. Dans ces cas exceptionnels, l'ADN est cible de processus chimiques plus lents tels l'hydrolyse et l'oxydation [Lindahl, 1993]. La plupart de ces processus est identique à ceux existant dans la cellule vivante, mais ne sont plus contrebalancés par les systèmes de réparation enzymatique après la mort. Ainsi, les lésions s'accumulent de manière irréversible jusqu'à la perte de l'information sur la séquence nucléotidique. Occasionnellement, si la désintégration de l'ADN n'est pas complète, la technique PCR permet de révéler une partie de l'information.

L'importance des processus de dégradation (enzymatique et chimique) touchant la molécule d'ADN est sous l'étroite dépendance des facteurs environnementaux (humidité, température, pH, etc.) [Pääbo, 1989a ; Thomas et Pääbo, 1993 ; Burger et al., 1999]. La préservation de la molécule d'ADN varie donc considérablement en fonction des conditions environnementales. Il n'existe pas de corrélation entre l'âge de l'échantillon et le degré de préservation de l'ADN [Höss et al., 1996 ; Poinar et al., 1996] en deçà d'une limite théorique de 100000 ans [Hagelberg et al., 1989 ; Lindahl, 1993]. Au delà, seules des conditions de conservation exceptionnelles permettraient de préserver l'information génétique [Lindahl, 1993]. Cependant, les échantillons analysés ayant permis d'obtenir une séquence d'ADN authentique et reconnue par la communauté scientifique avoisinent les 50000 ans d'âge (mammouth conservé dans le permafrost) [Höss et Pääbo, 1993 ; Hagelberg et al., 1994 ; Höss et al., 1994].

Dans certains cas, il est possible de pallier la perte d'information induite par la dégradation de l'ADN, par l'amplification de multiples fragments courts chevauchants suivie de leur clonage puis séquençage ou encore par l'analyse de marqueurs ponctuel tels les SNP [Pääbo et al., 2004].

Les acteurs et facteurs de dégradations mentionnés ci-dessous ne forment bien entendu pas une liste exhaustive. Le plus souvent, le degré de conservation d'un prélèvement biologique dépend de l'interaction de différentes conditions du milieu dans lequel il se trouve.

I.2.1. DEGRADATION DE LA MOLECULE D'ADN

I.2.1.1. Ruptures de brin

Le type de lésions le plus évident sur l'ADN dans les échantillons biologiques est sa dégradation en fragments de taille généralement inférieure à 500 pb [Hofreiter et al., 2001 ; Pääbo, 1989]. Cette réduction de taille des fragments est due à la fois à l'action des nucléases et à d'autres processus non enzymatiques.

Les réactions d'autolyse dues à la rupture des compartiments cellulaires, libérant les enzymes des lysosomes (nucléases, protéases, etc.), débutent dès la mort de la cellule. Ces enzymes s'attaquent entre autre à la molécule d'ADN et aboutissent très souvent à l'éclatement des cellules. Les protéases s'attaquent aux protéines associées à l'ADN nucléaire et mitochondrial les rendant plus accessibles à l'attaque des nucléases [Kelman et Moran, 1996]. Celles-ci vont s'attaquer aux liaisons spots (endonucléases) et ainsi réduire la molécule d'ADN en fragments de petite taille puis enlever des nucléotides à l'extrémité de ces fragments (exonucléases).

Cette dégradation enzymatique est accentuée par l'action des enzymes des micro-organismes qui participent à la poursuite du processus de décomposition.

Des processus non enzymatiques aboutissent aussi à la réduction de la quantité et de la taille de l'ADN. Par exemple, des clivages hydrolytiques non enzymatiques des liaisons spots génèrent des coupures monobrin [Lindhahl, 1993]. De même, les liaisons N-glycosidiques reliant les bases azotées aux désoxyriboses peuvent être sujettes aux clivages hydrolytiques entraînant la perte des bases ce qui affaiblit l'hélice de l'ADN et favorise sa fragmentation. Enfin, les radiations ultraviolettes peuvent également induire des coupures monocaténares dans les acides nucléiques.

I.2.1.2. Lésions hydrolytiques

La perte des groupements amines des bases azotées par hydrolyse, c'est-à-dire la déamination, les rend incapables de s'apparier avec leurs bases complémentaires [Eglinton et Logan, 1991] entraînant par exemple des erreurs de réplication lors de la PCR. Le pH et la température sont deux des principaux facteurs influençant l'hydrolyse. Par ailleurs, elle est favorisée par un milieu acide mais ralentie par des forces ioniques élevées.

I.2.1.3. Lésions oxydatives

Les attaques oxydatives affectent à la fois les bases (induisant des mutations) et les sucres (entraînant une coupure des brins d'ADN) [Lindahl, 1993]. Leurs cibles étant les doubles liaisons des noyaux penta- et hexamériques, l'oxydation induit leur rupture. Le compartiment nucléaire de la cellule est un milieu anoxique alors que la mitochondrie est le centre du métabolisme oxydatif. Ainsi, le processus d'oxydation agit en priorité sur le génome mitochondrial plus vulnérable au stress oxydatif que l'ADN nucléaire [Yakes et Van Houten, 1997 ; Bennetts et Aitken, 2005]. Ces facteurs pourraient être impliqués dans la supériorité du taux de mutation de l'ADN mitochondrial comparé à celui l'ADN nucléaire [Wallace, 1989].

I.2.2. FACTEURS ENVIRONNEMENTAUX

Si l'on connaît dans les grandes lignes les résultats des dégradations auxquelles est soumis l'ADN, il reste indispensable de connaître le plus précisément possible les facteurs favorisant ou limitant ces dégradations de manière à sélectionner le matériel à analyser et connaître ses limites. Il est reconnu que le degré de préservation de l'ADN est dépendant des facteurs environnementaux auxquels les restes humains ont été soumis [Pääbo, 1989a ; Thomas et Pääbo, 1993 ; Burger et al., 1999].

Que ce soit sur une scène de crime ou sur les sites de fouilles archéologiques, il est bien évident que la réunion de l'ensemble des conditions favorables à une conservation optimale de l'ADN est difficile.

I.2.2.1. La quantité d'oxygène

L'oxygène étant indispensable aux processus d'oxydation, à l'activité et la prolifération de micro-organismes aérobies, les dégradations de l'ADN induites par ces mécanismes sont directement dépendantes de sa quantité présente dans l'environnement de l'échantillon biologique.

I.2.2.2. La température

Les essais menés par de nombreuses équipes suggèrent que parmi les facteurs qui régissent la conservation de l'ADN pour des spécimens de quelques centaines à quelques milliers d'années, la température joue un rôle prépondérant. Une corrélation positive a été montrée entre la latitude et la probabilité de retrouver de l'ADN endogène [Reed et al., 2003]. Ces constatations sont argumentées par les résultats des analyses réalisées à partir de matériels provenant de tombes gelées [Leonard et al., 2000 ; Ricaut et al., 2004] et l'échec ou la suspicion concernant les études des restes anciens originaires de régions chaudes [Marota et al., 2002].

I.2.2.3. L'humidité

La présence d'eau ou d'un degré d'humidité important favorisent l'hydrolyse et l'activité enzymatique des micro-organismes [Lindahl, 1993 ; Burger et al., 1999]. En conséquence, un milieu sec mais tempéré, amenant à une dessiccation rapide des tissus, serait favorable à la préservation de l'ADN et à l'efficacité des amplifications.

I.2.2.4. Le pH

Un pH neutre ou basique dans les échantillons ou dans le lieu d'enfouissement favorise la préservation de l'ADN [Lindahl, 1993 ; Burger et al., 1999]. Cependant, il arrive que, malgré un pH basique du milieu, l'échantillon lui-même présente un pH acide.

I.2.2.5. Les UV

Les photons du rayonnement ultraviolet (des UVC aux UVA) sont fortement énergétiques et peuvent initier des réactions photochimiques dans les biomolécules [Freeman et al., 1989]. Les principales lésions de l'ADN induites par l'exposition aux UVB (290-320 nm), qui sont les rayons traversant la couche d'ozone atmosphérique (contrairement aux UVC) et fortement absorbés par les protéines et les acides nucléiques (contrairement aux UVA) [Freeman et al., 1989], sont notamment la formation de dimères de cyclobutane-pyrimidine et de photoproduits de pyrimidine-pyridone [de Laat et al., 1996 ; Anderson et Parrish 1981 ; Kanofsky et Sima 1991]. D'autres changements non-dimériques peuvent être induits par les radiations UV, tels les photohydrates de cytosine ou encore des coupures monobrin de l'ADN [Ananthaswamy et Pierceall, 1990 ; Hussein, 2005].

I.3. Avantages et inconvénients des techniques actuelles

Le pouvoir discriminant élevé des STR lié au polymorphisme de chaque locus est leur avantage principal et en fait des outils de choix pour l'identification génétique [Lygo et al., 1994 ; Budowle et al., 1999 ; Butler, 2001]. Grâce à ces marqueurs, il est possible de générer des profils génétiques à partir de très faibles quantités d'ADN au moyen de la technique PCR. Cependant, dans de fréquentes situations en criminalistique (mis à part les prélèvements *in vivo*) et en paléanthropologie, les échantillons, soumis à la dégradation par les mécanismes cités ci-dessus avant leur analyse sont loin d'être de très bons candidats pour l'analyse par STR [Baret et al., 1988]. En outre, des contaminants environnementaux peuvent se mélanger aux prélèvements à étudier, induisant une perte de signal des STR de taille plus importante. Cette perte de signal, résultant en la perte d'allèles et/ou de locus complet (*allelic/locus dropout*) [Whitaker et al., 1995], peut être la conséquence soit de la présence d'inhibiteurs de PCR soit de la fragmentation de l'ADN. Le problème est exacerbé lors de l'utilisation de réactions PCR à haut degré de multiplexage à cause de la vaste gamme de taille des amplicons produits. Les kits STR commercialement disponibles pour les applications en criminalistique peuvent générer des amplicons variant en taille de 100 à 450 pb [Krenke et al., 2002].

Dans les cas où l'ADN nucléaire est trop dégradé empêchant l'analyse de STR, les laboratoires ont recours à l'analyse de la séquence de l'ADNmt [Butler et Levin, 1998 ; Holland et Parsons, 1999]. Cependant, l'étude de l'ADNmt est longue et ne permet pas une discrimination individuelle à cause de son héritage non-mendélien.

Un autre désavantage des STR, particulièrement prononcé lors des tests de paternité, est leur taux de mutation relativement élevé puisque de l'ordre de 1 à 5 mutations par 1000 transferts alléliques ou générations [Brinkmann et al., 1998 ; Lins et al., 1998 ; Mornhinweg et al., 1998 ; Szibor et al., 1998 ; Henke et Henke 1999 ; Thomson et al., 1999 ; Butler, 2001].

I.4. Polymorphisme ponctuel de séquence

Les SNP (*Single Nucleotide Polymorphisms*) sont des substitutions d'un seul nucléotide qui se produisent à des positions ponctuelles spécifiques dans le génome (Figure 1a), générant des allèles différents [Brookes, 1999]. Par définition, la fréquence de l'allèle le moins représenté doit être au moins égale à 1% ; dans le cas contraire il serait considéré comme une mutation. Ainsi, les petites insertions et délétions (indels), souvent comprises dans la dénomination SNP utilisée de manière très large, ne correspondent pas strictement à la définition ci-dessus (Figure 1b) puisqu'un tel événement induit un décalage du cadre de lecture de la séquence (frameshift). Ces polymorphismes sont présents sur aussi bien l'ADN nucléaire que sur l'ADNmt.

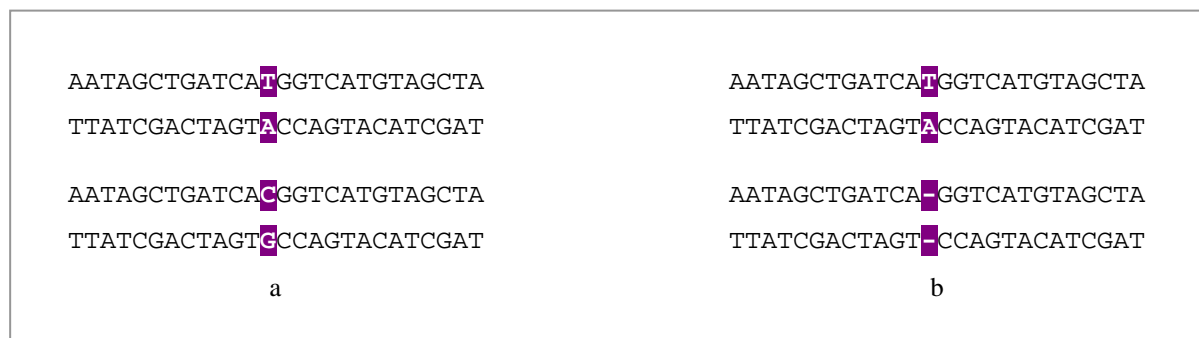


Figure 1 : Polymorphismes ponctuels de séquence.

a : Polymorphisme de séquence d'un seul nucléotide (SNP); b : Polymorphisme d'insertion ou de délétion d'un nucléotide (indel).

En principe, les SNP pourraient être des polymorphismes bialléliques, trialléliques ou tétraalléliques. Cependant, chez l'Homme, les SNP présentant 3 ou 4 allèles sont très rares, raison pour laquelle ils sont en général considérés comme des marqueurs bialléliques. Cette caractéristique s'explique par un taux de mutation faible [Nielsen, 2000] induisant une probabilité quasi nulle de double mutation sur un même locus. En moyenne, on trouve un SNP toutes les 500 à 1000 pb chez l'Homme et les SNP forment plus de 90% du polymorphisme génétique humain [Collins et al., 1998 ; International Human Genome Mapping Consortium, 2001 ; Venter et al., 2001].

A l'origine, les SNP sont issus de mutations survenues dans les cellules germinales et ayant échappé au système de réparation de l'ADN. Ainsi, la mutation a pu être transmise à la descendance, à condition bien sûr qu'elle ne soit pas létale. Parmi les substitutions possibles, on distingue les transitions, c'est à dire le remplacement d'une purine par l'autre purine ($G \leftrightarrow A$) ou d'une pyrimidine par l'autre pyrimidine ($C \leftrightarrow T$), et les transversions, c'est à dire le remplacement d'une purine par une pyrimidine ou inversement ($C \leftrightarrow A$, $C \leftrightarrow G$, $T \leftrightarrow A$, $T \leftrightarrow G$). Les polymorphismes les plus fréquents sont les transitions puisqu'elles n'impliquent pas le changement de noyau des bases azotées. Ainsi la taille et la structure globale du nucléotide ne sont pas modifiées et le système de réparation ne détecte pas le changement. De même, il semble que la prédominance de substitutions $C \leftrightarrow T$ ($G \leftrightarrow A$ sur le brin complémentaire) soit aussi liée aux réactions de déamination de méthylcytosines [Holliday et Grigg, 1993].

Selon l'endroit où ils apparaissent, ces polymorphismes peuvent avoir des effets différents sur le phénotype. La diversité nucléotidique, ou la proportion de nucléotides différant entre deux chromosomes choisis au hasard [Nei et Li, 1979], dans les exons est très inférieure à celle observée dans les régions non-codantes du génome. De plus, les SNP exoniques n'aboutissent à des changements non synonymes de codons que dans 50% des cas [Nickerson et al., 1998]. Par exemple, la diversité nucléotidique peut être très inférieure à 0,1% dans certaines régions hautement conservées du génome, alors que des régions particulières non-codantes des gènes HLA montrent des diversités nucléotidiques de 5 à 10%. Les SNP des régions codantes qui affectent la fonction ou la structure des protéines encodées peuvent avoir des répercussions cliniques. Les SNP des régions non-codantes peuvent aussi affecter la régulation des gènes ou présenter un déséquilibre de liaison avec une autre mutation ayant une conséquence phénotypique. Ces SNP sont, de ce fait, importants en médecine pour le diagnostic et le conseil génétique ou la spots [McCarthy et Hilfiker, 2000].

Cependant, la majorité des SNP est localisée dans les régions non codantes du génome et de ce fait n'a pas d'impact détectable sur le phénotype d'un individu. Ces SNP neutres sont très utiles en tant que marqueurs en génétique des populations [Nielsen, 2004 ; Marth et al., 2004], pour les études de l'évolution moléculaire [Zhao et al., 2003 ; Y chromosome consortium, 2002] et les analyses génétiques appliquées à la criminalistique [Gill, 2001 ; Gill et al., 2004].

I.4.1. INTERET DES MARQUEURS SNP AUTOSOMEAUX

La séquence d'analyse contenant les SNP peut être réduite à 50 à 60 nucléotides puisque le polymorphisme est ponctuel, alors que les STR utilisés aujourd'hui montrent des allèles allant jusqu'à 450 pb en taille [Krenke et al., 2002]. Cette caractéristique est d'un grand intérêt en criminalistique et surtout en anthropologie moléculaire puisque la taille des amplicons est critique pour l'étude d'ADN dégradés. Un autre exemple d'application de ces marqueurs est l'identification de fragments de corps de victimes de catastrophes de masse qui peuvent être fortement altérés. Les SNP offrent la possibilité d'approfondir les analyses lorsque l'étude des STR échoue, comme cela a été démontré dans l'identification de victimes de l'attentat du 11 septembre 2001 aux Etats Unis d'Amérique (*World Trade Center*).

Le taux de mutation des SNP, estimé à environ 10^{-9} - 10^{-8} chez l'Homme [Wang et al., 1998 ; Nielsen, 2000], est inférieur à celui des STR [Weber et Wong, 1993]. Ainsi, l'information fournie par les SNP peut porter sur des périodes plus longues que celles pouvant être étudiées par les STR ou les VNTR [Mountain et al., 2002 ; Brion et al., 2002]. Dans le cas particulier du chromosome Y, les SNP définissent des haplogroupes qui sont des groupes de séquences ou haplotypes génétiquement proches obtenus par l'analyse des STR.

Leur faible taux de mutation et leur dispersion très fréquente dans le génome [International Human Genome Mapping Consortium, 2001] font des SNP des outils attractifs pour les tests de paternité et les études en anthropologie [Reich et al., 2002]. En supposant l'absence d'allèles nuls, l'incompatibilité entre les génotypes sur un seul locus suffirait à rejeter l'hypothèse de paternité.

Les SNP étant des polymorphismes binaires, leur détection, analyse et interprétation à haut débit sont relativement aisées, adaptables à toutes les plateformes d'étude de l'ADN existantes et automatisables [Gut, 2001]. L'analyse en très haut débit rend plus faciles les études de populations nécessaires à l'estimation précise des fréquences alléliques. Ces dernières sont essentielles à l'interprétation correcte de l'analyse d'haplotypes dans le cas des analyses du chromosome Y ou de l'ADNmt.

I.4.2. TECHNOLOGIES : GENOTYPAGE DE SNP

Le développement de méthodes de génotypage, de chimie et de plateformes est permanent et très rapide. Chaque technologie présente des avantages et des inconvénients. De ce fait il est difficile d'en citer une comme étant idéale. Il est nécessaire de distinguer deux étapes dans toute réaction de génotypage de SNP : (1) la discrimination allélique, durant laquelle seront produit des analytes spécifiques de chaque allèle et (2) la détection des produits de celle-ci.

I.4.2.1. Technologies de discrimination allélique

I.4.2.1.1. Hybridation

L'hybridation de sonde allèle spécifique (*Allele specific oligonucleotide*, ASO) est basée sur la distinction de deux séquences ne différant que par une base par hybridation d'oligonucléotides complémentaires [Wallace et al., 1979]. La stringence de l'hybridation doit être optimisée afin d'obtenir la stabilité des hybrides sonde - cible parfaitement appariés seuls et non celle des hybrides contenant le mésappariement d'un nucléotide [Egholm et al., 1993 ; Nielsen, 2004].

I.4.2.1.1.1. PHARES MOLECULAIRES (MOLECULAR BEACONS)

Les sondes appelées phares moléculaires comportent trois régions : une région centrale, allèle spécifique et complémentaire de l'ADN cible, bordée de deux régions complémentaires entre elles portant à leurs extrémités les fluorophores (le fluorophore donneur (donor dye) en 5' et le fluorophore receveur (quencher dye) en 3'). Ainsi à l'état non hybridé, la sonde est repliée sur elle-même et les fluorophores étant rapprochés, aucun signal détectable n'est émis. Lors de l'hybridation avec l'ADN cible, la structure en épingle à cheveux s'ouvre en éloignant les deux fluorophores ce qui permet l'émission de fluorescence par le fluorophore donneur et sa détection [Tyagi et Kramer, 1996 ; Kostrikis et al., 1998 ; Tyagi et al., 1998] (Figure 2).

Cette technologie permet la détection de plusieurs séquences simultanément par l'utilisation de sondes couplées à différents fluorophores [Tyagi et al., 1998 ; 2000 ; Marras et al., 1999].

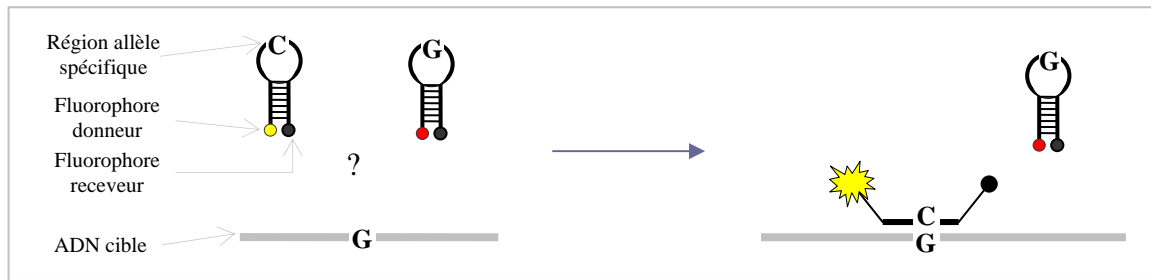


Figure 2 : Principe de la discrimination allélique par hybridation allèle spécifique de phares moléculaires.

Cependant, la capacité de détection de l'instrumentation est le facteur limitant du multiplexage. Le couplage de cette méthode à la technique d'amplification par PCR permet de suivre la production d'amplicons se traduisant par l'augmentation proportionnelle de la fluorescence (PCR quantitative).

I.4.2.1.1.2. TECHNOLOGIE 5'NUCLEASE OU TAQMAN

Cette méthode repose sur l'activité 5' nucléase de l'ADN polymérase impliquée dans la réaction de PCR. En effet, en progressant celle-ci va déplacer et cliver la sonde allèle spécifique hybridée à l'ADN cible (Figure 3).

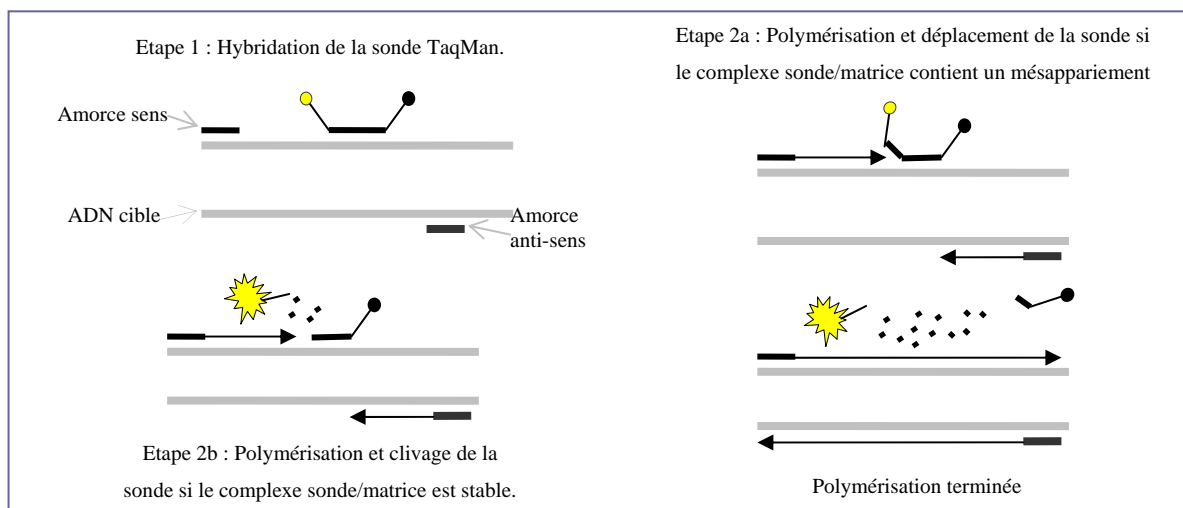


Figure 3 : Principe de la discrimination allélique par hybridation allèle spécifique de sondes TaqMan. Le fluorophore donneur est représenté en jaune et le fluorophore receveur en noir.

Cette sonde est dessinée de manière à permettre la capture de la fluorescence du fluorophore donneur (en 5' de la sonde) par le fluorophore receveur (en 3' de la sonde). Ainsi, lors du clivage de la sonde par l'ADN polymérase, les fluorophores donneur et receveur sont éloignés permettant la détection de fluorescence (PCR quantitative) [Holland et al., 1991 ; Livak et al., 1995 ; Kalinina et al., 1997 ; Lizardi et al., 1998 ; Livak, 1999]. La sonde hybridée avec un mésappariement à l'ADN cible (hybride instable) sera déplacée sans être clivée lors de la progression de l'ADN polymérase.

Cette technologie permet, dans la limite de la capacité de détection de l'instrumentation, la détection simultanée de plusieurs séquences par l'utilisation de sondes couplées à des fluorophores émettant de la fluorescence à différentes longueurs d'onde [Walker et al., 2005 ; Lee et al., 1999].

I.4.2.1.1.3. TECHNOLOGIE LIGHT CYCLER

Contrairement aux méthodes citées ci-dessus, cette technologie implique l'excitation et l'émission de fluorescence par le fluorophore receveur, dans ce cas nommé rapporteur. Elle implique l'hybridation d'une première sonde portant le fluorophore donneur et d'une deuxième sonde allèle spécifique portant le fluorophore rapporteur de manière adjacente (Figure 4).

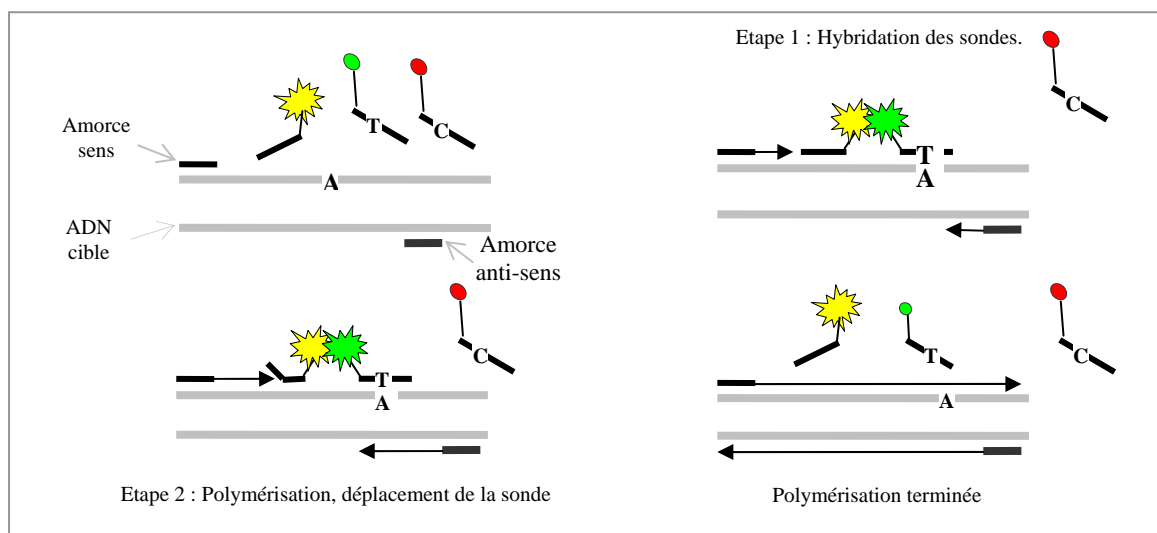


Figure 4 : Principe de la discrimination allélique par hybridation allèle spécifique et la technologie Light Cycler. Le fluorophore donneur est représenté en jaune et les fluorophores rapporteurs en vert et rouge selon l'allèle dont la sonde est spécifique.

En l'absence d'hybridation seule la fluorescence du donneur est détectée. Lorsque les deux sondes sont rapprochées dans l'hybride, l'émission du donneur excite le rapporteur qui à son tour émet de la fluorescence à une longueur d'onde différente. Lors de l'étape d'élongation de la réaction PCR, l'ADN polymérase déshybride les sondes, éloignant ainsi les fluorophores l'un de l'autre (PCR quantitative). Au cycle suivant les sondes vont reconnaître l'ADN génomique et les produits de l'amplification [Wittwer et al., 1997].

I.4.2.1.2. *Clivage invasif*

Le clivage invasif ou Invader Assay (Third Wave Technologies Inc., Madison, USA) est basé sur la spécificité de reconnaissance et de clivage enzymatique par la Flap endonucléase de la structure tridimensionnelle formée lors de l'hybridation parfaite de deux oligonucléotides chevauchants à l'ADN cible [Lyamichev et al., 1999 ; Lyamichev et Neri, 2003]. L'oligonucléotide invasif est complémentaire à la séquence en 3' du SNP. La sonde allèle spécifique est complémentaire à la séquence en 5' du SNP et à un allèle de ce dernier et porte en son extrémité 5' un fragment rapporteur non complémentaire à l'ADN cible. Ainsi le point de chevauchement des 2 oligonucléotides se situe au niveau du polymorphisme. Lorsque l'appariement entre la sonde et l'ADN cible est parfait, la sonde recouvre l'extrémité 3' de l'oligonucléotide invasif formant la structure tridimensionnelle reconnue et clivée par la Flap endonucléase, libérant ainsi le fragment rapporteur. Lorsque la base de la sonde n'est pas complémentaire à l'allèle de l'ADN cible, le mésappariement n'engendre pas la forme requise à la reconnaissance par la Flap endonucléase et le clivage n'a pas lieu. (Figure 5).

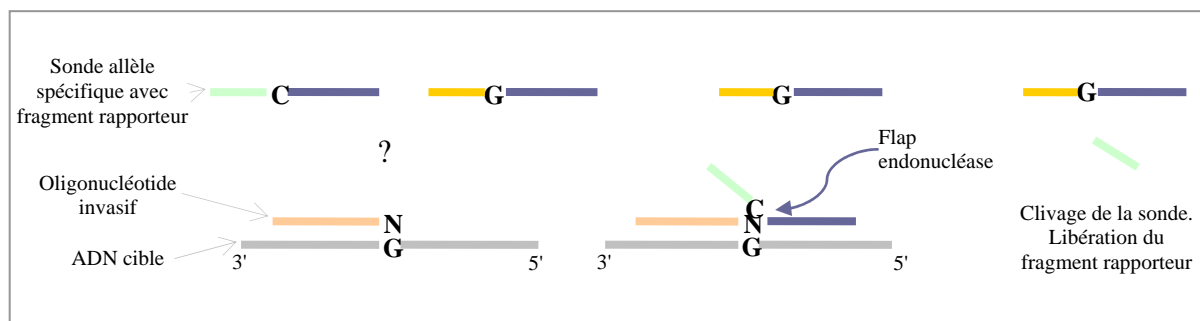


Figure 5 : Principe de la discrimination allélique par clivage invasif.

Le rapporteur ainsi libéré servira d'oligonucléotide invasif dans une deuxième réaction de clivage avec une sonde générant un signal détectable. La dénaturation rapide de la sonde clivée permet l'hybridation de sondes nouvelles et la production importante de fragments rapporteurs par la flap endonucléase. De part la double réaction, l'amplification du signal est exponentielle. Ainsi, cette méthode de détection ne nécessite pas de pré-amplification par PCR du fragment d'intérêt [Lyamichev et al., 1999 ; Ryan et al., 1999 ; Hall et al., 2000 ; Wilkins-Stevens et al., 2001 ; Rao et al., 2003] mais une quantité relativement importante d'ADN cible. Le couplage à une étape d'amplification en amont du clivage invasif permet d'augmenter la sensibilité de détection (PCR-Invader Assay) [Mein et al., 2000 ; Hsu et al., 2001]. Différentes méthodes peuvent être appliquées à la détection des produits de clivage, impliquant la construction de différentes sondes.

I.4.2.1.3. *Ligature*

La réaction de ligature ou *Oligonucleotide Ligation Assay* (OLA) est basée sur la liaison enzymatique de deux oligonucléotides adjacents lorsque les bases bordant le point de ligature sont parfaitement complémentaires à l'ADN cible [Landegren et al., 1988]. Dans le cadre de la discrimination allélique des SNP il s'agit de la ligature de l'oligonucléotide allèle spécifique et de la sonde spécifique de la séquence en 5' du polymorphisme par l'ADN ligase, enzyme spécifique de la réparation de liaisons phosphodiester (Figure 6).

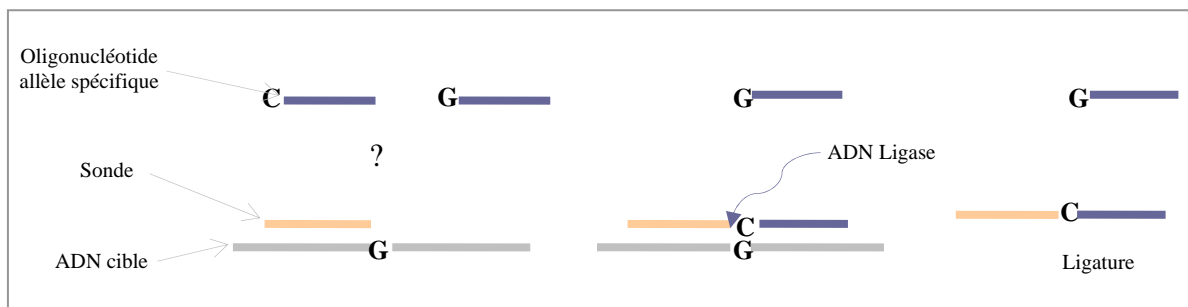


Figure 6 : Principe de la discrimination allélique par réaction de ligature.

L'utilisation d'une ADN ligase thermostable et de variations de température permet l'hybridation, la ligature et la déshybridation cyclique, ce qui engendre une croissance linéaire de la quantité de produits. A l'image du clivage invasif, cette méthode de détection ne nécessite pas de pré-amplification par PCR du fragment d'intérêt mais une quantité

relativement importante d'ADN cible. L'utilisation simultanée des deux brins d'ADN génomique comme cible permet une augmentation exponentielle de la quantité de produits. Cette variante de la technique est nommée *Ligation Chain Reaction* (LCR) [Barany, 1991 ; Benjamin et al., 2003] puisque les produits de la première réaction de ligature servent de cible au cycle suivant. L'utilisation de sondes Padlock permet aussi d'augmenter la sensibilité de détection. Dans ce cas, un des oligonucléotides est circularisé par ligature et peut être utilisé comme cible pour l'amplification circulaire ou *Rolling Circle Amplification* [Nilsson et al., 1994 ; 1997 ; Landegren et al., 1996 ; Baner et al., 1998 ; Lizardi et al., 1998].

I.4.2.1.4. RFLP

La méthode des RFLP, décrite au chapitre précédent, peut être appliquée à l'analyse de substitutions nucléotidiques puisque les enzymes de restriction sont spécifiques de séquences cibles [Pourzand et Cerutti, 1993 ; Parsons et Heflich, 1997]. L'utilisation d'enzymes de restriction spécifiques des séquences contenant le SNP d'intérêt engendre la digestion des produits PCR portant l'un des deux allèles uniquement (Figure 7).

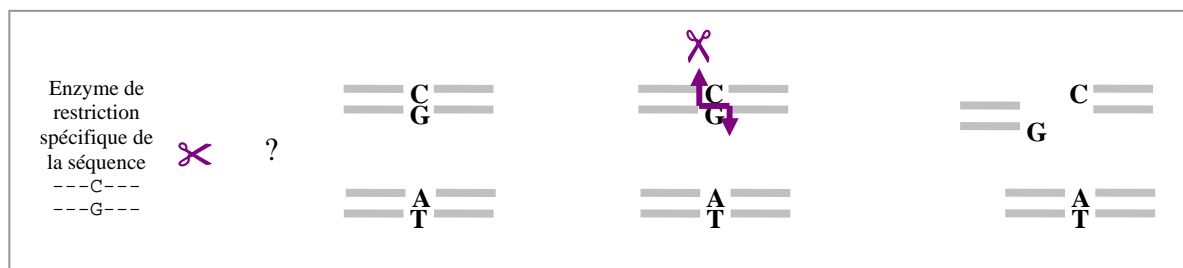


Figure 7 : Principe de la discrimination allélique par digestion spécifique de la séquence par une enzyme de restriction.

Cette méthode standard [Parsons et Heflich, 1997 ; Berger et al., 2002] ne permet pas un fort degré d'automatisation pourtant nécessaire à l'analyse multiplexée à haut débit des polymorphismes pour les raisons suivantes : (1) le nombre d'enzymes de restriction adaptées est limité ; (2) les fragments de restriction résultants peuvent engendrer des profils très complexes et difficiles à analyser et (3) des substitutions autres que le SNP d'intérêt peuvent apparaître dans le site de restriction et résulter en une non-digestion.

I.4.2.1.5. *Extension d'amorce*

L'extension d'amorce est l'une des méthodes les plus employées pour le génotypage de SNP et de mutations ponctuelles [Syvanen, 1999]. Cette technique robuste montre une grande flexibilité et nécessite le nombre d'oligonucléotides le moins important pour sa mise en oeuvre. Le marquage (p.e. fluorescence, radioactivité etc.) en vue de la détection est possible mais pas indispensable puisque les produits sont compatibles avec toutes les formes de détection. Les réactions d'extension sont réalisées de manière cyclique, ce qui engendre l'amplification de produits et l'augmentation du signal de façon linéaire [Tost et Gut, 2002].

I.4.2.1.5.1. EXTENSION DE SONDE ALLELE SPECIFIQUE

L'extension allèle spécifique implique l'hybridation de la sonde et son extension uniquement si son extrémité 3' est complémentaire à l'ADN cible. En utilisant deux sondes, chacune spécifique de l'un des allèles en 3', il est possible de déterminer laquelle a engendré la formation de produits et d'en déduire l'allèle présent sur l'ADN cible. Les produits ainsi générés peuvent être détectés par fluorescence sur microarray en utilisant des nucléotides marqués [Pastinen et al., 2000] ou des sondes marquées (Figure 8).

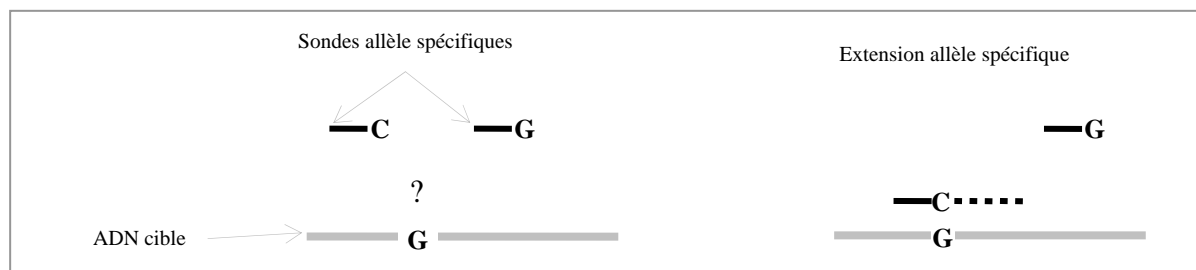


Figure 8 : Principe de la discrimination allélique par extension de sonde allèle spécifique.

Une variante de cette forme de discrimination allélique est la PCR allèle dépendante (*Amplification Refractory Mutation System, ARMS*) où le produit final est un amplicon double brin, dans le cas où la sonde allèle spécifique a trouvé son site d'hybridation [Newton et al., 1989 ; Ye et al., 2001 ; Chiapparino et al., 2004].

I.4.2.1.5.2. EXTENSION ALLELE SPECIFIQUE D'AMORCE

Le miniséquençage ou l'extension allèle spécifique implique l'hybridation de l'amorce à une base en amont du polymorphisme et son extension de manière complémentaire à l'allèle présent sur l'ADN cible. Cette méthode permet de distinguer les variants avec plus de précision que la différence de stabilité thermique des hybrides d'ADN cible avec des sondes allèle spécifiques. Le taux d'erreur d'incorporation de nucléotides par l'ADN polymérase est de l'ordre de 10^{-6} en moyenne et est insensible aux légères variations de stringence du milieu de réaction.

L'extension peut se faire sur une seule base (le SNP) ou sur un nombre réduit de bases. L'extension d'un nucléotide est achevée par la présence dans le milieu réactionnel de didésoxyribonucléotides (ddNTP) uniquement (Figure 9). Dans ce cas les produits de discrimination allélique sont distingués selon leur nature par différents formats de détection.

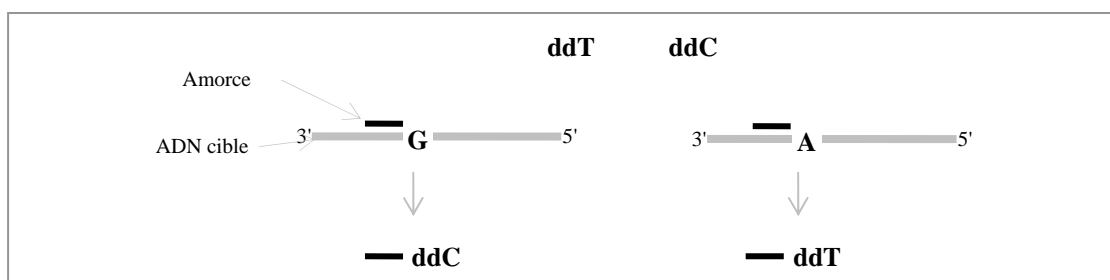


Figure 9 : Principe de la discrimination allélique par extension allèle spécifique d'amorce sur une seule base, selon le SNP uniquement.

Nous avons choisi pour cette étude de discriminer les allèles par extension allèle spécifique d'amorce en présence d'un mélange de dNTP et ddNTP [Braun et al., 1997 ; Fei et Smith, 2000 ; Buetow et al., 2001 ; Petkovski et al., 2005] (Figure 10).

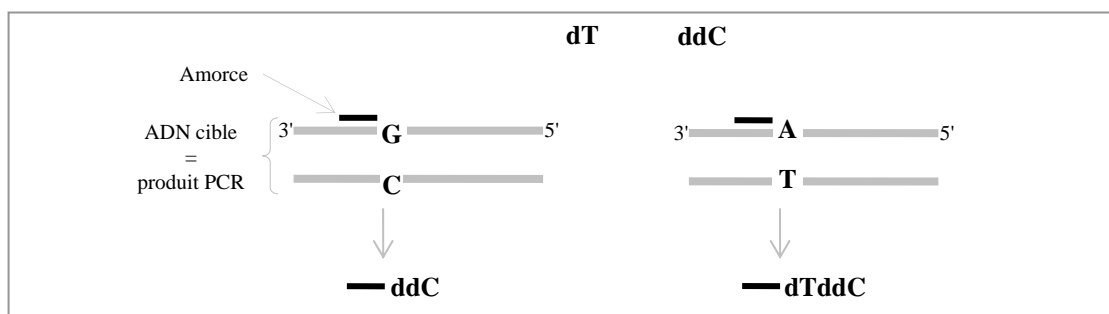


Figure 10 : Principe de la discrimination allélique par extension allèle spécifique d'amorce sur plusieurs base, selon le SNP et la séquence en 5' de ce dernier.

Cette méthode, bien que contraignante lors du design des réactions, apparaît (1) fiable, par la précision d'incorporation des nucléotides par l'ADN polymérase, (2) le plus directe, par la mesure de la masse moléculaire des produits qui ne nécessite aucun marquage, (3) le plus simple, permettant la distinction de tous les génotypes possibles sans ambiguïté et ne nécessitant aucune modification d'amorces ou de nucléotides et (4) modifiable sans contraintes.

I.4.2.2. Méthodes de détection des produits de la discrimination allélique

I.4.2.2.1. *Electrophorèse et détection directe de fluorescence*

L'utilisation de cette combinaison de méthodes permet de différencier les produits selon deux critères. L'électrophorèse permet de séparer les différentes molécules selon leur taille et la détection de fluorescence permet de distinguer des produits de même taille mais marqués par des fluorophores différents. Ceci engendre une méthode très flexible permettant la détection simultanée de différents produits (multiplexage). Cependant, des étapes de purification et de séparation efficace des réactifs non incorporés sont indispensables.

Lors de la détection de produits de ligature, il est possible de marquer les oligonucléotides allèle spécifiques par des modificateurs de mobilité et la sonde par un fluorophore [Grossman et al., 1994 ; Baron et al., 1996]. Les fragments obtenus par la méthode RFLP sont séparés selon leur taille et détectés par fluorescence en marquant les amorces PCR par des fluorophores ou en utilisant un intercalant de l'ADN lors de la migration (Bromure d'Ethidium ou SybrGreen). L'électrophorèse capillaire automatique couplée à la fluorescence balayant différentes longueurs d'ondes permet de déterminer la longueur de fragments RFLP de manière sensible et précise. L'incorporation de nucléotides marqués par des fluorophores permet de séparer et détecter par électrophorèse les produits d'extension d'amorce [Sanchez et al., 2003 ; Quintans et al., 2004 ; Brion et al., 2005].

I.4.2.2.2. *Puces à ADN et détection directe de fluorescence*

Les puces à ADN permettent de séparer les produits de la réaction de discrimination de manière spécifique et selon des critères de séquence alors que la détection proprement dite est basée sur la fluorescence [Leykin et al., 2005 ; Liljedahl et al., 2004 ; Wong et al., 2004]. En

effet, la réaction de discrimination peut se faire en milieu homogène en ajoutant aux produits une signature (queue de séquence) dont la séquence complémentaire se trouve fixée sur la puce. Ainsi, une fois la discrimination allélique achevée les produits sont "triés" et fixés sur la puce par hybridation de la signature à sa séquence complémentaire. D'autre part, la réaction de discrimination allélique peut se faire directement sur la puce si les sondes y sont fixées préalablement [Wang et al., 1998 ; Pease et al., 1994]. Les deux formats permettent une bonne séparation des produits mais la fixation des oligonucléotides (sondes ou complémentaires de la signature) sur la puce réduit sa flexibilité.

Lors de la détection des produits de ligature ou d'extension d'amorce sur microarray, il est possible d'utiliser des sondes biotinylées avec des oligonucléotides ou des nucléotides couplés à des fluorophores [Landegren et al., 1988 ; Barany, 1991 ; Samiotaki et al., 1994, Shumaker et al., 1996 ; Pastinen et al., 2000].

I.4.2.2.3. *Transfert résonant d'énergie de fluorescence*

Le transfert résonant d'énergie de fluorescence (*Fluorescence Resonance Energy Transfer*, FRET) se produit lorsque deux fluorophores sont très rapprochés et lorsque le spectre d'émission de l'un d'entre eux recoupe le spectre d'excitation de l'autre [Clegg, 1992]. En effet, la fluorescence du fluorophore donneur (donor dye) peut soit exciter un fluorophore rapporteur (reporter dye) qui émettra de la fluorescence à son tour, soit être absorbée par le fluorophore receveur (quencher dye). Ainsi la détection de la discrimination allélique est basée sur le passage des entités d'un état éteint à l'état fluorescent ou inversement. Malgré le coût important engendré par la synthèse de sondes marquées, le FRET est une forme de détection avantageuse. C'est une technologie très sensible qui, lorsqu'elle est couplée à l'amplification par PCR, permet de suivre l'évolution des réactions en temps réel ou la mesure *in fine*. De plus, elle est compatible avec toutes les méthodes de discrimination allélique. En effet, c'est la méthode la plus couramment utilisée pour détecter les discriminations par hybridation. Lors de la détection de produits de clivage la sonde porte à l'extrémité 5' (extrémité du fragment rapporteur) un fluorophore donneur et à son extrémité 3' complémentaire de l'ADN cible un fluorophore receveur. Lors du clivage par la flap endonucléase, les deux fluorophores sont éloignés l'un de l'autre permettant la détection de la fluorescence du donneur [Fors et al., 2000 ; Ledford et al., 2000]. L'augmentation de FRET suite aux réactions de ligature est mesurée grâce à l'utilisation d'une sonde portant à l'extrémité 5' un fluorophore donneur (donor dye) et d'oligonucléotides allèle spécifiques

portant des fluorophores rapporteurs (acceptor dye) à leurs extrémités [Delahunty et al., 1996 ; Chen et al., 1998 ; Iannone et al., 2000]. Sur le même principe il est possible de détecter des produits d'extension d'amorce marquée de fluorophore donneur par des ddNTP portant des fluorophores rapporteurs [Chen et al., 1999].

I.4.2.2.4. Polarisation de fluorescence (FP)

Lorsqu'une molécule fluorescente est irradiée par une lumière de longueur d'onde d'excitation donnée, une partie de cette lumière est absorbée. La lumière absorbée est réémise quelques nanosecondes plus tard, mais à une longueur d'onde plus élevée (longueur d'onde d'émission). La polarisation éventuelle de la lumière émise dépend de la liberté de rotation de la molécule en solution, c'est-à-dire de la température, de la viscosité du solvant et du volume de la molécule. Une petite molécule, telle qu'un fluorophore, peut effectuer une rotation rapide avant l'émission de la lumière, entraînant une dépoliarisation de cette lumière émise. En revanche, une macromolécule fluorescente, telle qu'un oligonucléotide marqué par un fluorophore, effectue une rotation beaucoup plus lente. Ainsi, dans le laps de temps séparant l'excitation de l'émission, la macromolécule n'effectue qu'une petite rotation et la lumière émise est polarisée. Le volume moléculaire étant proportionnel au poids moléculaire, la polarisation de fluorescence est une méthode adaptée à la détection de changements de ce dernier. En principe, toute réaction de discrimination allélique menant à un produit de taille inférieure ou supérieure au réactif fluorescent initial peut être analysée par FP.

La détection par FP, méthode très sensible, nécessite une quantité moins importante de fluorophores que le FRET ou la fluorescence directe, ainsi le coût des sondes est moins important. De plus, il est possible d'utiliser la totalité du spectre de la lumière visible pour la détection. Lors de la détection des produits de clivage, la sonde porte un fluorophore à son extrémité 5' et son clivage induit la diminution de masse moléculaire de l'entité marquée [Hsu et al., 2001 ; Kwok, 2002]. Lors de la détection des produits d'extension d'amorce la polarisation de fluorescence est induite par l'incorporation de nucléotides marqués [Chen et al., 1999a]. La détection de tous les produits implique une forte augmentation du bruit de fond lors de la formation de produits aspécifiques [Kwok, 2001 ; Di Gaetano et al., 2004] et représente le désavantage majeur de cette méthode.

I.4.2.2.5. *Luminescence*

La luminescence est émise dans une réaction enzymatique ATP (Adénosine TriPhosphate) dépendante. Lorsque la production d'ATP est couplée à la réaction de polymérisation d'ADN, chaque incorporation de désoxyribonucléoside engendre l'émission de lumière [Ronaghi et al., 1996]. Apparentée au séquençage, cette méthode, connue sous le nom de pyroséquençage, implique une cascade de quatre réactions enzymatiques [Ronaghi, 2001 ; Ching et Rafalski, 2002 ; Agah et al., 2004 ; Lavebratt et al., 2004]. L'incorporation d'un nucléotide par l'ADN polymérase engendre la libération d'un pyrophosphate. Ce dernier est transformé en ATP par l'ATP sulfurylase et l'ATP est utilisé par la luciférase pour générer de la lumière détectable. Enfin, l'excès de nucléotides est dégradé par l'apyrase. Cette méthode génère un bruit de fond très faible mais les étapes enzymatiques ainsi que les substrats additionnels compliquent la procédure expérimentale et se répercutent sur le coût.

I.4.2.2.6. *Spectrométrie de masse*

Contrairement aux autres méthodes de détection qui déduisent la nature du produit de la discrimination allélique en analysant le devenir d'un analyte, la spectrométrie de masse permet de mesurer une propriété intrinsèque du produit : sa masse (Figure 11). En particulier, la technologie MALDI-TOF, détaillée au chapitre suivant, trouve son utilité dans l'analyse génomique.

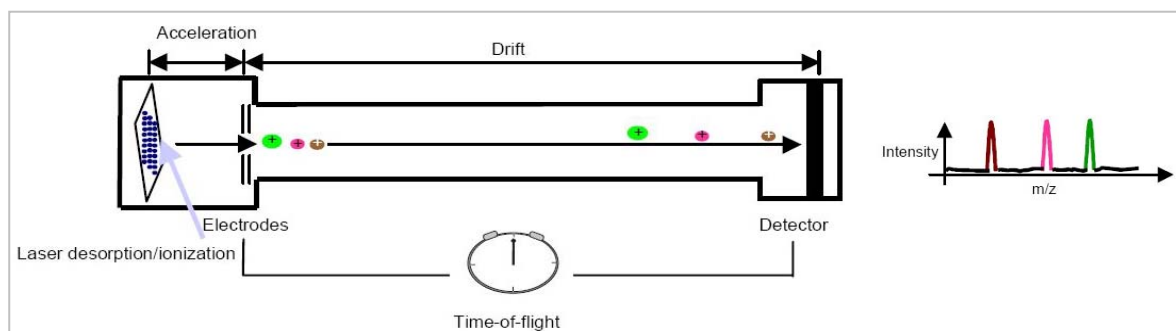


Figure 11 : Principe de la spectrométrie de masse MALDI-TOF [Pusch et al., 2002].

I.4.2.3. Possibilités d'enchaînement des technologies

La discrimination allélique et la détection des produits obtenus constituent deux étapes parfaitement individualisées. Les produits obtenus par des différents types de réaction peuvent être analysés par la même méthode et, inversement, les produits d'une même réaction peuvent être analysés par des méthodes différentes (Figure 12).

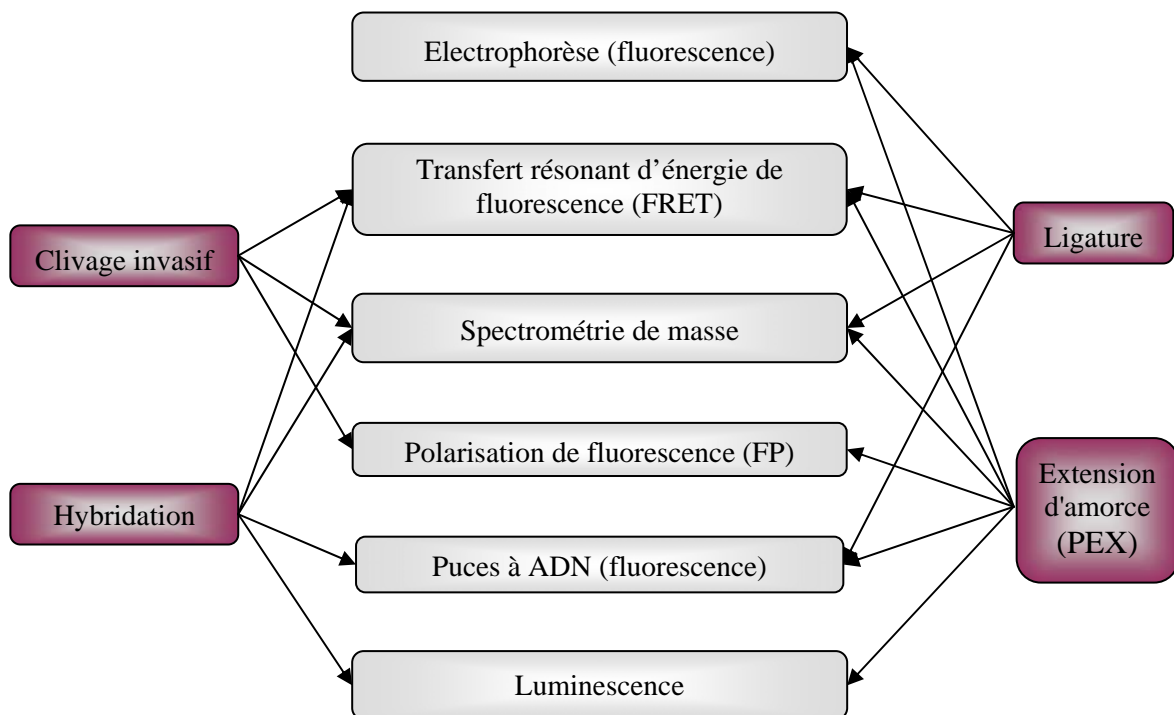


Figure 12 : Représentation schématique des différentes technologies de discrimination allélique (en violet) et de détection (en gris) de ses produits. Les flèches représentent les possibilités de compatibilité des deux étapes de l'analyse des SNP.

I.5. Spectrométrie de masse MALDI-TOF

L'étude structurale des molécules par spectrométrie de masse passe par une étape incontournable qui est leur ionisation: seules les formes ionisées peuvent être séparées et analysées selon leur rapport masse/charge à l'aide d'un champ électrique, variable ou non selon le type d'analyseur. L'étape de séparation fournit un spectre caractéristique des constituants.

I.5.1. HISTORIQUE

Jusqu'au milieu des années 80, l'ionisation induisait la fragmentation des molécules à analyser, soit par des processus très énergétiques (ionisation par impact électronique) soit par instabilité chimique (ionisation chimique). Ces méthodes d'ionisation permettaient l'obtention de nombreuses informations structurales, mais limitaient l'application de la spectrométrie de masse à l'analyse de petites molécules [Posthumus et al., 1978]. Au-delà de 1000 Da, les entités détectées étaient presque toujours des produits de dissociation de la molécule d'origine [Hillenkamp, 1983]. Cette restriction s'explique par la difficulté à vaporiser et ioniser les macromolécules par bombardement ou par chauffage. En effet, les macromolécules étaient le plus souvent pyrolysées avant même d'avoir pu être ionisées. L'utilisation d'une matrice comme médiateur pour les expériences de désorption/ionisation laser présentait des avantages évidents. Il fallait "simplement" rechercher des composés de petite taille, polaires, solubles dans les solvants adaptés aux macrobiomolécules, ayant une forte absorption à la longueur d'onde du laser utilisé et capables de co-désorber et de transférer leur charge à de plus gros composés bio-organiques.

En 1988, deux nouvelles méthodes d'ionisation dites "douces" ont été développées : l'ionisation ESI (*ElectroSpray Ionisation*) [Fenn et al., 1989] et l'ionisation au laser [Karas et al., 1987 ; Karas et Hillenkamp, 1988 ; Tanaka et al., 1988]. Grâce à ces deux techniques, l'analyse des macromolécules intactes est devenue possible, ouvrant un nouveau champ d'application à la spectrométrie de masse et plus particulièrement celui des biomolécules.

Le lien entre les deux méthodes trouve son origine dans le principe même des processus de désorption. Toutes deux conduisent à des espèces quasi microscopiques, les agrégats, mais qui résultent de conditions différentes. L'ionisation ESI produit des gouttelettes chargées

formées sous un champ électrique. L'ionisation MALDI engendre de la matière excitée par ablation laser d'une surface recouverte de matrice. Ces agrégats sont constitués d'un grand nombre de molécules de solvant ou de matrice qui stabilisent une (ou plusieurs) biomolécule(s) multichargée(s) selon leur taille et leur forme. Par le jeu de transferts d'énergie, une cascade de désolvatations conduit à la libération d'ions quasi moléculaires désolvatés, avec le minimum d'énergie interne. Ces étapes étant très rapides (de l'ordre de la microseconde), le processus de destruction moléculaire par pyrolyse se trouve limité dans cette période d'ionisation.

Les deux modes d'ionisation permettent une grande sensibilité (subfemtolaire) et une grande rapidité d'analyse (quelques secondes). Les deux principaux avantages de l'ionisation MALDI par rapport à l'ESI sont : (1) une meilleure tolérance aux sels non volatiles et aux tampons couramment utilisés en biologie moléculaire et (2) une moindre distribution de charges emportées par les ions produits.

La spectrométrie de masse MALDI-TOF (MALDI - Time of Flight) permet en pratique la mesure de masse de biopolymères de 1 à 500 kDa, même si en théorie sa gamme de masse est illimitée. Les molécules d'analytes sont incorporées à une matrice absorbant fortement à la longueur d'onde du laser. Les ions sont générés par excitation de la matrice par le laser. Cette source est généralement couplée à un analyseur TOF malgré le développement d'autres couplages par les constructeurs (MALDI-Ion Trap, MALDI-Q-TOF, MALDI-Q-Trap...).

La spectrométrie de masse MALDI-TOF s'adapte donc particulièrement bien à l'analyse de biomolécules tels que les peptides, les protéines, les oligosaccharides, les polymères synthétiques ou, comme dans le cadre de cette étude, à celle des oligonucléotides.

I.5.2. LA SOURCE MALDI

La technique MALDI utilise en général un faisceau laser pulsé travaillant dans l'UV pour désorber et ioniser un mélange de matrice et d'échantillon co-cristallisés sur une surface métallique. La matrice permet de minimiser la dégradation de l'échantillon provoquée par l'absorption de l'énergie du faisceau laser incident. Les ions obtenus sont ensuite accélérés afin d'entrer dans le tube de vol (Figure 13).

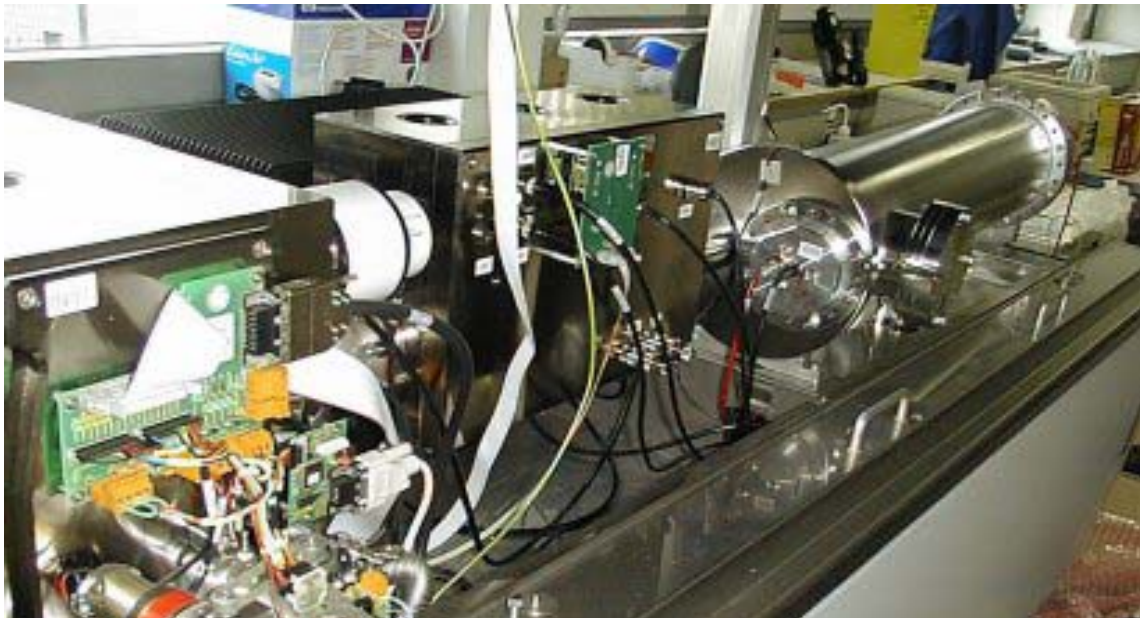


Figure 13 : Source et tube de vol de l'appareil Ultraflex TOF-TOF (Bruker Daltonics).

I.5.2.1. Ionisation

La désorption et l'ionisation MALDI comprennent l'absorption par la matrice de l'énergie fournie par le laser UV, l'ionisation de la matrice, la dissociation de la matrice, la sublimation, l'expansion de la matrice en un jet supersonique, l'entraînement de l'analyte dans le plasma d'expansion généré par la matrice et le transfert de charges vers les molécules d'analyte [Fenselau, 1997].

Si le processus d'ionisation reste peu connu, il est néanmoins bien établi que la protonation, la déprotonation et la cationisation se produisent en phase gazeuse [Glückman et Karas, 1999]. De plus, deux grandes phases dans le processus d'ionisation ont été décrites : l'ionisation primaire et l'ionisation secondaire [Knochenmuss et Zenobi, 2003].

Le résultat de l'ionisation primaire, qui est le produit de l'association entre un mécanisme de photo-ionisation dû aux protons absorbés par la matrice et un mécanisme thermique favorisé par la température élevée régnant dans le plasma d'expansion (environ 230°C), est la formation de molécules de matrice excitées (Figure 14). Les ions monochargés sont majoritaires dans les spectres MALDI obtenus à l'aide de laser à UV [Karas et al., 2000].

Le résultat de l'ionisation secondaire, qui aurait lieu dans le plasma d'expansion par transfert de protons en phase gazeuse [Zhigilei et al., 1998], est la formation de molécules d'analyte excitées (Figure 14).

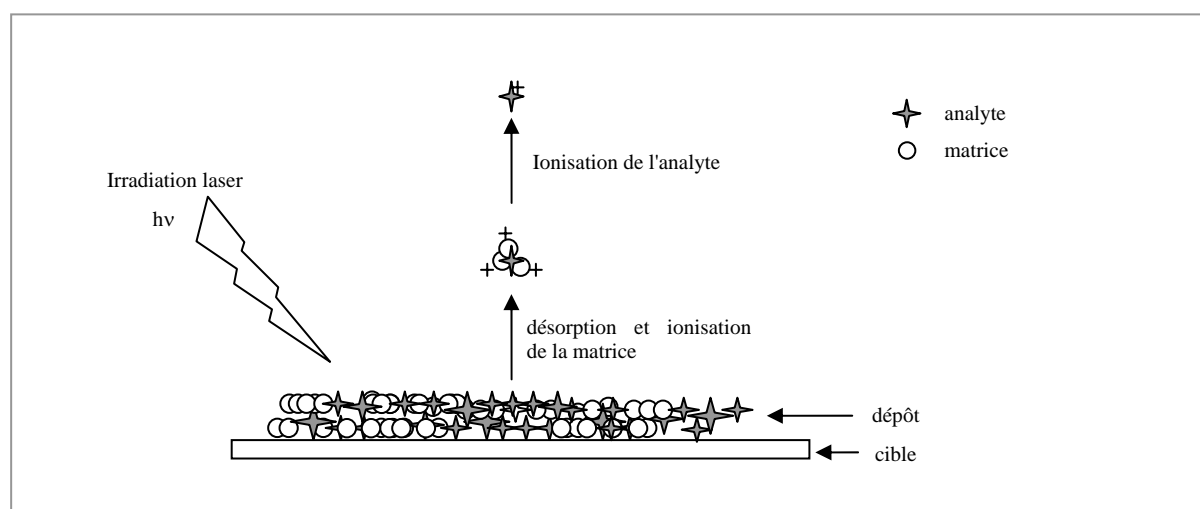


Figure 14 : Schématisation du processus de désorption / ionisation MALDI.

Durant ces processus, l'utilisation d'une matrice à affinité protonique inférieure à celle des analytes favorise thermodynamiquement la protonation de ces derniers. Cependant, l'énergie libérée par la réaction de transfert de protons en phase gazeuse peut induire la fragmentation des analytes. Pour les oligonucléotides, Zhu et al. [1995] stipulent que la perte de bases protonées suivie de la fragmentation du squelette mène aux fragments anioniques. Par conséquent, la différence d'affinité protonique de l'analyte et de la matrice détermine l'énergie disponible et, afin de réduire la fragmentation des oligodéoxynucléotides, le choix doit porter sur des matrices avec des affinités protoniques relativement élevées.

Les analytes ionisés sont accélérés par la différence de potentiel entre la cible (plaque porte-échantillon, IS1) et la lentille (grille d'extraction, IS2) en aval de la cible, typiquement 10 à 25 kV, pour finalement entrer dans le tube de vol qui est une zone de libre champ.

I.5.2.2. Extraction retardée

L'épaisseur du plasma implique une dispersion spatiale des ions à l'origine, ce qui provoque une distribution des temps d'arrivée.

La différence de localisation entre deux ions à l'application du champ d'accélération résultera en une différence de distance d'accélération, et donc en une vitesse différente dans le tube de vol. L'inversion transitoire (durant quelques nanosecondes) du champ en augmentant la tension de la lentille IS2 à une valeur légèrement supérieure à celle de IS1 permet une extraction retardée ou "delayed extraction", qui engendre la focalisation spatiale des ions et la correction de la résolution (Figure 14) [Brown et Lennon, 1995 ; Whittal et Li, 1995 ; Juhasz et al., 1996]. Cet équipement des spectromètres de masse MALDI a grandement amélioré la résolution des signaux de masse permettant la distinction de substitution de nucléosides dans la gamme de 1 à 8 kDa, correspondant à des longueurs d'oligonucléotides de 3 à 24 bases environ. Une fois focalisés, les ions sont ré-accélérés vers la lentille reliée à la terre avant de pénétrer dans le tube de vol sous vide et non soumis à une différence de potentiel (Figure 15).

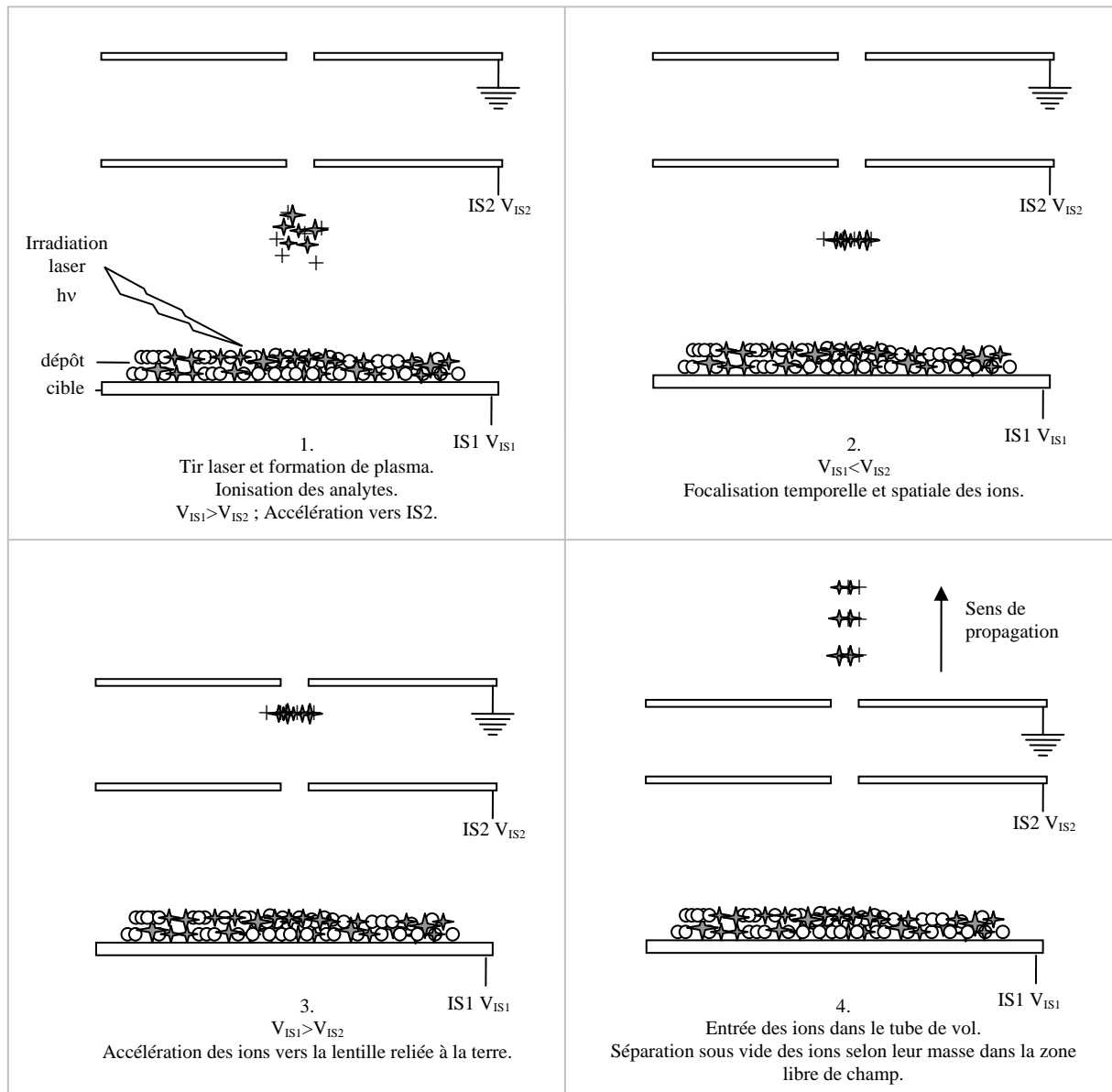


Figure 15 : Schéma du procédé de l'extraction retardée. Les analytes de masses différentes sont représentés par des étoiles de tailles différentes. La matrice est représentée par des cercles.

I.5.2.3. Matrices et modes de dépôt

La production d'ions par MALDI dépend de la bonne préparation du matériau composite, constitué de la matrice et du biopolymère analysé. Il est généralement admis que la qualité des spectres de masse dépend de certains facteurs liés à la préparation de l'échantillon :

- la méthode de préparation des cristaux,
- la sélection et la nature de la matrice,
- la présence de tensioactifs et/ou de tampons,

- le pH des solutions,
- les caractéristiques intrinsèques de la molécule d'analyte,
- le ou les solvants utilisés.

Un cristal est composé de molécules de matrice en large excès par rapport au nombre de molécules d'analyte présentes. Il peut être préparé de manière très simple : en laissant sécher sur une cible métallique porte-échantillon une gouttelette d'une solution aqueuse concentrée en matrice et diluée en analyte. Il existe cependant d'autres modes de préparation.

I.5.2.3.1. *Les matrices*

L'étape la plus délicate est la recherche d'une matrice. Le rôle de celle-ci est de permettre l'ionisation de l'analyte après qu'elle ait été excitée par les photons incidents du laser. Elle constitue un réservoir "tampon" d'énergie et transfère aux molécules d'analyte l'énergie nécessaire à leur ionisation. La sélection de matrices a été éminemment empirique car, à ce jour, nous ne connaissons pas de règles qui permettent de prédire, en fonction de la structure d'une molécule donnée, si celle-ci détiendra les propriétés d'une bonne matrice. Cependant, certains critères sont indispensables : ce petit composé organique acide, photosensible à cycle aromatique et de faible volatilité (10^{-7} mbar) doit former des cristaux incluant l'analyte. La matrice doit être soluble dans les solvants utilisés pour les macrobiomolécules et doit absorber (plus que l'analyte) à la longueur d'onde du laser (ici laser à azote qui émet dans l'ultraviolet à 337 nm). Le solide obtenu doit avoir une durée de vie dans le vide suffisamment longue pour permettre l'enregistrement du spectre c'est-à-dire qu'il ne doit pas sublimer dans les conditions de vide poussé qui règnent dans la source du spectromètre de masse. Le premier composé découvert fut l'acide nicotinique, compatible avec l'excitation à une longueur d'onde de 266 nm [Karras et al., 1987]. Par la suite, d'autres matrices adaptées à l'analyse des protéines ont été découvertes [Beavis et Chait, 1989 ; Strupat et al., 1992 ; Beavis et al., 1992]. L'acide 3-hydroxypicolinique (3-HPA) [Wu et al., 1993 ; 1994] seul ou mélangé à de l'acide pyrazinecarboxylique [Zhou et al., 2004] ou à de l'acide picolinique [Tang et al., 1994] ou encore la 2,4,6-trihydroxyacetophenone (THAP) conviennent à l'analyse des fragments d'ADN (figure 16).

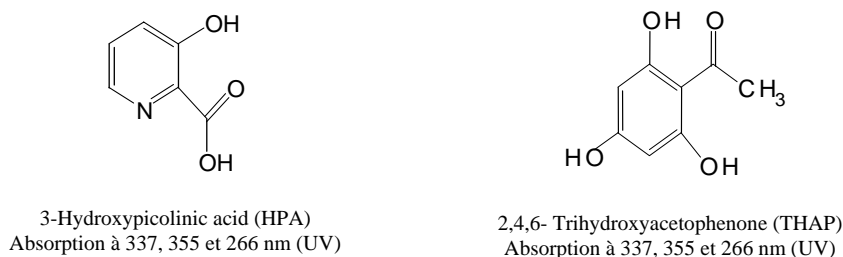


Figure 16 : Structure des matrices les plus utilisées pour l'analyse d'acides nucléiques.

Une limitation dans la détection des acides nucléiques par spectrométrie de masse est liée aux squelettes phospho-ribosés chargés négativement. Chaque liaison phosphate fournit une charge négative en solution et l'affinité des résidus phosphates pour les contre-ions utilisés classiquement dans les tampons PCR n'est pas suffisamment forte pour engendrer une saturation complète. Les sels interfèrent dans le procédé d'ionisation en induisant des adduits aux acides nucléiques, ce qui engendre la réduction significative des intensités du signal. L'ajout d'ions d'ammonium durant la préparation de la matrice permet de protéger les groupements phosphates de tels adduits [Pieles et al., 1993]. Cependant, la résolution satisfaisante s'appuie sur des procédés de purification stringente qui permettent d'éliminer ces ions et détergents pouvant perturber la cristallisation de la matrice.

I.5.2.3.2. Les modes de dépôt

Les méthodes de dépôts sont nombreuses. Depuis la préparation utilisée à l'origine [Karras et Hillenkamp, 1987], appelée méthode de la goutte séchée, qui permet de préparer un cristal de matrice dopé en analyte, de nombreux groupes ont proposé d'autres méthodes, telles la cristallisation lente pour former de grands cristaux de matrice enrichis en analytes [Xiang et Beavis, 1993], les préparations en "couches minces" pour former de larges films minces polycristallins [Vorm et al., 1994], ou les préparations en "sandwich" où l'échantillon est déposé sur un film de matrice avant d'être lui-même recouvert par une seconde couche de matrice [Kusmann et al., 1997]. L'utilisation du principe de l'électrospray pour réaliser des dépôts minces et réguliers en nébulisant sur la cible la solution d'échantillon a aussi été proposée [Axelsson et al., 1997].

Le dépôt doit être aussi régulier que possible. Des variations de l'épaisseur, dues à des irrégularités de surface, modifient la distance entre le plan du dépôt et la première grille d'extraction, ce qui provoque des variations de temps de vol des ions. Le même phénomène engendre des défauts de calibration externe ou pseudo-interne au dépôt.

Les matrices ont pour caractéristique de co-cristalliser sélectivement avec les biopolymères, ce qui permet de les séparer des contaminants (par exemple les sels).

Les dépôts d'échantillon sont hétérogènes. Dans le cas d'une hétérogénéité au niveau macroscopique, comme par exemple avec le 3-HPA, il est nécessaire de localiser les cristaux sur le dépôt. Cette recherche de "sweet spots" peut être un facteur limitant à l'automatisation.

I.5.3. L'ANALYSEUR TOF

I.5.3.1. Principe

Les sources MALDI sont en majorité couplées à des analyseurs TOF mais peuvent aussi être couplées à des trappes ioniques ou des analyseurs de type Q-TOF.

Le paramètre couramment utilisé pour caractériser les analyseurs est leur résolution R . Elle peut être définie par la capacité à distinguer une masse M par rapport à $M+\Delta M$ ou encore comme étant la précision de mesure de masse donnée par la relation $R=M/\Delta M$. Un analyseur TOF a une résolution de $R=20000$ et permet l'analyse dans la gamme de masse de $m/z=100000$, z étant la charge de l'analyte.

L'analyseur TOF sépare les ions en fonction de leur vitesse lorsqu'ils se déplacent sous vide dans un tube de vol. Dans la source, les ions sont accélérés par la même force F sur une même distance l entre les lentilles IS1 et IS2 et acquièrent tous la même énergie cinétique. Une population d'ions d'énergie cinétique commune, avec une distribution de masse donnée, a une distribution de vitesse définie. La vitesse moyenne de chaque ion est inversement proportionnelle à la racine carrée du rapport m/z de l'ion considéré.

I.5.3.2. Quelques formules fondamentales

Selon la relation fondamentale de la dynamique, le travail (W) de la force (F) appliquée sur un ion dans un champ (E) est donnée par la relation :

$$(1) \quad W = F \cdot l = z \cdot e \cdot E \cdot l = \frac{m \cdot (V_f^2 - V_0^2)}{2}$$

où ze représente la charge et m la masse de l'ion; l est la longueur du champ E .

V_f correspond à la vitesse de l'ion à la sortie de la zone d'accélération et V_0 celle à l'instant $t = 0$.

En considérant $V_f \gg V_0$, L la longueur du tube de vol libre de champ et t le temps mis par l'ion pour le traverser on obtient :

$$(2) \quad t = \frac{L}{\sqrt{\frac{2 \cdot z \cdot e \cdot E \cdot l}{m}}} \quad \text{ou} \quad t = \left(\frac{L}{\sqrt{2 \cdot e \cdot E \cdot l}} \right) \cdot \sqrt{\frac{m}{z}}$$

Les valeurs L , e , E et l étant communes et constantes pour tous les ions on peut simplifier la relation (2) :

$$t = b \cdot \sqrt{\frac{m}{z}} \quad \text{où } b \text{ est une constante.}$$

$$(3) \text{ Ainsi} \quad t \propto \frac{m}{z}$$

I.5.3.3. Mode linéaire

Une fois accélérés, les ions entrent dans un tube de vol (de longueur $L = 1 - 3$ m en général) libre de champ. Les ions se séparent dans ce tube de vol en fonction de la vitesse acquise dans la zone d'accélération, avant d'atteindre le détecteur placé à l'autre extrémité du tube de vol.

Pour une longueur de vol L et une tension accélératrice V_{acc} données, la durée écoulée t_{tot} , entre l'extraction des ions et leur arrivée au détecteur est proportionnelle à la racine carrée du rapport m/z des ions. Les ions de rapport m/z le plus petit arriveront les premiers.

Le spectre est obtenu en mesurant le signal enregistré par le détecteur comme une fonction temps-intensité pour chaque paquet d'ions envoyé. Comme tous les ions sont détectés, les analyseurs TOF ont une très haute sensibilité.

Une distribution d'énergie cinétique des ions à l'origine provoque une distribution des temps d'arrivée. Deux ions identiques, de vitesses longitudinales différents, accélérés à partir du même point, entreront dans le tube de vol à des vitesses et des temps différents. Il y a dispersion en temps et en énergie qui est l'une des limitations majeures de la résolution en mode linéaire.

I.5.3.4. Mode réflectron

La méthode de focalisation la plus efficace est le "réflectron" [Mamyrin et Shmikk, 1979]. Il est principalement constitué d'un miroir électrostatique formé par une série d'anneaux ou de grilles de potentiels croissants à l'extrémité du tube de vol, qui impose un champ électrique s'opposant à la progression des ions. Ce champ électrique est de direction opposée à la direction du champ accélérateur. Les ions les plus énergétiques pénètrent plus profondément dans le champ du réflectron et vont donc mettre plus de temps pour être réfléchis alors qu'ils sont arrivés avant les ions les moins énergétiques dans le réflectron. Ceci permet de focaliser tous les ions de même rapport m/z sur un plan. L'optique de l'instrument est ajustée pour que les ions soient légèrement déviés vers un autre détecteur hors de l'axe de déplacement longitudinal des ions. Le réflectron permet de corriger la dispersion en énergie cinétique des ions.

La focalisation de tous les ions de même rapport m/z permet aussi d'observer les ions métastables issus de la dissociation dans le tube de vol de certains ions ayant accumulé suffisamment d'énergie interne. Ces ions métastables ne peuvent être vus en mode linéaire puisqu'ils sont générés dans le tube de vol et gardent la même vitesse que l'ion parent. En revanche, le mode réflectron permet leur distinction car il implique leur focalisation énergétique [Chaurand et al., 1999].

I.5.3.5. Filtrage des ions de basse masse

La déflexion évite de surcharger le spectre obtenu en empêchant les ions de petite masse (typiquement 1 à 500 Da), telle la matrice, de parcourir le tube de vol. Elle a lieu à l'entrée du tube de vol et est achevée par l'application d'une tension durant un laps de temps très court entre deux plaques parallèles. Durant cet intervalle, tous les ions entrant dans le tube de vol seront attirés et capturés sur l'une des deux plaques.

L'option gating permet la non détection des ions de masse inférieure ou supérieure à un certain seuil. Dans ce cas, toutes les molécules atteignent le détecteur mais le signal correspondant à l'arrivée des ions n'est pas envoyé au digitaliseur. Ainsi l'analyse est centrée sur les analytes cibles tout en évitant le masquage de leur signal par celui d'un autre ion.

I.5.4. AVANTAGES ET INCONVENIENTS DE LA SPECTROMETRIE DE MASSE MALDI-TOF

- Compatibilité

La spectrométrie de masse MALDI-TOF est compatible avec tous les formats de discrimination allélique (Tableau 1, Figure 12).

Tableau 1 : Méthodes de discrimination allélique dont les produits peuvent être détecté par la spectrométrie de masse MALDI-TOF.

Discrimination allélique	Références
Ligature	Jurinke et al., 1996
Clivage	Liu et al., 1995 ; Griffin et al., 1999
Hybridation	Stomakhin et al., 2000 ; Griffin et al., 1997 ; Ross et al., 1997
Extension d'amorce	Haff & Smirnov ; 1997a ; 1997b ; Little et al., 1997 ; Sauer et al., 2000a ; 2000b ; Petkovski et al., 2005

- Mesure d'une propriété physique des analytes

Ne reposant que sur des propriétés intrinsèques des molécules, la spectrométrie de masse est une méthode de détection directe, ne nécessitant aucun marquage. Cette caractéristique offre une flexibilité plus importante et évite l'utilisation de marquage fluorescent (dont le coût est important) ou radioactif (dont l'utilisation est indésirable pour des raisons de sécurité). Il est cependant possible d'utiliser des nucléotides modifiés afin d'ajuster la gamme des masses à détecter.

- Possibilité d'analyse simultanée de différents produits (multiplexage)

Lors de l'analyse d'un grand nombre de SNP par échantillon, il est important de combiner leur détection en aussi peu de réactions que possible.

- Précision

Les spectromètres de masse MALDI-TOF sont des instruments de haute précision. Les masses de fragments courts peuvent être déterminées à seulement quelques unités de masse atomique près. En appliquant la calibration interne des spectres de masse, l'exactitude peut encore être améliorée menant à des résultats de génotypage sans aucune ambiguïté.

- Rapidité

Les spectromètres de masse MALDI-TOF disponibles aujourd'hui sont capables d'acquérir un spectre toutes les 2 à 5 secondes. L'acquisition de données peut être hautement automatisée et ne nécessiter aucune intervention du manipulateur. Le débit d'analyse peut monter à 12000 génotypes SNP par jour.

- Technologie adaptée aux approches de génomique fonctionnelle

L'analyse par spectrométrie de masse peut être adaptée à toute molécule. Ainsi, elle trouve son application en génomique et en protéomique. Cette pluralité de la méthode permet aussi la détection directe d'autres entités présentes dans le milieu réactionnel, tels les contaminants ou les inhibiteurs.

Pour autant, la spectrométrie de masse appliquée au génotypage présente également certains inconvénients.

- Susceptibilité aux adduits

Le fait que la technologie MALDI-TOF soit polyvalente et permette la détection de toute entité selon sa masse entraîne la nécessité d'étapes de purification rigoureuse du milieu réactionnel.

- Technologie innovante

L'instrumentation MALDI-TOF combine la haute performance avec la très haute sensibilité permettant son application dans différents domaines de la génomique et de la protéomique. Cependant, les réglages très fins nécessaires à l'adaptation à une application particulière nécessitent l'intervention de personnel expérimenté. Les appareils ne sont, pour le moment, pas adaptés à une utilisation en routine par des utilisateurs non spécialisés.

- Investissement important

Alors que l'analyse par spectrométrie de masse MALDI-TOF représente une méthode très abordable, l'investissement lors de l'acquisition et de la mise en place de l'environnement de l'appareil est très important.

- Cristallisation de la matrice

La cristallisation du dépôt de matrice et de l'échantillon, en particulier lors de l'analyse d'acides nucléiques avec l'acide hydroxypicolonique, est une étape qui n'est pas entièrement maîtrisée. Elle génère des "sweet spots" (des points de concentration de l'ADN) et un dépôt très hétérogène, ce qui présente un facteur limitant à l'automatisation.

I.6. *Problématique*

Le projet présenté dans ce mémoire est le fruit de la collaboration entre l'équipe d'accueil EA 3428 de paléoanthropologie moléculaire (Institut de Médecine Légale, Université Louis Pasteur) et du laboratoire Codgene (laboratoire d'identification génétique).

Dans le cadre de leurs travaux, ces deux entités sont fréquemment confrontées à des échantillons contenant de l'ADN dégradé et en très faible quantité, à partir desquels, dans le meilleur des cas, il n'est possible d'obtenir que des profils STR partiels. Afin de repousser les limites de l'identification et de la filiation génétiques, nous avons opté pour le développement d'une méthode qui complèterait celles appliquées en routine.

L'objectif de ce travail est la sélection et la validation d'un ensemble de 50 SNP autosomaux adaptés à l'identification et la filiation génétiques d'individus. Nous avons opté pour le développement de leur détection par extension allèle spécifique d'amorce ou miniséquençage (PEX), suivie de leur analyse par spectrométrie de masse MALDI-TOF. Bénéficiant de l'accumulation des avantages de ces marqueurs (taille du fragment d'intérêt), de la spécificité d'hybridation et d'amplification (PCR et PEX), de la tolérance des réactions aux légères variations des conditions de réalisation (mesure d'une propriété intrinsèque des analytes sans marquage) et de la sensibilité subfemtolaire de l'appareillage, cette combinaison de technologies trouve son application en criminalistique et anthropologie moléculaire. Le choix de méthodes compatibles avec, et pouvant être substituées par, toutes les autres techniques de détection et d'analyse actuellement connues, laisse une ouverture et de la flexibilité lors de l'étude des marqueurs sélectionnés et sa mise en application dans tout laboratoire.

Le travail sur la technologie MALDI-TOF a été possible grâce à une étroite collaboration entre l'institut de Médecine Légale de Strasbourg dirigé par le Pr. Bertrand Ludes et le Laboratoire de Spectrométrie de Masse Bioorganique (LSMBO) situé à l'Ecole de Chimie des Polymères et des Matériaux (Université Louis Pasteur), dirigé par le Dr. Alain Van Dorsselaer.

II. Matériel et Méthodes

II.1. *Choix des SNP*

II.1.1. SNP AUTOSOMAX

II.1.1.1. Critères de sélection

Le premier tri des SNP a été fait en fonction des fréquences alléliques (qui devaient être comprises entre 0,2 et 0,8). Les marqueurs adaptés à l'identification génétique doivent être silencieux, c'est-à-dire présents sur l'ADN mais sans expression ou influence phénotypique, et localisés dans des régions introniques non-codantes. Ensuite ont été retenus les SNP éloignés les uns des autres par au moins 4 Mpb. Enfin, la disponibilité des informations complètes, en particulier sur la variabilité de la séquence bordant le polymorphisme, était indispensable au dessin d'amorces pour l'amplification.

II.1.1.2. Sources

Les SNP candidats aux fins d'évaluation ont été principalement sélectionnés dans les bases de données disponibles publiquement tels le SNP Consortium Ltd (<http://www.snp.cshl.org/>) [International Human Genome Mapping Consortium, 2001 ; Thorisson et Stein, 2003] ou la base de données construite par Applied Biosystems exploitable par l'environnement de recherche "myScience" (<https://myscience.appliedbiosystems.com/genotype/search.jsp>) ou sa version téléchargeable SNPbrowserTM.

II.1.1.3. Aspects techniques

Les étapes de sélection et de validation des marqueurs SNP dans ce travail sont schématisées dans la Figure 17.

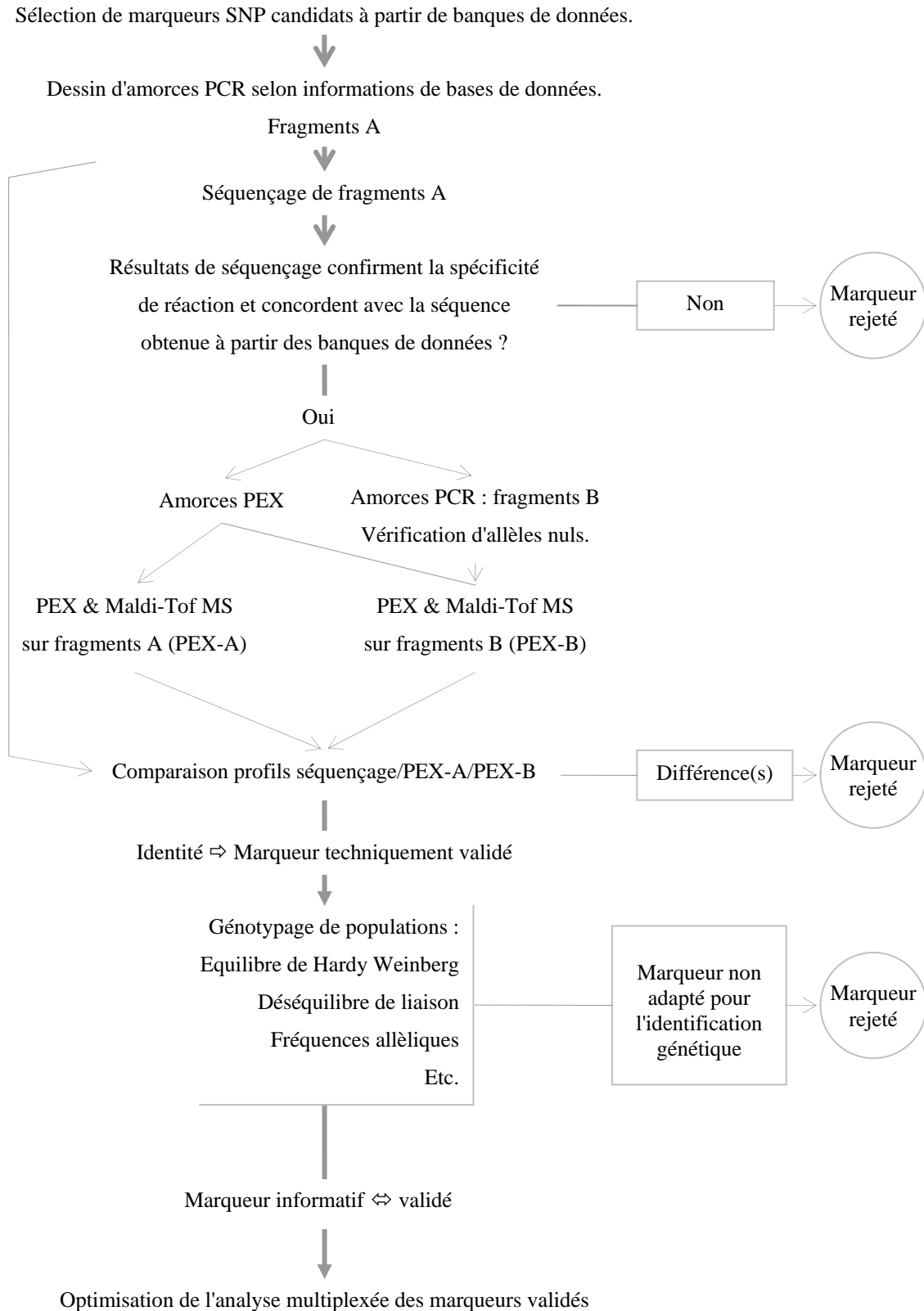


Figure 17 : Présentation schématique des étapes de sélection et de validation des marqueurs SNP utilisées dans ce travail.

II.1.2. SNP DU CHROMOSOME X (SNP-X)

Les marqueurs du chromosome X (SNP-X) ont été sélectionnés parmi des SNP rapportés comme étant dimorphes et présentant des fréquences alléliques se rapprochant de 0,5 [Taillon-Miller et Kwok, 2000] afin d'être évalués en tant que marqueurs en médecine légale et en anthropologie [Szibor et al., 2003]. Les marqueurs du chromosome X ont été étudiés par séquençage uniquement.

II.1.3. SNP DU CHROMOSOME Y (SNP-Y)

L'étude de populations anciennes à l'aide des marqueurs classiques (microsatellites nucléaires) permettent de déterminer des relations de proche parenté entre individus retrouvés dans des sites d'inhumation, révélant ainsi des pratiques funéraires. L'utilisation de marqueurs bialléliques qui présentent un taux de mutation inférieur à celui des polymorphismes de répétition permet de cibler des événements correspondant à des périodes plus longues. L'équipe d'accueil 3428 travaillant sur des spécimens anciens de Mongolie, Sibérie, Yakoutie et d'Amérique du sud, l'étude des SNP-Y a été concentrée sur 3 marqueurs (TAT, M242 et RPS4Y) présentant un intérêt pour étudier l'histoire et l'évolution de ces populations du passé (Figure 18) [Zerjal et al., 1997 ; Seielstad et al., 2003 ; Bergen et al., 1998].

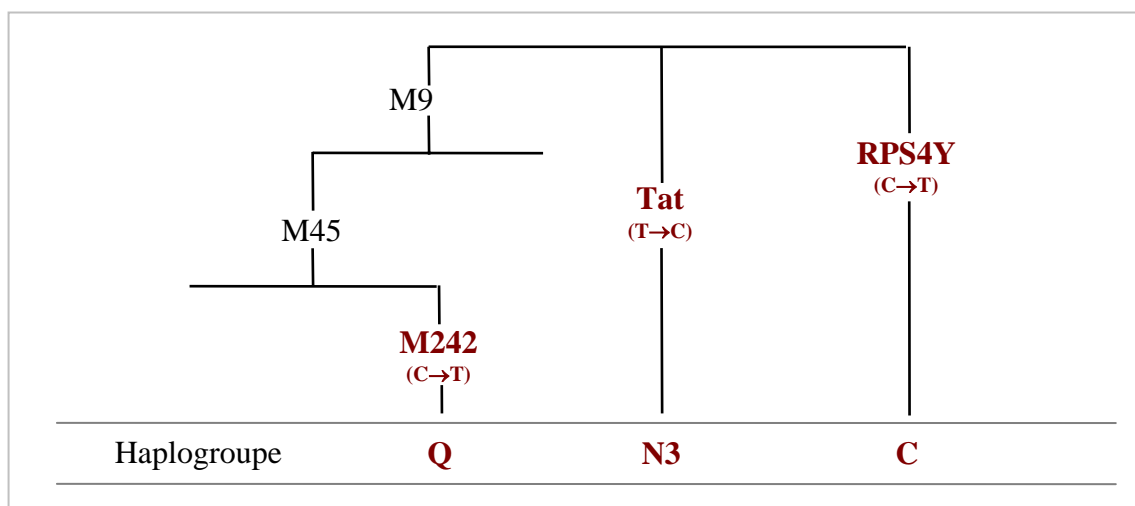


Figure 18 : Arbre phylogéographique montrant les SNP du chromosome Y étudiés ainsi que les haplogroupes qu'ils définissent.

II.2. Echantillons

II.2.1. CONSENTEMENT ECLAIRE

Ce travail a été réalisé conformément aux lois du 29 juillet 1994 et du 6 août 2004 relatives à la bioéthique.

Selon l'Article 16-10 du Code Civil, *"L'examen des caractéristiques génétiques d'une personne ne peut être entrepris qu'à des fins médicales ou de recherche scientifique". "Le consentement exprès de la personne doit être recueilli par écrit préalablement à la réalisation de l'examen, après qu'elle a été dûment informée de sa nature et de sa finalité. Le consentement mentionne la finalité de l'examen. Il est révocable sans forme et à tout moment"*.

L'article 16-11 du Code Civil dispose : *"L'identification d'une personne par ses empreintes génétiques ne peut être recherchée que dans le cadre de mesures d'enquête ou d'instruction diligentées lors d'une procédure judiciaire ou à des fins médicales ou de recherche scientifique. En matière civile, cette identification ne peut être recherchée qu'en exécution d'une mesure d'instruction ordonnée par le juge saisi d'une action tendant soit à l'établissement ou la contestation d'un lien de filiation, soit à l'obtention ou la suppression de subsides. Le consentement de l'intéressé doit être préalablement et expressément recueilli. Lorsque l'identification est effectuée à des fins médicales ou de recherche scientifique, le consentement de la personne doit être au préalable recueilli."*

Selon l'article L.145 du Code de la Santé Publique, *"L'examen des caractéristiques génétiques d'une personne ou son identification par empreintes génétiques, lorsqu'elle n'est pas réalisée dans le cadre d'une procédure judiciaire, ne peut être entrepris qu'à des fins médicales ou de recherche scientifique. Le consentement doit être recueilli par écrit après qu'une information loyale et intelligible ait été donnée (Article L145 du Code de la Santé Publique alinéa 2)"*.

Le consentement éclairé des personnes permettant la réalisation de ces travaux a été précédé d'une information appropriée sur la nature de l'examen proposé et mentionne la finalité de l'étude. Recueilli par écrit il est révocable sans forme et à tout moment.

Un exemple de consentement éclairé est présenté en annexe 1.

II.2.2. ECHANTILLONS D'OPTIMISATION ET VALIDATION TECHNOLOGIQUE

II.2.2.1. SNP autosomaux et du chromosome X

Les développements et optimisations ont été effectués sur 2 ADN féminins, 2 ADN masculins et un mélange des 4 ADN en proportions égales extraits à partir de prélèvements buccaux sur cytobrosses de personnes travaillant au laboratoire. Toutes les expériences ont été faites sur des témoins d'extraction (extraction sur eau ultra pure) et des témoins négatifs d'amplification. Après extraction, l'ADN a été quantifié sur gel d'agarose.

II.2.2.2. SNP du chromosome Y

Les développements et optimisations ont été effectués sur 11 prélèvements de différents types (prélèvements sanguins dans des tubes contenant de l'EDTA, prélèvements sanguins déposés sur support FTA[®] (Whatman Biosciences, Inc), prélèvements buccaux sur cytobrosse, etc.) sur des personnes de sexe masculin, dont l'ADN a été extrait à différents moments puis conservé dans des conditions adaptées (solution d'extrait à -20°C ou sur substrat à température ambiante). Toutes les expériences ont été faites sur des témoins d'extraction (extraction sur eau ultra pure) et des témoins négatifs d'amplification (eau ultra pure et ADN féminin).

II.2.3. ECHANTILLONS DE POPULATION FRANÇAISE

Les analyses SNP ont été réalisées sur des cas médico-légaux (71 prélèvements dont 20 cas de paternité confirmés et 5 exclusions) en complément de l'analyse des marqueurs STR. Ces familles françaises ne présentent pas de lien de parenté entre elles. En complément de cette population dont les descendants ont dû être exclus des calculs statistiques (fréquences, déséquilibres de liaison, etc.), les SNP ont aussi été étudiés sur 29 prélèvements buccaux de personnes sans lien de parenté. Tous les marqueurs SNP, autosomaux et du chromosome X, techniquement validés ont été étudiés sur un échantillon de 100 individus.

Il s'agit de 2 types de prélèvements, sanguins sur EDTA (anticoagulant) ou sur carte FTA[®] et buccaux sur cytobrosse. Afin d'uniformiser le type de matrice utilisée, les échantillons de sang ont été déposés sur du papier FTA[®] et les analyses effectuées à partir de pastilles.

II.2.4. ECHANTILLONS ANCIENS : ETUDE DES SNP-Y

Les SNP du chromosome Y ont été analysés sur des échantillons d'ADN masculin anciens, obtenus à partir de restes osseux découverts dans la nécropole de la vallée d'Egyin Gol en Mongolie du nord [Keyser-Tracqui et al., 2003]. Selon les résultats de la datation au carbone 14, ce site funéraire a été utilisé durant toute la période Xiongnu, du 3^{ème} au 2^{ème} siècle avant Jésus Christ.

II.3. De l'ADN dans l'échantillon à l'ADN dans le tube

II.3.1. EXTRACTION D'ADN

Quel que soit le type de matrice, le principe des techniques d'extraction d'ADN est sensiblement le même et comprend quatre grandes étapes : la lyse des cellules, la dégradation des protéines libérées, la séparation des acides nucléiques des autres composants et la purification de l'ADN.

Le protocole ci-dessous est utilisé pour l'extraction d'ADN à partir d'environ 1 mL de sang total, d'un prélèvement buccal sur cyto Brosse. Les cellules sont lysées par incubation de l'échantillon en présence de 3 mL de tampon de lyse (tampon Tris-HCl 10 mM pH 8, NaCl 300 mM, EDTA 10 mM, SDS 2% (p/v), urée 8 M, protéinase K 300 µg / mL) sous agitation douce à 37°C pendant la nuit. Après l'ajout de 20 µg / mL de RNase A, le mélange est incubé sous agitation douce à 56°C pendant une heure afin de digérer les acides ribonucléiques.

L'ADN ainsi libéré est extrait selon un protocole standard d'extraction organique (phénol/chloroforme/alcool isoamylique (25 : 24 : 1, v/v/v) [Ausubel et al., 2002].

II.3.2. PURIFICATION DE L'ADN

II.3.2.1. Précipitation à l'éthanol

La précipitation alcoolique permet d'ôter toute trace de chloroforme de la phase aqueuse, contenant l'ADN, prélevée suite à l'extraction organique. L'ADN est précipité selon un protocole standard [Ausubel et al., 2002], séché puis resuspendu dans 400 µL (pour le sang)

ou 200 µL (pour les cytobrosses ou les taches) de tampon TE light 1X (tampon Tris HCl 10 mM pH 8, EDTA 0,1 mM) ou d'eau ultra pure.

La purification par précipitation est utilisée lors du travail sur des échantillons contenant de l'ADN en abondance et non dégradé tels des prélèvements de sang frais ou de salive sur cytobrosses.

II.3.2.2. Purification à l'aide du kit Clean-Mix™ (Talent)

Reposant sur la grande affinité de l'ADN pour la silice [Höss et Pääbo 1993] en présence de thiocyanate de guanidine, le kit Clean-Mix™ (Talent) permet de laver l'ADN fixé sur la résine selon le protocole du fournisseur et de l'éluer, une fois purifié de tout contaminant, dans de l'eau ultra pure. Cette méthode permet de purifier des fragments d'ADN dont la taille est comprise entre 100 pb et 500 kpb, à partir du surnageant issu de l'extraction organique de matrices complexes tels les ossements, les dents, les prélèvements anciens ou dont la conservation dans des conditions optimales n'est pas garantie.

II.3.3. QUANTIFICATIONS DE L'ADN PAR PCR EN TEMPS REEL

Au vu de la disparité qualitative et quantitative des échantillons à partir desquels l'ADN peut être extrait, il apparaît indispensable de quantifier les extraits avant leur analyse.

La détermination quantitative d'ADN se base sur le nombre de cycles minimum (Ct) nécessaire à l'atteinte du seuil correspondant à une fluorescence donnée dans le tube réactionnel. Le seuil est déterminé pour correspondre à la zone d'amplification exponentielle de la réaction de PCR.

On distingue deux types de quantifications, l'une absolue se référant à une gamme de concentration connue et la seconde relative se référant à une autre cible interne à l'échantillon tel un gène domestique.

Dans cette étude, les quantifications absolues ont été effectuées grâce au kit Quantifiler™ Human DNA Quantification Kit (Applied Biosystems) basé sur la technique 5' Nucléase de la technologie TaqMan (décrit dans la section I.4.4.2.1.2). La gamme de concentration pouvant être détectée est comprise entre 23 pg et 50 ng pour 2 µL d'extrait. Les amplifications et lectures ont été faites au moyen de l'appareil ABI PRISM® 7000 Sequence Detection System (Applied Biosystems).

II.3.4. CONCENTRATION ET PURIFICATION PAR ULTRAFILTRATION

L'ultrafiltration sur microconcentrateur repose sur l'utilisation de membranes semi-perméables pour la séparation de molécules en solution en fonction de leur poids moléculaire. D'autres facteurs, telles la forme et la charge de la molécule, peuvent aussi jouer un rôle lors de la séparation. Les molécules de taille plus importante que celle des pores de la membrane seront retenues à sa surface et concentrées par ultrafiltration. L'avantage de ce système est de ne pas induire de dégradation mécanique des molécules tout en permettant leur purification et leur lavage. Les propriétés de rétention de la membrane sont exprimées par le seuil d'arrêt moléculaire (*Molecular Weight Cut Off*, MWCO) qui indique que les molécules dont le poids moléculaire lui est supérieur sont retenues à plus de 90%.

Par ultrafiltration, il est possible de concentrer des suspensions de manière très efficace sans induire de dégradation mécanique des acides nucléiques (par exemple lorsque l'extraction d'ADN n'a pas permis d'obtenir une suspension suffisamment concentrée pour son application).

Les microconcentrateurs utilisés dans cette étude sont les microcon® (Millipore) de MWCO adaptés à la taille des fragments cibles selon le protocole du fournisseur.

II.3.5. UTILISATION DE LA TECHNOLOGIE FTA® (WHATMAN, INC)

FTA® est une chimie brevetée de la société Whatman. Cette chimie sur matrice de cellulose permet d'une part la lyse des membranes des cellules déposées et de leurs organelles suivie du largage des acides nucléiques qui seront enchâssés et protégés dans les fibres du support (Figure 19). D'autre part, elle inactive les phages, bactéries et virus rendant ainsi inertes (pour le manipulateur et pour les acides nucléiques) les échantillons collectés. Une fois secs, les acides nucléiques sont protégés de la dégradation enzymatique, microbienne, oxydative ou par les radicaux libres, et peuvent être conservés durant plusieurs années à température ambiante. Un prélèvement sur carte FTA consiste en la découpe, à l'aide d'un emporte-pièce, d'une rondelle (*punch*) de 1,2 mm de diamètre imprégnée de sang (ou de salive) séché. Le punch est lavé à 3 reprises par 500 µL de solution de lavage FTA®, selon les recommandations du fournisseur, puis à 2 reprises par 200 µL de tampon TE light. Le punch est ensuite séché à température ambiante puis conservé transitoirement à 4°C ou utilisé directement en PCR.

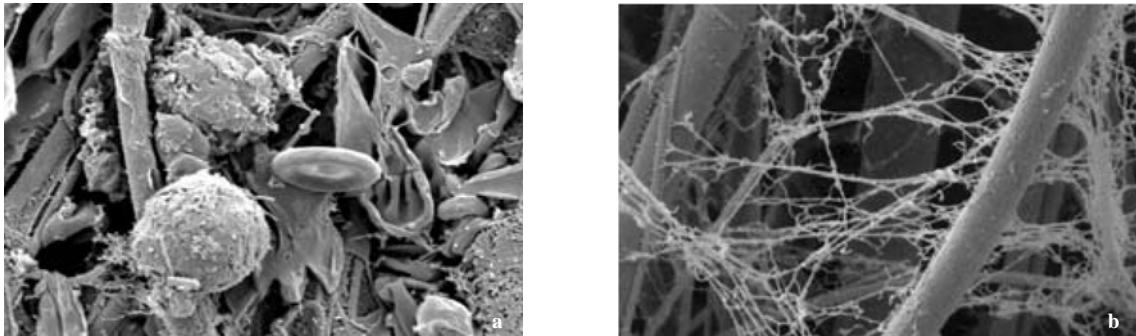


Figure 19 : Vue en microscopie électronique d'une portion d'un dépôt de sang sur un support FTA® (Whatman, Inc). a - dépôt frais de sang (1500 X); b - dépôt de sang sur une carte FTA® (Whatman Biosciences, Inc) après un lavage au réactif FTA (10000 X). Source : Whatman

II.4. Amplification par PCR

II.4.1. DESSIN ET CONSTRUCTION D'AMORCES PCR

La mise au point d'une réaction PCR spécifique et efficace dépend fortement du choix des amorces devant l'initier [Sanchez et al., 2005]. La recherche d'amorces a été faite manuellement avec le support du logiciel OLIGO-4.0-s [Rychlik et al., 1989] pour les calculs de température de fusion et d'énergie libre des structures secondaires en temps réel.

- La longueur des amorces doit être comprise entre 18 et 30 nucléotides pour assurer la spécificité lors de l'hybridation. En théorie, un oligonucléotide de 18 bases représente une séquence unique parmi 7×10^{10} nucléotides et devrait s'hybrider à une seule position dans la majorité des génomes eucaryotes comportant 10^9 à 10^{10} pb.

- Deux couples d'amorces non chevauchants pour la PCR ont été dessinés pour l'amplification de chaque marqueur, le premier de température de fusion (T_m) comprise entre 56 et 58°C et l'autre de T_m comprise entre 64 et 66°C [Borer et al., 1974 ; Suggs et al., 1981 ; Breslauer et al., 1986 ; Rychlik et al., 1990]. Lorsque la taille du fragment le permettait (minimum 50 pb avec le SNP positionné au centre de l'amplicon), les produits d'amplification obtenus ont été comparés par séquençage. Dans le cas contraire les résultats ont été comparés suite à l'extension d'amorce.

- Les structures secondaires induites par l'auto-, l'inter- et l'intra-complémentarité des amorces doivent être facilement rompues lors de l'étape de dénaturation, pour ne pas empêcher l'amplification. Ainsi, les amorces ont été sélectionnées de manière à présenter des énergies libres des liaisons prédites supérieures à -6 kcal [Breslauer et al., 1986].
- La composition des amorces en A et T d'une part et en G et C d'autre part doit être équilibrée (entre 40 et 60%) et homogène sur la longueur de l'oligonucléotide [Henegariu et al., 1997 ; Markoulatos et al., 2002].
- L'hybridation de l'extrémité 3' de l'amorce étant capitale pour l'initiation de l'amplification une "GC-clamp" y est ajoutée (1 à 3 guanines ou cytosines complémentaires à la séquence cible). Il est recommandé d'éviter une terminaison de l'amorce par une thymine en son extrémité 3' [Kwok et al., 1990 ; Sommer et Tautz, 1989].
- La taille des fragments à amplifier a été limitée à un maximum de 150 pb. La région des 11 nucléotides, comprenant le SNP et 5 bases de part et d'autre de celui-ci, a été exclue du champ de sélection des amorces. La perte de stringence due à l'hybridation des oligonucléotides pour la discrimination allélique sur ceux de l'amplification est ainsi minimisée. La limite inférieure de la longueur des amplicons a été fixée à environ 60 pb afin d'éviter le fort potentiel de contamination observé avec des fragments plus courts [Børsting et al., 2004].

II.4.2. OPTIMISATION DES REACTIONS DE PCR

Dans un premier temps, les fragments contenant les SNP d'intérêt ont été amplifiés individuellement (PCR single) dans le but de valider la spécificité de l'amplification, la séquence du fragment (par séquençage), le polymorphisme du SNP et la faisabilité de la réaction d'extension d'amorce. Ce n'est qu'après vérification de ces points que l'optimisation de l'amplification simultanée de plusieurs fragments d'intérêt a pu être envisagée.

Les réactions d'amplifications ont été développées dans des conditions identiques pour tous les fragments d'intérêt. D'une part, ceci a été accompli par le dessin d'amorces de T_m voisins, de pourcentages de GC équilibrés et de complémentarité faible. D'autre part, au vu de l'indépendance des facteurs primordiaux (la température d'hybridation (T_a) et la concentration en magnésium), le milieu de réaction a été optimisé selon un plan expérimental factoriel complet à 3 niveaux. Il en résulte une estimation non biaisée de tous les effets principaux et

des interactions dues à l'évaluation de chaque facteur sous un ensemble variable de conditions avec des répétitions internes.

II.4.2.1. Plan d'expérience

Les niveaux du facteur "concentration Mg^{2+} " ont été définis autour de la valeur recommandée par le fournisseur de la polymérase (2 mM, 2,5 mM et 3 mM). Les valeurs de T_a testées ont été définies autour de la valeur médiane de la fourchette préconisée au paragraphe II.4.1. (49°C, 51°C et 53°C pour le premier et 56°C, 58°C et 60°C pour le second couple d'amorces).

Cette première amplification devant servir à la validation des marqueurs, le résultat des tests, déterminé par électrophorèse sur gel d'agarose, peut être encodé de manière binaire. Le résultat est considéré positif si, pour toutes les amplifications testées dans les conditions définies, l'amplicon attendu est présent, des produits d'amplification aspécifique sont inexistantes et les dimères d'amorces sont minoritaires. Le résultat est considéré négatif si une de ces exigences n'est pas satisfaite.

Les expériences ont été faites en deux séries. Une première série complète d'expériences a été réalisée sur 15 couples d'amorces. Au vu des résultats obtenus, la deuxième série d'expériences sur les couples d'amorces restants était réduite aux tests de conditions ayant permis un résultat positif sur la première série.

II.4.2.2. Les autres paramètres de réaction PCR

Les fragments ont été amplifiés dans un volume réactionnel de 30 μ L à l'aide d'une unité de Taq polymérase (MBI fermentas, Lituanie) en présence du tampon de réaction sans magnésium (1X), de 0,32 μ M de chaque amorce, 200 μ M de chaque dNTP et de $MgCl_2$ à la concentration déterminée par le plan d'expérience. Les réactions de PCR ont été réalisées sur un appareil GeneAmp PCR System 9600 (Perkin Elmer) avec un programme composé de 5 minutes de dénaturation initiale suivies de 30 cycles de 30 secondes à 95°C, 45 secondes à T_a (cf. II.4.2.1.), 30 secondes à 72°C et finalisé par une extension de 2 minutes à 72°C.

II.4.2.3. Détection des amplicons

Les produits d'amplification ont été observés sous lumière UV après électrophorèse sur gel à 3% d'agarose standard contenant du BET, dans du tampon TBE 1X (89 mM Tris base, 89 mM acide borique, 2 mM EDTA, pH 8,3) [Ausubel et al., 2002].

II.5. Séquençage

Le séquençage des fragments contenant les SNP a été effectué d'une part afin de vérifier la séquence bordant le polymorphisme dans le but de la synthèse d'amorces pour la réaction d'extension d'amorce (premier lot d'amorces), d'autre part afin de comparer les résultats obtenus par les deux méthodes (les deux lots d'amorces) et enfin pour vérifier l'existence d'allèles nuls dus à une éventuelle mutation dans le site de fixation des amorces (séquençage des fragments amplifiés par les amorces du premier lot, grâce aux amorces du deuxième lot).

Les réactions de séquençage ont été réalisées grâce au kit BigDye® Terminator v 1.1 Cycle Sequencing Kit (Applied Biosystems) sur un appareil GeneAmp PCR System 9600 (Perkin Elmer) selon les recommandations du fournisseur. Les deux brins des fragments, amplifiés puis purifiés par ultracentrifugation sur colonne microcon®-30 (Millipore) et élués dans 20 µL de tampon TE light, ont été séquencés.

Les produits de séquençage ont été purifiés soit par précipitation à l'éthanol selon les recommandations du fournisseur du kit de séquençage et élués dans 20 µL d'eau ultra pure, soit grâce au kit Genopure oligo (Bruker Daltonics) auquel cas les produits ont été élués dans 10 µL d'eau ultra pure. Cette dernière technique permet d'obtenir un degré de pureté comparable à celui de la précipitation mais avec un gain de temps important. Dans ce cas, seul 2 µl d'éluât ont été déposés dans 18 µL d'eau en vue de l'électrophorèse contre 4 dans 16 µL d'eau pour les produits précipités à l'éthanol.

Les données brutes de séquençage obtenues après électrophorèse capillaire sur un appareil ABI PRISM® 3100 genetic analyzer (Applied Biosystems) sont interprétées grâce au logiciel DNA Sequencing Analysis (Applied Biosystems) puis analysées à l'aide du logiciel Sequence Navigator (Applied Biosystems) basé sur le programme d'alignements multiples Clustal-V.

II.6. Extension d'amorce (*Primer Extension, PEX*)

II.6.1. DESSIN D'AMORCES PEX

Le choix des amorces pour l'extension allèle spécifique d'amorce est limité par rapport à celui des amorces PCR. En effet, une amorce PEX doit s'hybrider immédiatement en amont (du côté 5') du polymorphisme. La longueur de l'amorce étant limitée par la température de fusion imposée, le choix entre les orientations sens et anti-sens sera guidé par la "qualité" de la séquence uniquement.

L'extension d'amorce (PEX) pour plusieurs SNP devant pouvoir se faire simultanément en une seule réaction (multiPEX), la même T_m a été imposée pour toutes les amorces. Les produits PEX étant analysés sans modification (par exemple coupure par enzyme de restriction ou photoclivage pour réduire leur taille), une T_m de 56 à 60°C a été choisie. Ainsi la masse des produits a été limitée à la gamme comprise entre 4500 et 8500 Da. Cette T_m implique une hybridation lors de la réaction de PEX à une T_h entre 48 et 56°C assurant ainsi une spécificité suffisante, en particulier lorsqu'il s'agit d'hybridation sur des produits PCR et non sur tout un génome.

En vue de la réaction d'extension d'amorce, les produits PCR ont été purifiés grâce au kit GenopureTM ds (Bruker Daltonics) selon les recommandations du fournisseur et élués dans 8 μ L d'eau ultra pure.

II.6.2. OPTIMISATION PEX

Dans un premier temps, les SNP d'intérêt ont été analysés individuellement (singlePEX) dans le but de valider la spécificité d'extension. Les PEX simultanées (multiPEX) n'ont été optimisées qu'après validation de la faisabilité technique de l'analyse des SNP et de leur informativité et utilité pour l'identification individuelle.

Au vu de l'indépendance des principaux facteurs de l'amplification, le milieu de réaction a été optimisé selon un plan expérimental factoriel complet à 3 niveaux.

II.6.2.1. Plan d'expérience

Les facteurs primordiaux dans une réaction PEX simple sont la concentration en magnésium, la température d'hybridation et la quantité d'ADN cible.

Les niveaux du facteur "concentration Mg²⁺" ont été définis selon les recommandations du fournisseur de la polymérase. Les valeurs de Ta testées ont été définies autour de la valeur médiane de la fourchette préconisée au vu des Tm des amorces (Tableau 2).

Tableau 2 : Facteurs et leurs niveaux évalués dans le plan d'expérience pour l'optimisation des réactions PEX.

Facteur	Niveau	-1	0	+1
[Mg ²⁺]		5 mM	7,5 mM	10 mM
Ta		49°C	51°C	53°C
Quantité de produit PCR purifié		2 µL	4 µL	6 µL

Le résultat est considéré positif si, pour toutes les extensions testées dans les conditions définies, les produits attendus sont présents et correspondent aux produits de séquençage, des produits aspécifiques sont inexistantes, l'excès d'amorce n'est pas majoritaire et les produits correspondant aux deux allèles sont équilibrés. Le résultat est considéré négatif si une de ces exigences n'est pas satisfaite.

II.6.2.2. Les autres paramètres de réaction PEX

Le volume réactionnel est de 10 µL et contient 1 unité de TerminiPol® polymérase (Solis BioDyne, Estonie), du tampon de réaction sans magnésium (1X), 1 µM d'amorce, 200 µM de chaque dNTP et ddNTP nécessaire à l'extension et du MgCl₂ à la concentration déterminée par le plan d'expérience. Les réactions de PEX ont été réalisées sur un appareil GeneAmp PCR System 9600 (Perkin Elmer) avec un programme comprenant 1 minute de dénaturation initiale suivie de 35 cycles de 10 secondes à 95°C, 30 secondes à Ta (cf. Tableau 2), 30 secondes à 72°C et finalisé par une extension de 2 minutes à 72°C.

II.6.3. PURIFICATION DES PRODUITS PEX

Les produits d'extension ont été purifiés grâce au kit Genopure™ oligo (Bruker Daltonics) selon les recommandations du fournisseur et élués dans 5 µL d'eau ultra pure.

II.7. *Les réactions simultanées (PCR multiplex et multiPEX)*

L'optimisation des réactions simultanées a été faite lorsque tous les marqueurs ont été validés. Les deux étapes de réactions multiples sont liées puisque les produits de chaque PCR multiplexée doivent être analysés dans des réactions PEX multiplexées. Cependant, tous les produits d'une PCR ne doivent pas impérativement être analysés en une seule PEX, ce qui permet un moindre degré de multiplexage dans la discrimination allélique et l'utilisation d'amorces de taille similaire. Ce dernier point permet des températures de fusion voisines et une faible dispersion des spectres de masse sans perte de résolution.

Tout d'abord les SNP ont été regroupés selon les dNTP nécessaires à leur analyse c'est à dire selon la composition du milieu réactionnel PEX (Tableau 3).

Ensuite, les SNP de chaque milieu ont été triés selon la masse moléculaire des amorces servant à leur discrimination allélique et des produits d'extension et regroupés par 2 à 4 de manière à ce que les entités attendues (amorces et oligonucléotides SNP et allèle spécifiques) différent en masse de au moins 60 Da. Ceci a donné lieu à des "groupes PEX". Enfin, plusieurs "groupes PEX" ont été regroupés en "groupes PCR".

Tableau 3 : Composition en nucléotides des différents milieux PEX

Milieu	Désoxyribonucléotide	Di-désoxyribonucléotides
A	dATP	ddTTP, ddGTP, ddCTP
T	dTTP	ddATP, ddGTP, ddCTP
G	dGTP	ddTTP, ddATP, ddCTP
C	dCTP	ddGTP, ddTTP, ddATP
AT	dATP, dTTP	ddGTP, ddCTP
GC	dGTP, dCTP	ddTTP, ddATP

Ensuite, les SNP de chaque milieu ont été triés selon la masse moléculaire des amorces et des produits d'extension puis regroupés par 2 à 4 de manière à ce que les entités diffèrent en masse d'au moins 60 Da. Ceci a donné lieu à des "groupes PEX". Enfin, plusieurs "groupes PEX" ont été regroupés en "groupes PCR".

Dans un premier temps, les réactions simultanées ont été optimisées selon les mêmes facteurs que les réactions simples. Ensuite, l'équilibre des quantités de produits générées a été ajusté en variant la concentration des amorces de manière empirique. Cependant, lorsque des fragments contenant des SNP d'intérêt ne pouvaient pas être produits par cette optimisation simple, de nouvelles amorces ont dû être dessinées et synthétisées jusqu'à obtention de la compatibilité avec des réactions simultanées.

II.8. Spectrométrie de masse MALDI-TOF

II.8.1. PREPARATION DE LA MATRICE

L'analyse d'oligonucléotides en spectrométrie de masse MALDI-TOF nécessite leur co-cristallisation avec une matrice afin d'être protégés lors de leur ionisation.

Dans cette étude, la matrice utilisée comporte de l'acide 3-hydroxypicolinique (3-HPA) [Wu et al., 1994] et de sels d'ammonium afin d'améliorer le signal (diammonium hydrogènocitrate, DAHC) [Pieles et al., 1993].

La cristallisation de la matrice est un mécanisme mal contrôlé. En effet, le 3-HPA cristallise le plus souvent en anneau hétérogène rendant une recherche de *sweet spots* de cristaux incontournable. Afin d'optimiser la cristallisation, nous avons testé plusieurs facteurs (Tableau 4) selon un plan expérimental factoriel à 2 niveaux.

Tableau 4 : Facteurs et leurs niveaux évalués dans le plan d'expérience pour l'optimisation de la cristallisation des dépôts.

Facteur \ Niveau	-1	+1
[3-HPA]	5 g / L	10 g / L
[DAHC]	2 g / L	4 g / L
Quantité de matrice déposée	0,5 µL	1 µL
Quantité d'échantillon déposée	0,5 µL	0,75 µL

II.8.2. PREPARATION DU DEPOT

La méthode de dépôt sélectionnée est la méthode de la goutte séchée sur une cible Anchor-Chip™ (284-400 µm) (Bruker Daltonics). Le dépôt de l'échantillon se fait sur le dépôt de matrice cristallisée. Différents modes de séchage, 20 minutes à température ambiante, 5 minutes à l'étuve à 35°C et 30 secondes dans une cloche à vide, ont été testés afin d'observer leur influence sur la qualité et la rapidité de la cristallisation.

II.8.3. ANALYSE PAR SPECTROMETRIE DE MASSE MALDI-TOF

Les analyses ont été faites sur un instrument Ultraflex™ (Bruker Daltonics) au LSMBO en mode linéaire positif (Tableau 5). L'excès d'amorces non étendues a été utilisé comme standard interne à chaque dépôt pour la calibration de l'appareil.

Tableau 5 : Valeurs de paramètres de la méthode d'analyse appliquées à la détection des oligonucléotides par l'instrument Ultraflex™ (Bruker Daltonics).

Paramètres de l'instrument	Valeurs pour l'analyse d'oligonucléotides (3 kDa < M < 9 kDa)
Pression N2	~1400 mbar
Puissance Laser	65-75%
Nombre de tirs	100
Spectromètre	
Source ionique 1	20 kV
Source ionique 2	18,3 kV
Lentille	8 kV
Extraction d'ion pulsée	150 ns
Polarité	Positive
Détecteur	Linéaire
Filtrage d'ions de basse masse	2000 Da
Détection	
Gamme de masse	Basse : 3 kDa – 9 kDa
Gain électronique	3X
Lissage en temps réel	oui
Paramétrage	
Fréquence Laser	20 Hz
Atténuateur de laser	minimum 60 / gamme 30
Digitaliseur	127 cnt

II.8.4. DETERMINATION DE LA MASSE

Les valeurs de masse pour chaque pic détecté à un temps de vol donné ont été attribuées grâce au logiciel X-MASSTM (Bruker Daltonics) en mode d'annotation centroïde. Cette fonction calcule la masse de l'entité détectée selon la position des 10% de la partie supérieure du pic qu'elle engendre.

Afin de déterminer les fenêtres de lecture des temps de vol de chaque produit, c'est-à-dire les valeurs d'intervalles de masses dans lesquels les pics correspondant à un produit donné doivent être attendus, nous avons comparé d'une part les valeurs de masses attendues aux valeurs de masses observées et nous avons d'autre part étudié la déviation standard de la masse observée pour chaque produit. Ces évaluations ont été faites par comparaison de 96 spectres pour chaque produit attendu.

II.9. Analyse des SNP du chromosome Y

L'analyse des SNP TAT, M242 et RPS4Y a été réalisée de manière simultanée. Les fragments les contenant ont été amplifiés dans un volume réactionnel de 30 μ L à l'aide de 1 unité de Taq polymérase (MBI fermentas, Lituanie) en présence du tampon de réaction sans magnésium (1X), 0,32 μ M de chaque amorce TAT et RPS4Y, 0,16 μ M de chaque amorce M242 (Tableau 6), 200 μ M de chaque dNTP et de 3 mM MgCl₂. Les réactions de PCR ont été réalisées sur un appareil GeneAmp PCR System 9600 (Perkin Elmer) avec un programme composé de 5 minutes de dénaturation initiale à 96°C suivies de 30 cycles de 15 secondes à 96°C, 15 secondes à 60°C, 30 secondes à 72°C et finalisé par une élongation de 2 minutes à 72°C.

Tableau 6 : Séquence des amorces utilisées pour l'amplification simultanée des fragments contenant les 3 SNP-Y sélectionnés.

Marqueur	Amorce	Séquence amorce (5'→3')	Taille du fragment amplifié
M242	Sens	GCT TTC TAC GGC ATA GAA AGT TTG TGC	143 pb
	Antisens	CCT AGA ACA ACT CTG AAG CGG TGG	
TAT	Sens	GAC TCT GAG TGT AGA CTT GTG A	112 pb
	Antisens	GAA GGT GCC GTA AAA GTG TGA A	
RPS4Y	Sens	CTA TTG CAG GGC AAT AAA CCT TGG	70 pb
	Antisens	CTT CAG CAA CAG TAA GTC GAA TGC	
		

Les séquences des amorces et produits PEX possibles lors de l'analyse de ces trois marqueurs sont résumés dans le tableau 7. Le volume réactionnel de PEX de 10 μ L contient 1 unité de TermiPol® polymérase (Solis BioDyne, Estonie) en présence du tampon de réaction sans magnésium (1X), 1 μ M des amorces TAT-S et RPS4Y-S et 0,5 μ M d'amorce M242-S, 200 μ M de chaque dNTP et ddNTP nécessaire à l'extension et 10 mM $MgCl_2$. Les réactions de PEX ont été réalisées sur un appareil GeneAmp PCR System 9600 (Perkin Elmer) avec un programme composé de 1 minute de dénaturation initiale à 96°C suivie de 35 cycles de 10 secondes à 95°C, 30 secondes à 54°C, 30 secondes à 72°C et finalisé par une élongation de 2 minutes à 72°C.

Tableau 7 : Séquence et masse des produits attendus lors de l'analyse des marqueurs TAT, M242 et RPS4Y. L'orientation de l'amorce (s pour sens et as pour antisens) est aussi indiqué.

Marqueur	Produit	Séquence	Masse
RPS4Y-S	Amorce (s)	CAA TAA ACC TTG GAT TTC	5459 Da
	Allèle T	CAA TAA ACC TTG GAT TTC ddT	5763 Da
	Allèle C	CAA TAA ACC TTG GAT TTC dCdCddT	6340 Da
M242-S	Amorce (s)	CAA AAA GGT GAC CAA GGT GCT	6489 Da
	Allèle T	CAA AAA GGT GAC CAA GGT GCT dT	6777 Da
	Allèle C	CAA AAA GGT GAC CAA GGT GCT dCddT	7066 Da
TAT-S	Amorce (as)	CTC TGA AAT ATT AAA TTA AAA CAA C	7626 Da
	Allèle T	CTC TGA AAT ATT AAA TTA AAA CAA Cdda	7939 Da
	Allèle C	CTC TGA AAT ATT AAA TTA AAA CAA CdGddT	8259 Da

II.10. Analyse des SNP du chromosome X

Les fragments contenant les SNP du chromosome X ont été amplifiés en réactions individuelles dans un volume réactionnel de 30 μ L à l'aide de 1 unité de Taq polymérase de MBI fermentas (Lithuanie) en présence du tampon de réaction sans magnésium (1X), 0,32 μ M de chaque amorce (Tableau 8), 200 μ M de chaque dNTP et de 2 mM $MgCl_2$. Les réactions de PCR ont été réalisées sur un appareil Perkin Elmer GeneAmp PCR System 9600 avec un programme composé de 5 minutes de dénaturation initiale à 96°C suivies de 30 cycles de 15 secondes à 96°C, 15 secondes à 58°C, 30 secondes à 72°C et finalisé par une élongation de 2 minutes à 72°C.

Les fragments ont été analysés par séquençage grâce au kit BigDye® Terminator v 1.1 Cycle Sequencing Kit (Applied Biosystems) sur un appareil Perkin Elmer GeneAmp PCR System 9600 selon les recommandations du fournisseur (paragraphe II.5.). Les deux brins des amplicons ont été séquencés.

Tableau 8 : Les SNP du chromosome X étudiés, leur fréquences alléliques dans la population caucasienne et les séquences des amorces (de 5' en 3') permettant l'amplification et le séquençage de fragments les contenant.

dbSNP N°	séquence de l'amorce sens séquence de l'amorce antisens	allèles
rs618	TCC TAA GAG ATG TTT TGT TC CTT TCC TGT TAC TTC CCG	G/A
rs16282	CTG GGC TCA ATG AAT GGC GTG AAG AGA GAA ATG AAA TG	C/T
rs616	TCC ATC AGA ACT ACA GCG TTC TCA CTT ATC AAC AAA CC	C/T
rs17379	TTT TAC ATT GAA GAT TTA CAC A TGC TAA TGA GAC AAA TGC C	A/G
rs578	CCA TCC ATC ACT TCC TTC TGC CCT TAG TTC CCT CAG ATG AC	C/T
	

II.11. Analyses statistiques

II.11.1. PARAMETRES SPECIFIQUES DES LOCI

- Le degré d'hétérozygotie

Le degré d'hétérozygotie représente la fréquence observée du génotype hétérozygote.

- Probabilité d'identité et pouvoir de discrimination

La probabilité d'identité (pM) représente la probabilité pour que deux individus pris au hasard dans une population présentent des génotypes identiques pour le locus considéré. Elle est calculée par la formule suivante :

$$pM = \sum_{i=1}^n pGi^2$$

où pGi correspond à la fréquence du i^{ème} génotype, i couvrant les différents génotypes considérés.

Le pouvoir discriminant (Pd) correspond à la probabilité inverse de pM : $Pd = 1 - pM$. Plus un marqueur génétique possède de variants alléliques, plus le pouvoir de discrimination est élevé. Lors de l'étude de plusieurs loci, il est possible de calculer un pouvoir de discrimination combiné à partir des pM des n marqueurs selon la formule suivante, où pM_i correspond à la probabilité d'identité du i^{ème} marqueur et n le nombre de marqueurs considérés.

$$P_{d-combiné} = 1 - \prod_{i=1}^n pM_i$$

Ce pouvoir discriminant traduit la confiance que l'on peut avoir dans l'utilisation d'un ensemble de marqueurs pour distinguer deux individus dans une population. Il permet d'évaluer la puissance de l'analyse que l'on pratique mais ne donne pas de renseignement sur l'indice biologique lui-même. Ce renseignement peut être obtenu à partir de la fréquence du profil qui correspond au produit des fréquences des génotypes observés sur les loci considérés. Correspondant au risque de correspondance fortuite elle permet d'évaluer la valeur probante de l'indice. Plus ce risque est faible, plus la valeur probante sera grande.

II.11.2. UTILISATION DU LOGICIEL GENEPOP

Les divergences de l'équilibre de Hardy-Weinberg et les déséquilibres de liaison des différents loci situés sur le même chromosome ont été estimés grâce au logiciel Genepop [Raymond et Rousset, 1995], basé sur la méthode de la chaîne de Markov [Guo et Thompson, 1992]. Les paramètres de la chaîne de Markov ont été fixés à 1000 mémorisations, 1000 batch et 5000 itérations par batch pour tous les tests, afin d'obtenir des déviations standard (*Standard Error*, S.E.) des estimations inférieures à 0,001.

Une population est un groupe de reproduction tel que la majorité des individus choisit son conjoint à l'intérieur du groupe. La loi de Hardy Weinberg pose de nombreuses conditions : la population doit pouvoir être assimilée à une population panmictique et pangamique infinie, exempte de mutation, migration ou sélection. Dans ce cas, si à la génération n, les allèles a et b ont pour fréquence p_a et p_b, les fréquences des trois génotypes aa, ab et bb à la génération n+1 sont p_a², 2p_ap_b et p_b² respectivement. Si les fréquences suivent cette loi, il est admis qu'elles sont en équilibre de Hardy-Weinberg pour le locus étudié.

On appelle déséquilibre de liaison, le paramètre qui mesure l'écart à l'association aléatoire de deux allèles situés à des locus différents dans les gamètes. Il mesure le degré de dépendance statistique entre deux allèles à deux loci différents.

II.11.3. PROBABILITES DE PATERNITE

Lors des calculs de paternité les rapports de vraisemblance (*likelihood ratio*, LR) connaissant le duo mère/enfant opposent les probabilités de deux hypothèses. D'une part l'hypothèse H_0 où le père présumé est le père biologique sachant le résultat de l'analyse des marqueurs génétiques E ($p(H_0|E)$) est évaluée. D'autre part est calculée la probabilité de l'hypothèse H_1 où un autre homme sans lien de parenté avec le père présumé est le père biologique de l'enfant sachant le résultat E ($p(H_1|E)$). Les LR ont été calculés grâce au génotypage des 25 trios (père présumé, mère et enfant) où la paternité a été confirmée ou exclue préalablement par analyse de microsatellites. Les calculs des LR pour chaque locus ont été faits après modification des formules utilisées pour l'interprétation des analyses STR (Tableau 9). Les LR cumulés pour chaque profil ont été calculés selon la formule indiquée ci-dessous où n est le nombre de marqueurs dont le génotype est inclus dans le profil et LR_i est le LR pour le génotype du $i^{\text{ème}}$ marqueur.

$$LR_{\text{cumulé}} = \frac{\prod_{i=1}^n LR_i}{1 + \prod_{i=1}^n LR_i}$$

Enfin la probabilité de paternité $p(\text{Pat})$ est définie par le rapport $LR_{\text{cumulé}} / (1+LR_{\text{cumulé}})$ en tenant compte de la probabilité *a priori* de 0,5.

Tableau 9 : Rapports de vraisemblance correspondant aux possibilités de combinaisons de génotypes de la mère, de l'enfant et du père présumé dans les tests de paternité. Ces rapports présupposent l'absence de lien de parenté entre le père présumé et le père biologique. a et b correspondent aux allèles de chaque marqueur et $f(a)$ à la fréquence de l'allèle transmis par le père.

Génotype de la mère	Génotype de l'enfant	Génotype du père présumé	LR
a/a	a/a	a/a	1/f(a)
a/a	a/a	a/b	0,5/f(a)
b/b	a/b	a/a	1/f(a)
b/b	a/b	a/b	0,5/f(a)
a/b	a/a	a/a	1/f(a)
a/b	a/a	a/b	0,5/f(a)
a/b	a/b	a/a	1
a/b	a/b	a/b	1

III.Résultats

III.1. *Les marqueurs autosomaux*

III.1.1. SELECTION THEORIQUE

Lors de l'étape de la sélection selon les critères de localisation, fonction et fréquences alléliques, 162 SNP ont été retenus comme marqueurs potentiels pour l'identification génétique. Les loci étaient caractérisés par une localisation non associée à de l'ARN messager et par des fréquences alléliques comprises entre 0,2 et 0,8 dans trois populations différentes. Cependant, cette dernière donnée, obtenue le plus souvent sur la banque de données de Applied Biosystems, restait à vérifier puisque les évaluations de fréquences ne portaient que sur un faible nombre d'individus de chaque population (caucasienne, afro-américaine, asiatique).

Le tableau résumant les SNP considérés, leurs localisations [Furey et Haussler, 2003], leurs allèles observés, leurs fréquences moyennes actuellement disponibles et leurs taux d'hétérozygotie moyens, se trouve en Annexe 2.

III.1.2. TRI "EXPERIMENTAL"

Cette étape de tri est basée sur la faisabilité de l'amplification par PCR dans les conditions souhaitées, la compatibilité avec une réaction d'amplification simultanée de plusieurs fragments et la faisabilité de la réaction de discrimination allélique.

III.1.2.1. La PCR simple

Selon le plan d'expérience détaillé dans le paragraphe II.4.2.1., la température d'hybridation (Th) et la concentration en magnésium, permettant l'amplification spécifique de tous les fragments, ont été déterminées à 51°C et 2 mM. Une fois les conditions permettant l'amplification spécifique de la majorité des fragments définies, certains marqueurs ont été sélectionnés selon leur compatibilité avec celles-ci afin de simplifier les expériences.

Les séquences des amorces dessinées pour l'amplification des fragments contenant les 50 SNP suite à toutes les étapes de sélection (expérimentales et statistiques) ainsi que la taille des amplicons sont résumés en Annexe 3.

III.1.2.2. Le séquençage

La méthode de discrimination allélique retenue pour cette étude implique l'utilisation d'amorces s'hybridant de manière adjacente au SNP d'intérêt. Ainsi, la première étape de tri est basée sur la vérification d'absence d'autres polymorphismes parmi les 25 bases de la séquence de part et d'autre du locus d'intérêt selon les données disponibles dans les bases de données ou selon les résultats obtenus par séquençage. Lorsque l'information sur la présence de variations dans la région d'hybridation de l'amorce était disponible dans les bases de données, l'exclusion du marqueur a été faite sans vérification expérimentale.

La majorité des séquences obtenues à partir de banques de données a été confirmée expérimentalement. De même, cette vérification a permis de valider la spécificité des réactions PCR. Cependant, la présence de certains polymorphismes non répertoriés dans les bases de données lors des développements initiaux a été révélée. Par exemple, le séquençage du fragment contenant le SNP rs1321158 (alias SP85) sur un faible nombre d'ADN (échantillons de validation) a permis de détecter des polymorphismes à 12 pb en amont et à 15 pb en aval du SNP d'intérêt (Figure 20).

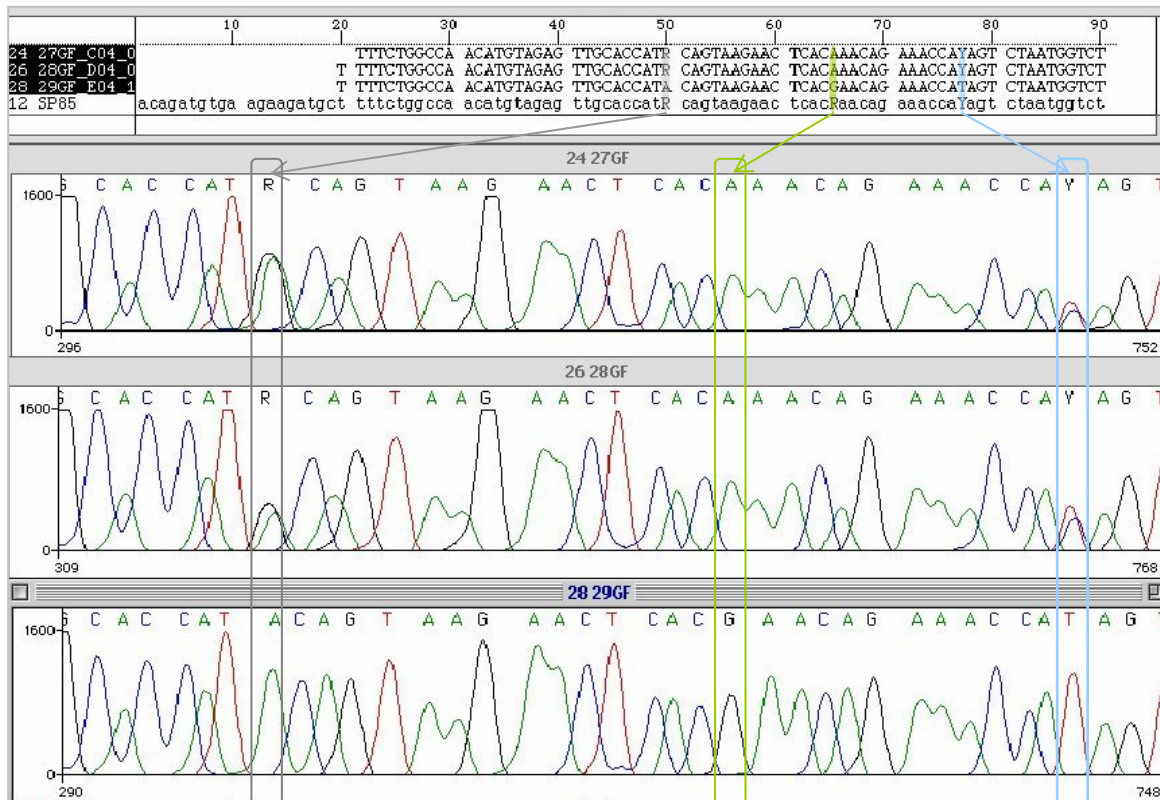


Figure 20 : Alignement de séquences du fragment contenant le marqueur rs1321158 (alias SP85, surligné en vert) sur 3 échantillons révélant la présence de polymorphismes (surlignés en gris et bleu) dans les régions bordant le SNP d'intérêt.

III.1.2.3. La PEX simple

La température d'hybridation et la concentration en magnésium permettant l'extension spécifique de la majorité des amorces ont été déterminées à 51°C et 10 mM. La quantité de produit PCR contenue dans 2 µL d'éluât après purification à l'aide du kit Genopure™ ds (Bruker Daltonics) a été retenue pour la réalisation des PEX. Ce volume permet de compenser de légères fluctuations d'efficacité de PCR dues à l'ADN matrice d'origine. Cependant, lorsque l'efficacité de la réaction PCR est sensiblement faible, la quantité d'amplicon cible a été augmentée soit en diminuant le volume d'éluât (de 8 à 4,5 µL) ou en augmentant le volume d'éluât introduit dans le milieu PEX. Une fois les conditions de réaction permettant une discrimination allélique spécifique de la majorité des SNP définies, certains marqueurs ont été sélectionnés selon leur compatibilité avec celles-ci afin de simplifier les expériences. Les séquences des amorces dessinées pour la discrimination allélique des SNP, leurs masses molaires et températures de fusion sont résumées en Annexe 4.

III.1.2.4. Analyse en spectrométrie de masse MALDI-TOF

III.1.2.4.1. Préparation de la matrice et du dépôt

Parmi les différentes méthodes de préparation, le dépôt de 0,5 µL de matrice à 20 g/L de 3-HPA et 2 g/L de DAHC co-cristallisé avec 0,75 µL d'éluât après purification à l'aide du kit Genopure™ oligo (Bruker Daltonics) par séchage pendant 5 à 10 minutes à l'étuve à 35°C maximum a été retenu. La méthode de la goutte séchée, simple à mettre en œuvre, permet la formation de cristaux relativement grossiers et hétérogènes identifiables lors de l'observation du dépôt grâce à la caméra de l'appareil (Figure 21 - gauche).

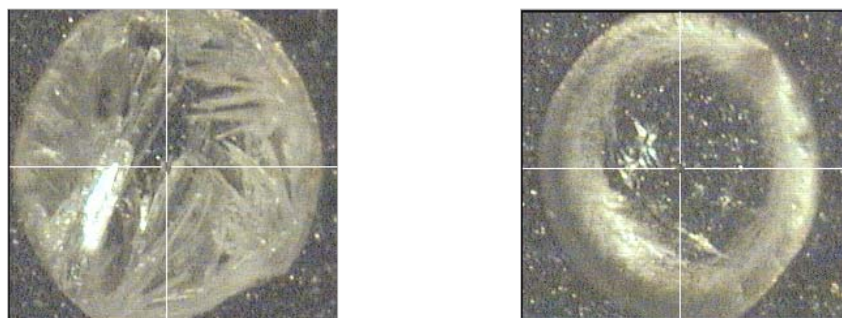


Figure 21 : Aspect du dépôt de la matrice et de l'échantillon par la méthode de la goutte séchée après séchage à l'étuve (à gauche) et dans une cloche à vide (à droite).

Le séchage sous vide favorise la cristallisation du dépôt en forme de bague laissant un espace vide au centre du dépôt sans formation de cristaux visibles (Figure 21 - droite).

Cependant, malgré l'optimisation de la composition de la matrice et du mode de préparation du dépôt, la cristallisation semble une étape difficilement contrôlable. En dupliquant systématiquement les dépôts sans changer de matériel (de cône) il est apparu que deux dépôts identiques préparés dans des conditions identiques peuvent cristalliser de manière différente (Figure 22).

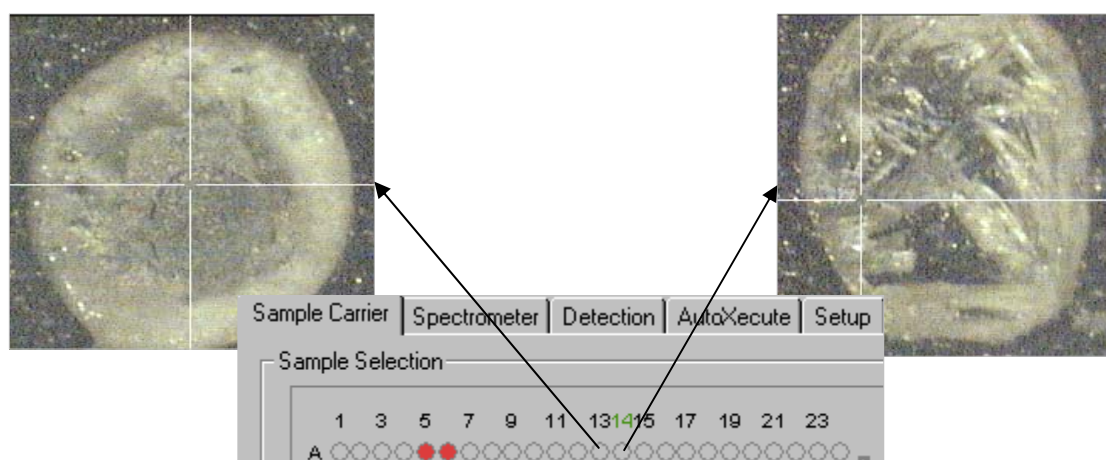


Figure 22 : Aspects de deux dépôts d'un même échantillon réalisés consécutivement dans les conditions identiques. Le dépôt A13 n'a formé aucun cristal visible ce qui rend la recherche d'analytes difficile alors que le dépôt A14 a formé des cristaux de taille importante dont la localisation est aisée.

III.1.2.4.2. Analyse des produits d'extension d'amorce

Une fois les réactions optimisées aucun produit secondaire non spécifique n'a été observé dans la fenêtre de mesure (3-9 kDa). Fréquemment, l'amorce est présente en excès dans le milieu PEX ce qui permet sa détection lors de l'analyse en spectrométrie de masse et son utilisation comme standard interne. Ainsi, un génotype homozygote est généralement caractérisé par 2 pics et un hétérozygote par 3 pics sur les spectres de masse (Figure 23). Des résultats similaires ont été obtenus pour tous les marqueurs sélectionnés. Seule la présence d'excès d'amorce diffère selon l'efficacité de réaction. En réaction simple, la température d'hybridation peut représenter un paramètre pour moduler l'efficacité et se traduit par l'augmentation ou la diminution de la quantité d'amorce résiduelle. Cependant, cette stratégie

ne peut pas toujours être appliquée dans le cas de réactions multiplexées puisqu'elle peut permettre l'amélioration d'une réaction au détriment des autres.

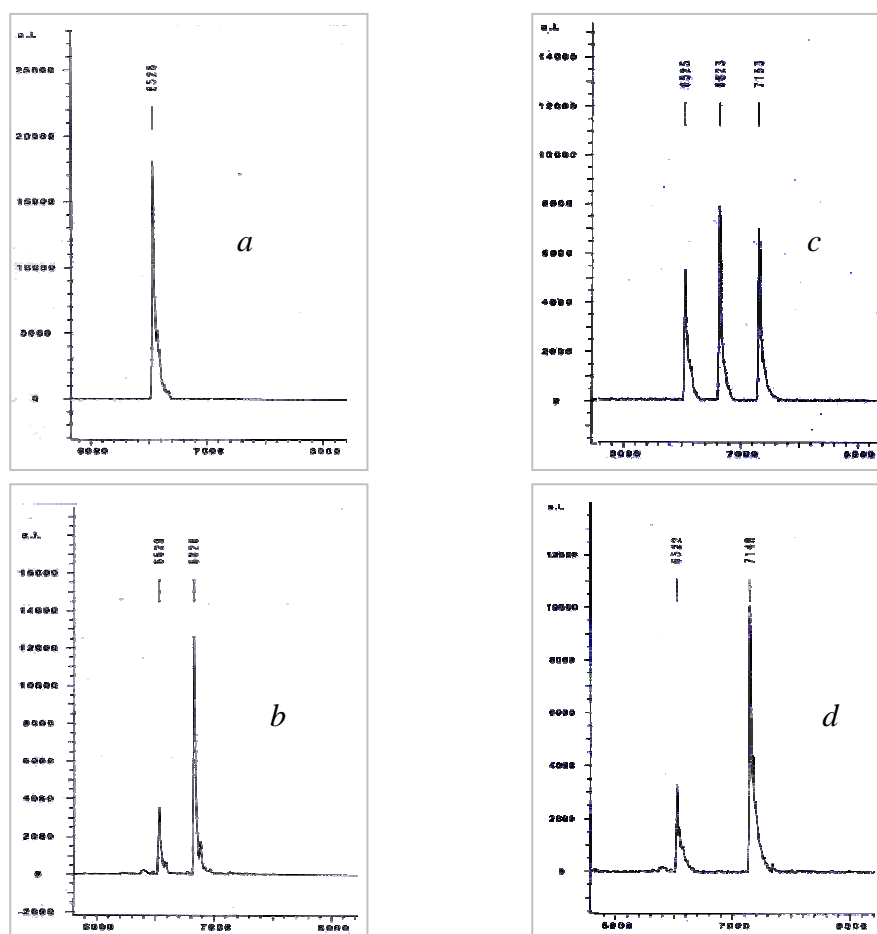


Figure 23 : Exemples de spectres obtenus lors de l'analyse du polymorphisme rs820129 (alias AB2P). a - témoin négatif ; seule l'amorce est détectée, b – échantillon lab 6 ; génotype homozygote A/A, c - échantillon lab 4 ; génotype hétérozygote A/G, d - échantillon lab 21 ; génotype homozygote G/G. Echelle de l'abscisse : m/z.

III.1.2.4.3. Détermination de la masse

Lors de cette évaluation, chaque produit a été représenté dans 96 spectres. Les résultats obtenus lors de l'analyse des amorces étendues ou non sont similaires. Les valeurs de masse ont été déterminées par la fonction centroïde du logiciel X-MASSTM (Bruker Daltonics).

Les déviations numériques, entre la valeur attendue et la moyenne des valeurs observées, et les valeurs de masse attendue présentent une corrélation positive et exponentielle ($r^2 = 0,9537$) (figure 24). Ces déviations sont comprises entre 1,6‰ et 33‰ pour des valeurs de masse allant de 4700 Da à 7600 Da.

Les masses observées pour chaque produit concernant chaque marqueur présentent une distribution selon la loi normale, ainsi la précision de la détermination des masses peut être évaluée par l'analyse de leur déviation standard (Figure 24). Pour tous les produits celle-ci est comprise entre 2,03 Da et 3,85 Da et elle présente une corrélation positive et faiblement linéaire par rapport à la masse ($r^2= 0,5626$).

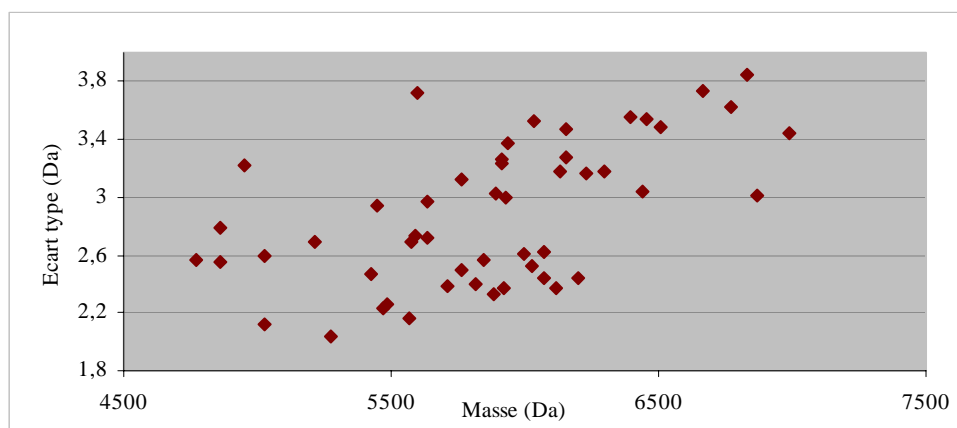


Figure 24 : Déviation standard de la distribution des masses observées en fonction de la masse observée. Afin d'alléger le graphique, seules les valeurs correspondant aux amorces non-étendues sont représentées car elles servent de référence dans chaque spectre. Des valeurs équivalentes ont été obtenues pour les produits de l'extension d'amorce.

III.1.2.5. Les réactions "multiplexes"

Les fragments contenant les 50 SNP et la différence nucléotidique dans le gène de l'amélogénine ont pu être amplifiés en 5 réactions PCR (tableau 10). Les génotypes ont été déterminés en 16 réactions PEX (tableaux 11a à 11g) en fonction des nucléotides nécessaires à la discrimination allélique.

Tableau 10 : Les groupes d'amplification des fragments contenant les SNP sélectionnés.

Groupe PCR	Groupes PEX correspondants	Nombre de fragments amplifiés
1	A - 1, A - 2, A - 3	11
2	T - 1, T - 2, T - 3	9
3	T - 4, T - 5, T - 6	9
4	G - 1, G - 2, G - 3, C - 1	11
5	C - 2, A/T, G/C	11

L'amplification de fragments de longueur relativement homogène (minimisant l'effet d'amplification préférentielle de fragments courts) rend difficile la distinction des différents fragments sur gel. De ce fait, la présence de certains produits n'a pu être déterminée qu'après analyse en spectrométrie de masse.

Tableau 11a : Réactions PEX en milieu C (dCTP, ddTTP, ddGTP, ddATP). Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm (Da)
PEX C - 1	SP39 - amorce	GTGCCAGACCGCTTCTTG	
	SP39 - A	GTGCCAGACCGCTTCTTGddA	297,21
	WI2032 - amorce	CAGAAGCAATTACAGGAAG	115,11
	SP39 - C	GTGCCAGACCGCTTCTTGdCddT	165,06
	WI2032 - C	CAGAAGCAATTACAGGAAGddG	148,15
	WI2032 - G	CAGAAGCAATTACAGGAAGdCddT	264,17
PEX C - 2	AB2Q - amorce	GTCAGCGGAGCCTCTC	
	AB2Q - C	GTCAGCGGAGCCTCTCddG	313,21
	AB2Q - G	GTCAGCGGAGCCTCTCdCddG	289,19
	SP80 - amorce	GCTGTTCAAATTCCTCTCTC	533,36
	SP80 - A	GCTGTTCAAATTCCTCTCTCddT	288,20
	SP21 - amorce	ATTGTCAGTCTCAAAGCACA	116,10
	SP80 - G	GCTGTTCAAATTCCTCTCTCdCddA	182,10
	SP21 - T	ATTGTCAGTCTCAAAGCACAddT	106,10
	WI1349 - amorce	CAGAAAATATGATTAGAAAAAG	146,13
	SP21 - C	ATTGTCAGTCTCAAAGCACAdCddT	143,06
	WI1349 - A	CAGAAAATATGATTAGAAAAAGddA	154,15
	WI1349 - C	CAGAAAATATGATTAGAAAAAGdCddG	305,18

Tableau 11b : Réactions PEX en milieu A/T (dATP, dTTP, ddGTP, ddCTP). Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX A/T	SP28 - amorce	CAAACCAGGGAAGGGAGG	131,08
	AB2H - amorce	TCTATTCAGCAGGCACTTA	81,04
	SP92 - amorce	CAGGTAGGGTCACAGTTTC	101,09
	SP28 - G	CAAACCAGGGAAGGGAGGddG	91,06
	AB2H - C	TCTATTCAGCAGGCACTTAddC	81,04
	SP92 - G	CAGGTAGGGTCACAGTTTCddC	141,12
	SP28 - A	CAAACCAGGGAAGGGAGGdAddG	122,07
	AB2H - T	TCTATTCAGCAGGCACTTAdTddG	81,04
	SP92 - A	CAGGTAGGGTCACAGTTTCdTddG	131,08

Tableau 11c : Réactions PEX en milieu A (dATP, ddTTP, ddGTP, ddCTP). Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX A - 1	AB2I - amorce	GAGCTGACTTGCCGCC	
	AB2I - C	GAGCTGACTTGCCGCCddG	313,21
	AB2I - T	GAGCTGACTTGCCGCCdAddG	313,21
	Or62 - amorce	GTGTAGTGGGAGCACTGG	151,09
	Or29 - amorce	ACACTGGATAACCAAGGAC	179,16
	Or62 - C	GTGTAGTGGGAGCACTGGddG	134,05
	AB1H - amorce	TGAACCATTTCCTCCTGGAC	79,06
	Or29 - C	ACACTGGATAACCAAGGACddG	100,10
	Or62 - T	GTGTAGTGGGAGCACTGGdAddT	109,04
	AB1H - C	TGAACCATTTCCTCCTGGACddG	104,07
	Or29 - T	ACACTGGATAACCAAGGACdAddT	75,09
	AB1H - T	TGAACCATTTCCTCCTGGACdAddC	198,10
PEX A - 2	WI2054 - amorce	GGCTGGTGGAGGCAGA	
	WI2054 - C	GGCTGGTGGAGGCAGAddG	313,21
	SP26 - amorce	GCAGGGGGAGTCCTTAAC	224,15
	WI2054 - T	GGCTGGTGGAGGCAGAdAddG	89,06
	SP26 - C	GCAGGGGGAGTCCTTAACddG	224,15
	SP26 - T	GCAGGGGGAGTCCTTAACdAddT	288,20
	WI1305 - amorce	GTATACTTATTACTTCTCCCT	131,08
	WI1305 - C	GTATACTTATTACTTCTCCCTddG	313,21
	WI1305 - T	GTATACTTATTACTTCTCCCTdAddC	273,19
PEX A - 3	SP67 - amorce	CTTCTGTCCACCTTTCC	
	SP67 - A	CTTCTGTCCACCTTTCCddT	288,20
	SP67 - T	CTTCTGTCCACCTTTCCdAddT	313,21
	WI1121 - amorce	AGAGGAGAGTAAAAACAGC	303,24
	SP29 - amorce	CAACAACAGTGAAAAGGATC	224,15
	WI1121 - C	AGAGGAGAGTAAAAACAGCddG	89,06
	SP29 - G	CAACAACAGTGAAAAGGATCddG	224,15
	WI1121 - T	AGAGGAGAGTAAAAACAGCdAddT	64,05
	Or54 - amorce	CTGTTTTCTGTTATATTCTGCT	134,02
	SP29 - A	CAACAACAGTGAAAAGGATCdAddG	115,14
	Or54 - C	CTGTTTTCTGTTATATTCTGCTddG	198,07
	Or54 - T	CTGTTTTCTGTTATATTCTGCTdAddT	288,20

Tableau 11d : Réactions PEX en milieu T (dTTP, ddATP, ddGTP, ddCTP) 1^{ère} partie. Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX T - 1	AB1M - amorce	AACTAATGGGCACCCTTC	
	Or68 - amorce	CCGCTGTGAGACAAAGGG	130,08
	AB1M - G	AACTAATGGGCACCCTTCddC	143,11
	Or68 - C	CCGCTGTGAGACAAAGGGddC	130,08
	AB1M - A	AACTAATGGGCACCCTTCdTddG	214,15
	Or68 - T	CCGCTGTGAGACAAAGGGdTddG	130,08
PEX T - 2	WI1306 - amorce	AGGTAGCCCACAGAAGC	
	WI1306 - G	AGGTAGCCCACAGAAGCddC	273,19
	WI1306 - A	AGGTAGCCCACAGAAGCdTddA	328,22
	WI563 - amorce	TAGGTAGACATGATCAGGG	102,05
	Or33 - amorce	GCAGCATACTCATTCACAGA	153,11
	WI563 - G	TAGGTAGACATGATCAGGGddC	120,08
	Or33 - C	GCAGCATACTCATTCACAGAddC	153,11
	WI563 - A	TAGGTAGACATGATCAGGGdTddA	175,11
Or33 - T	GCAGCATACTCATTCACAGAdTddG	169,11	
PEX T - 3	Or09 - amorce	CAGGCAGGTGGTCAGGC	
	WI1803 - amorce	GTGATGCTCTACCCTACT	149,11
	Or09 - C	CAGGCAGGTGGTCAGGCddC	124,08
	WI1803 - G	GTGATGCTCTACCCTACTddC	149,11
	Or09 - T	CAGGCAGGTGGTCAGGCdTddA	179,11
	WI1803 - A	GTGATGCTCTACCCTACTdTddA	149,11
	AB2T - amorce	GAGAGAACAATTAACCTCAC	100,10
	SP02 - amorce	GTTGTTAATGAGGATGATGC	100,02
	AB2T - C	GAGAGAACAATTAACCTCACddG	213,19
	SP02 - G	GTTGTTAATGAGGATGATGCddC	60,00
	AB2T - A	GAGAGAACAATTAACCTCACdTddG	244,20
SP02 - A	GTTGTTAATGAGGATGATGCdTddG	100,02	
PEX T - 4	SP12 - amorce	AGGCTACTCTTCTTCTTAGG	
	AB2J - amorce	GAAATGACAACCTGGTGAAAT	124,13
	SP12 - C	AGGCTACTCTTCTTCTTAGGddC	149,06
	AB2J - A	GAAATGACAACCTGGTGAAATddA	148,15
	SP12 - T	AGGCTACTCTTCTTCTTAGGdTddA	180,07
	AB2J - T	GAAATGACAACCTGGTGAAATdTddC	100,10
	Or05 - amorce	AGTTCAAATATACTAAACTATTC	214,15
	Or05 - C	AGTTCAAATATACTAAACTATTCddC	273,19
	Or05 - T	AGTTCAAATATACTAAACTATTCdTddC	304,20

Tableau 11e : Réactions PEX en milieu T (dTTP, ddATP, ddGTP, ddCTP) 2^{ème} partie. Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX T - 5	WI1341 - amorce	AACAGATCCATCACAATCC	
	WI597 - amorce	GAGTCATGGAGGGTCATTA	198,10
	WI1341 - G	AACAGATCCATCACAATCCddC	75,09
	AB1C - amorce	GCATCAAAACAAAGGCTCAA	129,09
	WI597 - G	GAGTCATGGAGGGTCATTAddC	69,01
	WI1341 - A	AACAGATCCATCACAATCCdTddG	146,12
	AB1C - C	GCATCAAAACAAAGGCTCAAddG	98,08
	WI597 - A	GAGTCATGGAGGGTCATTAdTddC	60,00
	AB1C - A	GCATCAAAACAAAGGCTCAAdTddA	228,20
PEX T - 6	Or11 - amorce	TAGAACCAATCAAGGCTCC	
	SP56 - amorce	CTCTCCTATCCATTTTCCAC	148,07
	Or11 - C	TAGAACCAATCAAGGCTCCddC	125,12
	WI681 - amorce	CAAATGGTGCCATAGTATAA	110,07
	SP56 - T	CTCTCCTATCCATTTTCCACddA	62,02
	Or11 - T	TAGAACCAATCAAGGCTCCdTddA	156,13
	WI681 - G	CAAATGGTGCCATAGTATAAddC	55,04
	SP56 - A	CTCTCCTATCCATTTTCCACdTddC	69,01
	WI681 - A	CAAATGGTGCCATAGTATAAdTddA	259,21

Tableau 11f : Réactions PEX en milieu G/C (dGTP, dCTP, ddATP, ddTTP). Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX G/C	AmelSD - amorce	CTGCAGGGTGTGAGTC	
	AmelSD - C	CTGCAGGGTGTGAGTCddG	313,21
	AmelSD - T	CTGCAGGGTGTGAGTCdAddG	313,21
	SP78 - amorce	TGAGAATGTGGACCTGAGG	353,24
	AB1D - amorce	ATCTTAGCCTGTACCTTTTA	100,07
	SP78 - A	TGAGAATGTGGACCTGAGGddT	188,13
	AB1D - A	ATCTTAGCCTGTACCTTTTAddA	109,08
	SP87 - amorce	AGTAGGCTTTTCATCAATTAG	105,06
	SP78 - G	TGAGAATGTGGACCTGAGGdCddT	75,04
	AB1D - C	ATCTTAGCCTGTACCTTTTAdCddA	109,08
	SP87 - T	AGTAGGCTTTTCATCAATTAGddT	104,08
	SP87 - G	AGTAGGCTTTTCATCAATTAGdGddT	329,21

Tableau 11g : Réactions PEX en milieu G (dGTP, ddATP, ddTTP, ddCTP). Δm correspond à l'écart de masse du produit par rapport au produit précédent (monoprotoné).

"lot" PEX	produit	séquence du produit	Δm
PEX G - 1	SP19 - amorce	AAGAGACCTGTAGGAGCG	
	SP19 - T	AAGAGACCTGTAGGAGCGddA	297,21
	SP19 - C	AAGAGACCTGTAGGAGCGdGddT	320,20
	AB2P - amorce	ATAAAGAAAGAACCTGGAGAC	290,22
	AB2P - A	ATAAAGAAAGAACCTGGAGACddA	297,21
	AB2P - G	ATAAAGAAAGAACCTGGAGACdGddA	329,21
PEX G - 2	AB1T - amorce	CGCACCCCGCAGACCT	
	AB1T - T	CGCACCCCGCAGACCTddA	297,21
	AB1T - C	CGCACCCCGCAGACCTdGddT	320,20
	SP44 - amorce	CTCCGTGTGTCTGTGTGC	99,04
	SP44 - G	CTCCGTGTGTCTGTGTGCddC	273,19
	SP64 - amorce	AATTGATGTTGGGCTCTGG	128,10
	SP44 - C	CTCCGTGTGTCTGTGTGCdGddC	201,11
	SP64 - A	AATTGATGTTGGGCTCTGGddA	96,10
	SP64 - G	AATTGATGTTGGGCTCTGGdGddC	305,19
	WI1327 - amorce	CCCAAGAACTACATTTGGGGAG	276,20
	WI1327 - G	CCCAAGAACTACATTTGGGGAGddC	273,19
	WI1327 - C	CCCAAGAACTACATTTGGGGAGdGddC	329,21
PEX G - 3	WI1732 - amorce	ATGTGGCAGAGACTGAAT	
	WI1732 - T	ATGTGGCAGAGACTGAATddA	297,21
	WI1732 - C	ATGTGGCAGAGACTGAATdGddA	329,21
	WI1126 - amorce	CAAGAAATAAAAGACAAGTCC	235,19
	WI1126 - T	CAAGAAATAAAAGACAAGTCCddA	297,21
	Or36 - amorce	GTAGATAGAATTGAATAGAAGT	124,05
	WI1126 - C	CAAGAAATAAAAGACAAGTCCdGddA	205,16
	Or36 - T	GTAGATAGAATTGAATAGAAGTddA	92,05
	Or36 - C	GTAGATAGAATTGAATAGAAGTdGddA	329,21

La figure 26 représente des exemples de spectres obtenus lors de l'analyse des polymorphismes des groupes PEX G - 1 (figure 26a) et G - 2 (figure 26b).

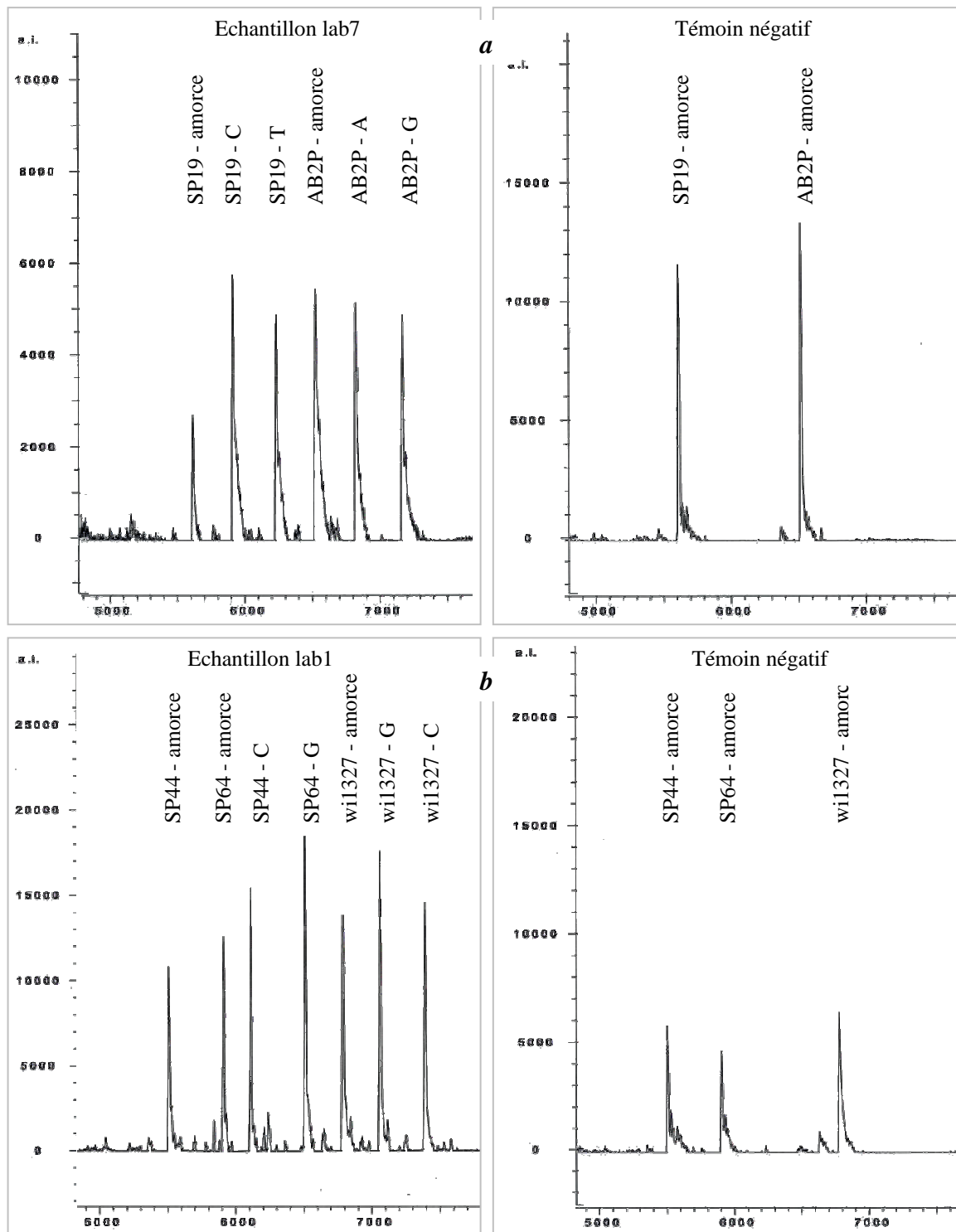


Figure 26 : Exemples de spectres de masses lors d'analyses simultanées de plusieurs SNP. Echelle de l'abscisse : m/z .

(a) Analyse des SNP du groupe PEX G-1 chez un individu hétérozygote pour les 2 marqueurs et celui du témoin négatif.

(b) Analyse des SNP du groupe PEX G-2 chez un individu homozygote C/C pour le marqueur SP44, homozygote G/G pour le SNP SP64 et hétérozygote (G/C) pour le locus wi1327 et celui du témoin négatif.

III.1.2.6. Les hauteurs de pics

La différence de hauteur des pics correspondant aux deux allèles d'un individu hétérozygote peut atteindre 40%. La même différence de hauteur de pics a été observée lors de l'analyse de deux dépôts consécutifs d'un même échantillon (Figure 27).

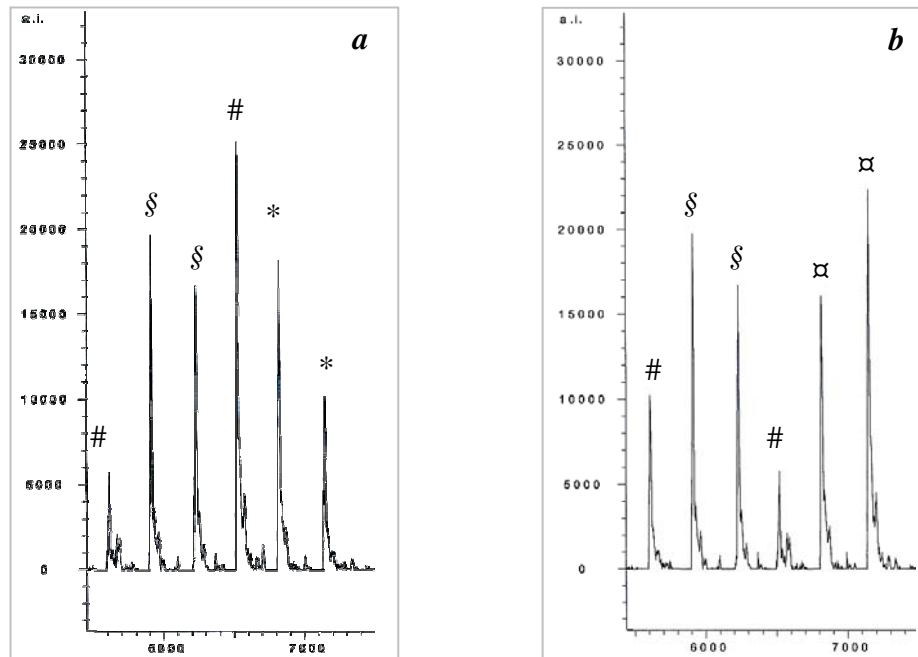


Figure 27 : Spectres obtenus à partir de deux dépôts consécutifs d'un même échantillon lors de l'analyse simultanée de deux marqueurs SNP. Dans le spectre a, la différence de hauteur des pics correspondant aux deux allèles du second marqueur atteint 40% (*). Les proportions sont inversées dans le spectre b (≡). De même, la proportion d'amorce non étendue varie grandement entre les 2 spectres (#). Par contre les hauteurs des pics correspondant au premier marqueurs sont constantes (§).

La balance ou équilibre d'hétérozygotie (*heterozygous balance*, Hb) se réfère au rapport de hauteur ou intensité des pics correspondant aux deux allèles d'un ADN hétérozygote. Nous avons observé des rapports très variables selon les marqueurs et selon les dépôts. Un rapport minimal de 1:3 entre les hauteurs des pics correspondant aux deux allèles d'un même marqueur a été utilisé comme seuil pour l'attribution du génotype hétérozygote.

III.1.3. RESULTATS DE L'EVALUATION STATISTIQUE

III.1.3.1. Les marqueurs

Les marqueurs ont été classés selon leurs positions génomiques et numérotées de 1 à 50 (Tableau 12). Les marqueurs 51 et 52 sont des exemples de loci dont l'évaluation statistique n'a pas permis la sélection en tant que marqueurs pour l'identification.

Tableau 12 : Correspondance des numéros attribués aux SNP (Nm_q) selon leur position génomique avec leurs numéros de référence dans la base données dbSNP (dbSNPrs) et la distance en paires de bases du marqueur le plus proche situé sur le même chromosome (NtDist).

Nm _q	dbSNPrs	labcode	NtDist (Mpb)	Nm _q	dbSNPrs	labcode	NtDist (Mpb)
1	rs991191	SP02	101,4	27	rs936250	SP44	18,7
2	rs2013526	Or05	101,4	28	rs997750	Or36	50,2
3	rs1020636	Or09	26,8	29	rs241	WI1349	50,2
4	rs1019264	Or11	26,8	30	rs243	WI2054	20,1
5	rs222	WI1121	38,7	31	rs544021	AB2Q	27,5
6	rs1551995	SP12	6,5	32	rs882937	AB1M	27,5
7	rs227	WI1341	6,5	33	rs1013582	SP56	21,2
8	rs225	WI1327	188,2	34	rs1793286	AB2J	17,3
9	rs763443	SP19	25,1	35	rs2123	WI563	27,9
10	rs279536	SP21	25,1	36	rs811	WI597	27,9
11	rs230	WI1803	57	37	rs1924609	AB2H	
12	rs724784	AB1D	15,6	38	rs757655	SP64	
13	rs1371370	SP26	15,6	39	rs767453	SP67	
14	rs889012	AB1C	10,9	40	rs727206	Or54	42,2
15	rs910170	AB1H	4,7	41	rs820129	AB2P	42,2
16	rs833069	SP28	4,7	42	rs276922	AB2T	6,2
17	rs713053	SP29	118,7	43	rs948438	SP78	6,2
18	rs997556	Or29	44,3	44	rs768058	SP80	26
19	rs2260	WI1126	44,3	45	rs1674139	AB2I	
20	rs234	WI1732	9,1	46	rs1003204	Or62	34,6
21	rs725	WI681	35,2	47	rs964224	SP87	15,3
22	rs235	WI1305	24,9	48	rs228047	AB1T	15,3
23	rs892503	Or33	24,9	49	rs738518	Or68	6,5
24	rs874746	SP39	37,1	50	rs737612	SP92	6,5
25	rs240	WI2032	9,9	51	rs868432	Or57	5,3
26	rs237	WI1306	9,9	52	rs754	WI1325	34,6

III.1.3.2. Les génotypes

Les calculs des paramètres de génétique des populations permettant d'estimer les déséquilibres de liaison, l'équilibre de Hardy Weinberg, le degré de polymorphisme et l'hétérozygotie, ont été effectués sur un échantillon constitué de 75 individus sans liens de parenté.

Le caractère binaire des marqueurs a été exploité, afin de faciliter la compréhension, la comparaison, l'exploitation et l'archivage des génotypes. En désignant par 0 le génotype homozygote de l'allèle de fréquence la plus élevée, par 1 le génotype homozygote de l'allèle de fréquence la moins élevée et par 2 le génotype hétérozygote, un code ternaire propre à chaque individu a été obtenu. Ce code ternaire (*Ternary Identification Code*, TIC) peut facilement être transformé en code décimal (*Decimal Code*, DC) en appliquant la formule ci-dessous avec n le nombre de marqueurs, i le numéro du marqueur et TIC_i le code ternaire correspondant au $i^{\text{ème}}$ marqueur.

$$DC = \sum_{i=1}^n TIC_i \cdot 3^{(i-1)}$$

Les TIC correspondants sont résumés dans le tableau 13.

Tableau 13 : Représentation sous forme de code ternaire des génotypes obtenus par analyse des SNP décrits dans le tableau 12. Les marqueurs sont représentés de gauche à droite dans l'ordre croissant des numéros qui leur ont été attribués. x indique que le génotype n'a pas été déterminé.

Individu	Génotype sous forme de code ternaire	Individu	Génotype sous forme de code ternaire
M1	12011122222022222000022210121220222020211022100220	G9	1202020211221201122021001002222122222002022201210200
G1	2202112222122221020222200010112002221012221002122222	P9	2222022012222201122221021202202120202000000022022220
P1	2200122202122001000112202012212001022221201022221111	M10	2122021202222120120202220011102221202202210020220220
M2	2200012220010201221022001010202210022102221022210120	G10	2122002221122120100001122211200221222102212012200100
G2	2022012212210202121222001212202112002201212221222100	P10	2222000112202200102001121211220011012201222221200120
P2	0021012011212022102220022212001211021002222101222200	M12	0222220012221012220220200212202222221002010022020220
M3	222001002001202022222000222221220122100222222020200	G12	2211020222212012212022000221022111211002012011000210
G3	0122022022010200222212020112120221101002022201220000	P12	2012002222022011122201002222011121221000011012020020
P3	2201022002012200200212212211102222101222022002120000	M13	2012002012000210020110210111202120002122022021222000
M4	0110020211121220122200100210020222022210222000221200	F13	2022000012202222220122210211201102001202010001022000
G4	011022222222222111020202222200201221112222022102000	P13	1022022020122102102221220211121122221022210202220220
P4	2222212020020222212220002022202201101211202012220200	M14	2220010010121202000201110020122001211022200022202220
M5	022202100222112020222222212222102022221202020122200	F14	0120220010102102000202220020222222111202100010022220
G5	2121022221021212101022002212022101022221021210200202	P14	02202222101001220222200202220202202222120100012222110
P5	2201220222002222122202202210210121000021021210202021	M15	222012200222002222022202010222021002220222022010202
M6	222222102102021200120222110222202021212022202202220	F15	2100121220222200022202022220022021202022221022020122
G6a	2222202200100022200212000120220202021010220022122210	P15	2100201220222200022202220122200021222001012212220110
G6b	20200022002020222102110202222112102012220221222201120	M16	1002212101001202220121212221102100021012000212200110
P6	0020002020220202222222000222110102022022122210122210	F16	200222222022222000201212202100122202210002010120110
M7	2022100212222002202002112222022202001022021202210220	P16	200002002221202120002101012212012112012002102222220
G7	1020202222020000001000211020021222021001002002212100	M17	2220222220112020100222201211211121201022001001000220
P7	122020210002022002220202121222102002222020021222220	F17	021012100012220210010122221222221001020002002202100
M8	1202202022001220021120020220010122201021202022221112	P17	2210201000222202100201120012220202002220002200202120
G8	1020222020202022201202202022220202102002102222022122	M18	2210020220202222220210020222000201222000210022222202
P8	2020012220202002121002102221220002100200220221002220	F18	022002000220020102002222022222201000000220002222102
M9	2102220121211202212002022022120102222002221121112100	P18	2020002202102121022202102010211202200000220000212100

Tableau 13 (suite) : Représentation sous forme de code ternaire des génotype obtenus par analyse des 52 SNP décrits dans le tableau 12. Les marqueurs sont représentés de gauche à droite dans l'ordre croissant des numéros qui leur ont été attribués. x indique que le génotype n'a pas été déterminé.

Individu	Génotype sous forme de code ternaire	Individu	Génotype sous forme de code ternaire
M19	102002211022221012222202211212020120222222022020200	LAB 6	2220012012222002122201122212120200002121021221221110
F19	2020222210022212202022002112220122201211222020210200	LAB 7	212000012210211020002222022110210220222222001210210
P19	022022201002212200220000221220011022221012012220200	LAB 8	0111220112200212122222011212220110200002222021100102
M20	1xxx2x010x20x2012x0201xx212x221102221x0x00122x22x2x0	LAB 9	0102002200212200122202200211022220221000012022120220
F20	1xxx2x010x20x2022x0001xx222x211220001x0x22222x11x2x0	LAB 10	21010222222120020200222202202210022012122002210200
P20	2xxx0x020x02x2002x2202xx20xx222022002xxx12x02x22xxx0	LAB 11	0020000202221000212022020210221212022020002022122110
M21	0210222101222211002100121220200202202000012122121001	LAB 12	2022020022001022122120201212110012012222021001212120
G21	222002220120022202220222020120201001000212221121222	LAB 13	2121010020220200122020012210212212002212222001200210
O21c	2220002202220200201210002221212002002012010202102210	LAB 14	2221122022021011000200100011022202012221200011202200
M22	0222020012121221111220202222200012201201201202020000	LAB 15	0211012020021202222022122212221221222212212122200200
F22	0002010210212202221220000012120022001102201121210220	LAB 16	1012020201011000100202202222101002011100222110002220
P22	2020222100121022122120202112210100002210002021220220	LAB 17	0222012020102020220202022201212100221220222020012200
M23	0002002021021020212022000020220202201200002202122210	LAB 18	0011022202002202102200100222012210220022020121222122
F23	2202000222221220122222222101121012020220220222220	LAB 19	2122012220012212102102221012020222212100202022122010
G23	222202222002200121001020012210120001222020222222010	LAB 20	122001002002212212122200001122202002122222222100202
P23	2222210222220200202200200210102100000220022222202000	LAB 21	2122100110222221020001202012222100001221102222021122
M24	0010220222220202222202020012212222001112121220122100	LAB 22	0212000202022220111122000222202022002000022121210120
G24	2010000221222002210222020222010012001110212202212200	LAB 23	0110020211121200001222202112202200012220211220021200
P24	0120112210221002020100221100020222021210011022210221	LAB 24	0221202122021122212202202221012222221220122211000202
LAB 1	2201020200201020222200212111210102122121022001012010	LAB 25	2102002020022221110212101222012000001200020212200000
LAB 2	2002021212010222221022112211102200022222202010220110	LAB 26	1102220221212202120002200212222221102011122022020220
LAB 3	2010210220000000202022000212221100221221110012000202	LAB 27	2122000222122220022122120211222020211202120022020000
LAB 4	2020022210222122102210020212011101012202212212100000	LAB 28	212020211210222222222002021222220000122221021220110
LAB 5	1020200202122222120222022210012101001100221002210202	LAB 29	020201002120202222022002202022122102220002221120100

III.1.3.3. Evaluations statistiques

III.1.3.3.1. Les marqueurs

Cette étape de tri des marqueurs présélectionnés comprend l'évaluation du polymorphisme des loci défini par les fréquences alléliques, l'évaluation de leur écart de l'équilibre de Hardy Weinberg et la quantification de leurs pouvoirs discriminatoires définis par les fréquences des différents génotypes observés. Le faible effectif de la population étudiée a été pris en considération par l'application aux calculs de la correction de Levene [Levene, 1949].

III.1.3.3.1.1. FREQUENCES ET DEGRE D'HETEROZYGOTIE

Parmi les marqueurs sélectionnés la majorité présente des fréquences alléliques comprises entre 0,35 et 0,65 (Figure 28).

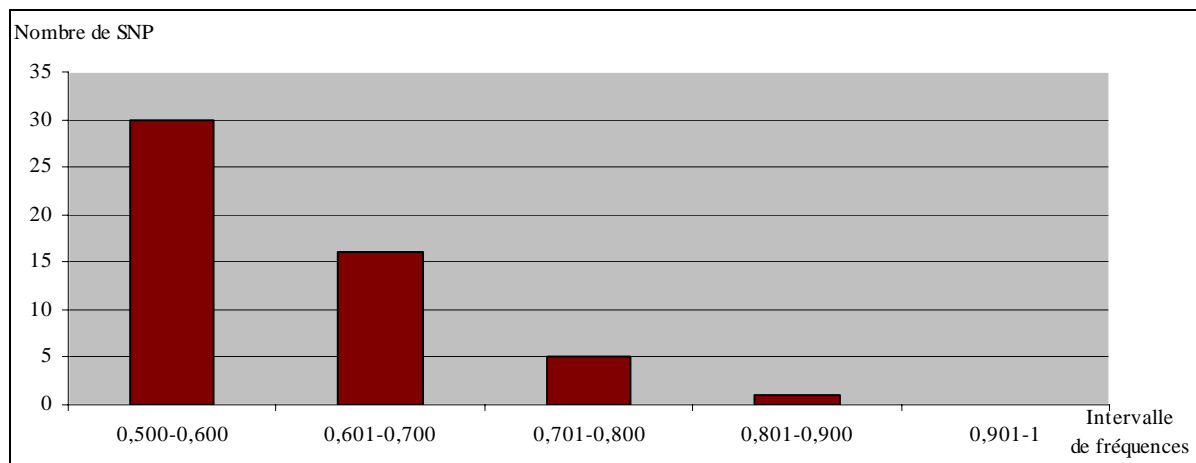


Figure 28 : Représentation graphique de la distribution du nombre de marqueurs en fonction des fréquences de l'allèle le plus représenté observées sur l'échantillon de population étudié.

Les degrés d'hétérozygotie observés (tableau 14) correspondent aux degrés d'hétérozygotie moyens rapportés dans les banques de données (Annexe 2).

Tableau 14 : Fréquences des allèles et degré d'hétérozygotie (H_o) observés pour chacun des SNP.

SNP	allèles	Fréquences des allèles	H_o	SNP	allèles	Fréquences des allèles	H_o
1	A / G	0,573 / 0,427	0,489	27	G / C	0,23 / 0,77	0,354
2	C / T	0,534 / 0,466	0,498	28	C / T	0,486 / 0,514	0,500
3	T / C	0,527 / 0,473	0,499	29	A / C	0,5 / 0,5	0,500
4	C / T	0,637 / 0,363	0,462	30	T / C	0,507 / 0,493	0,500
5	T / C	0,747 / 0,253	0,378	31	C / G	0,567 / 0,433	0,491
6	C / T	0,527 / 0,473	0,499	32	A / G	0,447 / 0,553	0,494
7	G / A	0,673 / 0,327	0,440	33	A / T	0,653 / 0,347	0,453
8	C / G	0,587 / 0,413	0,485	34	A / T	0,527 / 0,473	0,499
9	C / T	0,513 / 0,487	0,500	35	G / A	0,713 / 0,287	0,409
10	C / T	0,623 / 0,377	0,470	36	A / G	0,653 / 0,347	0,453
11	G / A	0,56 / 0,44	0,493	37	C / T	0,367 / 0,633	0,465
12	A / C	0,553 / 0,447	0,494	38	A / G	0,61 / 0,39	0,476
13	T / C	0,466 / 0,534	0,498	39	A / T	0,601 / 0,399	0,480
14	C / A	0,587 / 0,413	0,485	40	C / T	0,582 / 0,418	0,487
15	C / T	0,647 / 0,353	0,457	41	A / G	0,64 / 0,36	0,461
16	A / G	0,6 / 0,4	0,480	42	A / C	0,52 / 0,48	0,499
17	G / A	0,413 / 0,587	0,485	43	A / G	0,493 / 0,507	0,500
18	T / C	0,616 / 0,384	0,473	44	G / A	0,74 / 0,26	0,385
19	T / C	0,6 / 0,4	0,480	45	C / T	0,513 / 0,487	0,500
20	C / T	0,533 / 0,467	0,498	46	C / T	0,445 / 0,555	0,494
21	A / G	0,673 / 0,327	0,440	47	T / G	0,513 / 0,487	0,500
22	C / T	0,593 / 0,407	0,483	48	C / T	0,553 / 0,447	0,494
23	C / T	0,603 / 0,397	0,479	49	T / C	0,712 / 0,288	0,410
24	A / C	0,658 / 0,342	0,450	50	A / G	0,439 / 0,561	0,493
25	C / G	0,607 / 0,393	0,477	51	C / T	0,603 / 0,397	0,479
26	A / G	0,56 / 0,44	0,493	52	A / G	0,88 / 0,12	0,211

III.1.3.3.1.2. EQUILIBRE DE HARDY-WEINBERG

Les effectifs théoriques ont été calculés sur la base des fréquences alléliques obtenues selon les formules de la loi de Hardy-Weinberg. Les données ont été évaluées par un test du χ^2 qui permet d'estimer la probabilité p que l'écart entre les valeurs calculées et les valeurs observées soit aussi important par hasard. Au-delà de la valeur p_{critique} on considère que l'écart n'est pas lié à une raison biologique et que la population étudiée est à l'équilibre pour le marqueur considéré. Ainsi, en utilisant 2 degrés de liberté (en raison de l'existence de 2 allèles) et la valeur de p_{critique} à 0,05, on ne constate d'écart significatif par rapport aux proportions de Hardy-Weinberg que pour deux marqueurs (marqueurs 51 (rs868432) et 52 (rs754)) sur cette population (tableau 15).

Tableau 15 : Résultats des tests exacts de divergence de l'équilibre de Hardy-Weinberg pour les marqueurs étudiés. Les valeurs des erreurs types des calculs sont toutes inférieures à 0,001 (valeurs non présentées). Les marqueurs montrant des valeurs de p inférieures à la valeur de p_{critique} (0,05) sont surlignés en gris.

Marqueur	p	Marqueur	p	Marqueur	p	Marqueur	p
1	0,249	14	0,237	27	0,746	40	0,335
2	0,638	15	1,000	28	1,000	41	1,000
3	0,351	16	0,635	29	0,255	42	0,646
4	0,217	17	0,342	30	0,489	43	0,816
5	0,220	18	0,223	31	1,000	44	0,239
6	0,482	19	0,227	32	0,644	45	0,493
7	0,187	20	0,065	33	0,314	46	0,342
8	0,635	21	0,793	34	0,164	47	0,493
9	0,355	22	0,634	35	1,000	48	0,483
10	1,000	23	0,091	36	1,000	49	0,574
11	0,489	24	0,444	37	0,454	50	0,351
12	0,242	25	1,000	38	1,000	51	0,030
13	0,815	26	0,060	39	1,000	52	0,008

III.1.3.3.1.3. POUVOIR DISCRIMINATOIRE

Les valeurs de pouvoir discriminatoire de chaque marqueur (Tableau 16), correspondant à la probabilité que deux individus pris au hasard dans la population possèdent un génotype différent, sont toutes supérieures à 0,5 à part pour le marqueur 52 (rs754, cases grisées).

Tableau 16 : Valeurs de pouvoir discriminatoire (Pd) des marqueurs calculé à partir des fréquences des génotypes observés. Le marqueur présentant le pouvoir discriminant le plus faible est surligné en gris.

Marqueur	Pd	Marqueur	Pd	Marqueur	Pd	Marqueur	Pd
1	0,620	14	0,617	27	0,520	40	0,618
2	0,624	15	0,601	28	0,625	41	0,603
3	0,624	16	0,614	29	0,625	42	0,625
4	0,604	17	0,617	30	0,625	43	0,625
5	0,542	18	0,610	31	0,620	44	0,547
6	0,624	19	0,614	32	0,622	45	0,625
7	0,590	20	0,624	33	0,598	46	0,622
8	0,617	21	0,590	34	0,624	47	0,625
9	0,625	22	0,616	35	0,567	48	0,622
10	0,608	23	0,614	36	0,598	49	0,568
11	0,621	24	0,596	37	0,605	50	0,621
12	0,622	25	0,613	38	0,612	51	0,614
13	0,624	26	0,621	39	0,614	52	0,355

III.1.3.3.2. L'ensemble des marqueurs

Au vu du faible pouvoir de discrimination du marqueur 52 et du déséquilibre des marqueurs 51 et 52 par rapport à la loi de Hardy Weinberg, ils ont été exclus des évaluations de l'ensemble des marqueurs pour identification génétique.

III.1.3.3.2.1. DESEQUILIBRE DE LIAISON

Sur la base des résultats sur cette population (tableau 17), on ne constate pas de déséquilibre de liaison entre les marqueurs situés sur le même chromosome ($p < 0,05$). Les allèles des loci sélectionnés sont considérés comme transmis de manière indépendante à la descendance.

Tableau 17 : Résultats des tests exacts de déséquilibre de liaison pour les marqueurs situés sur le même chromosome. Les valeurs des erreurs types des calculs sont toutes inférieures à 0,001. P critique à 0,05 et 2 degrés de liberté.

Chromosome	1	2	2	3	3	4
Paire de loci	1 & 2	3 & 4	4 & 5	6 & 7	7 & 8	9 & 10
p	0,676	0,153	0,733	0,786	0,695	0,539
Chromosome	4	5	5	6	6	7
Paire de loci	10 & 11	12 & 13	13 & 14	15 & 16	16 & 17	18 & 19
p	0,995	0,708	0,210	0,706	0,346	0,521
Chromosome	7	7	8	8	9	9
Paire de loci	19 & 20	20 & 21	22 & 23	23 & 24	25 & 26	26 & 27
p	0,615	0,340	0,594	0,627	0,933	0,782
Chromosome	10	10	11	11	11	12
Paire de loci	28 & 29	29 & 30	31 & 32	32 & 33	33 & 34	35 & 36
p	0,831	0,238	0,710	0,614	0,738	0,553
Chromosome	17	18	18	21	22	
Paire de loci	40 & 41	42 & 43	43 & 44	47 & 48	49 & 50	
p	0,600	0,454	0,900	0,766	0,493	

III.1.3.3.2.2. QUALITE DE L'ENSEMBLE DES SNP

Le degré de polymorphisme ou pourcentage de marqueurs polymorphes dans le lot considéré est de 100%, puisque tous les loci présentent un degré d'hétérozygotie non nul. Le taux moyen d'hétérozygotie de cet ensemble de marqueurs sur l'échantillon de population considéré est de 47,4%. Ce paramètre est une mesure plus précise de la diversité génétique que le degré de polymorphisme.

La probabilité d'identité obtenue par l'analyse combinée des 50 marqueurs, à l'exclusion du marqueur de détermination du sexe, atteint $3,75 \cdot 10^{-21}$ et le pouvoir de discrimination atteint aisément celui des lots de marqueurs microsatellites analysés actuellement en routine.

III.1.3.3.2.3. FREQUENCE DU PROFIL

Les fréquences des profils, obtenues par le génotypage des 100 individus caucasiens, sont comprises entre $4,14 \cdot 10^{-28}$ et $2,90 \cdot 10^{-21}$ pour les profils complets (50 SNP). Les fréquences des profils partiels, établis à partir de 36 marqueurs pour les échantillons M20 et G20 et de 32 marqueurs pour l'échantillon PP20, n'excèdent pas $5,86 \cdot 10^{-13}$.

III.1.4. RESULTATS SUR LES PATERNITES

Les probabilités soutenant l'hypothèse de paternité, obtenues selon les formules explicitées au paragraphe II.9.2. pour les 20 cas de paternité confirmés, sont comprises entre 0,9996937 et 0,9999997. Seule la famille 20 présente une probabilité de paternité moindre (0,9979720), celle-ci étant calculée selon les résultats du génotypage de 32 SNP seulement.

Les probabilités de paternité de 0 ont été obtenues selon les génotypes de 7 loci pour le trio 23 (mère, garçon, fille et père présumé) et 8 loci pour les trios 22 et 24. Ces loci présentent des génotypes incompatibles avec l'héritage d'allèles du père présumé.

Pour le trio 21, l'exclusion n'est basée que sur les génotypes de 4 loci, le père présumé dans ce test de filiation étant le frère du père biologique de l'enfant. Ainsi s'opposent deux hypothèses: (1) les individus PP21 et G21 sont respectivement oncle et neveu et (2) les individus PP21 et G21 n'ont pas de lien de parenté. Selon ces hypothèses, le LR combiné atteint 5 109 609 799 147. Avec des probabilités *a priori* de 50%, ceci engendre la probabilité de l'hypothèse de la relation oncle et neveu de 0,99999999999998.

III.2. Marqueur pour la détermination du sexe

Le marqueur utilisé pour la détermination du sexe n'est en réalité pas un polymorphisme mais une différence ponctuelle de séquence entre les gènes de l'amélogénine des chromosomes X et Y (figure 29). Un individu de sexe féminin est caractérisé par la détection d'un seul allèle tandis qu'un individu de sexe masculin est caractérisé par une hétérozygotie (Figure 30).

Le choix de l'amorce pour la réaction de discrimination allélique de ce locus est très restreint. Selon la séquence représentée dans la figure 29, l'utilisation d'une amorce sens qui se termine par 5 cytosines consécutives du côté 3' pourrait engendrer des mésappariements très stables. Le nucléotide présent sur le chromosome X à ce locus étant aussi une cytosine, l'appariement de la suite de 5 C de l'amorce à une suite de 6 G de la séquence cible, pourrait aussi bien se

faire de manière très stable sur les cytosines 1 à 5 que sur celles de 2 à 6 empêchant ainsi l'extension. La sélection de l'amorce antisens est restreinte puisque la zone d'hybridation est bordée de nucléotides différant selon le chromosome étudié (annotés par des x dans la figure 29).

```

LFasta [Pearson et Lipman, 1988] v2.1u00 Mar, 2001

AMELX : séquence sur le chromosome X (AY040206)
AMELY : séquence sur le chromosome Y (BC074976)

AMELX      CATGGGTGGATGGCTGCACCACCAAATCATCCCCGTGCTGTCCCAACAGCA
            ::::::::::::::::::::::::::::::::::::::x::::::::::::::::::
AMELY      CATGGGTGGATGGCTGCACCACCAAATCATCCCCGTGGTGTCCCAACAGCA

AMELX      CCCCCCGACTCACACCCTGCAGCCTCATCACCACATCCCAGTGGTGCCA
            :::::x::::::::::::::::::::x::::::::::::::::::::::::::
AMELY      CCCCCTGACTCACACCCTGCAGTCTCATCACCACATCCCAGTGGTGCCA
    
```

Figure 29 : Alignement LFASTA [Pearson et Lipman, 1988] des séquences de fragments contenant la différence de séquence entre les copies du gène de l'amélogénine sur les chromosomes X et Y. Les séquences sont extraites des séquences génomiques contenues dans les bases de données nucléotidiques. Le nucléotide d'intérêt est accentué en gris. Les sites de fixation des amorces pour la réaction PCR correspondent aux séquences soulignées. Le site d'hybridation de l'amorce pour la réaction PEX est représenté en gras.

L'amorce ainsi choisie présente une température de fusion compatible avec une température d'hybridation de 51°C déterminée pour les réactions simultanées. Ainsi, ce marqueur nommé AmelSD, peut être analysé simultanément avec des SNP autosomaux du lot sélectionné.

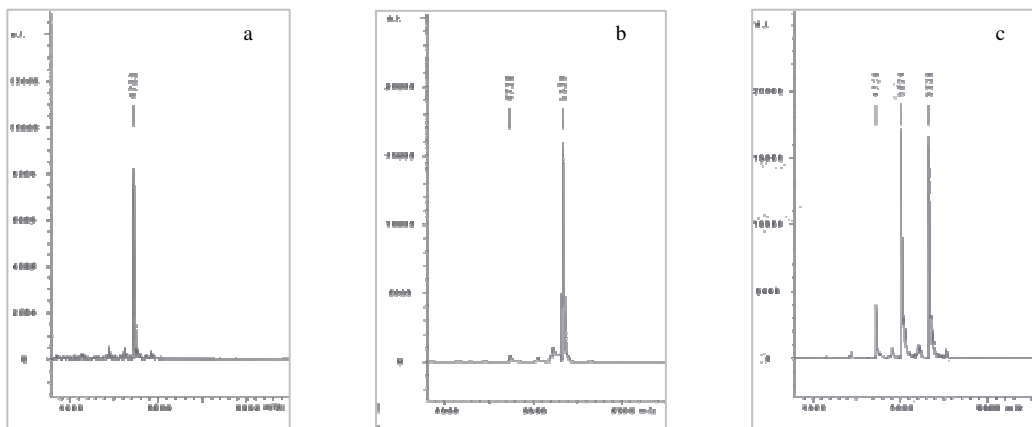


Figure 30 : Exemples de spectres obtenus lors de l'analyse du marqueur AmelSD sur le témoin négatif (a), un échantillon féminin (b) et masculin (c).

L'analyse de tous les échantillons décrits dans la section II.2. a permis la détermination du sexe biologique des individus sans ambiguïté. Aucune divergence entre sexe biologique et morphologique n'a été observée.

III.3. Les SNP du chromosome X

L'échantillon de population étudié révèle le polymorphisme de ces loci avec des fréquences alléliques observées qui concordent avec celles ayant servi de critères de sélection. Sur la base des résultats de l'analyse statistique, on ne constate pas de déséquilibre de liaison significatif entre les différents marqueurs ($p > 0.05$, tableau 18). Le génotypage de 120 hommes caucasoïdes sans lien de parenté sur ces 5 marqueurs a généré 29 profils différents, alors que 23 combinaisons alléliques différentes ont été observées lors de l'étude de 28 ADN féminins sans liens de parenté.

Tableau 18 : Résultats des tests exacts de déséquilibre de liaison pour les marqueurs situés sur le chromosome. Les valeurs des erreurs types des calculs sont toutes inférieures à 0,001. P critique à 0,05 et 2 degrés de liberté.

Paire de loci	p	Paire de loci	p
rs618 & rs16282	0,55454	rs616 & rs17379	0,50707
rs618 & rs616	0,07509	rs618 & rs17407	0,18628
rs16282 & rs616	0,77089	rs16282 & rs17407	0,92234
rs618 & rs17379	0,91078	rs616 & rs17407	0,67171
rs16282 & rs17379	0,44498	rs17379 & rs17407	0,24060

III.4. Les SNP du chromosome Y

III.4.1. LA DETECTION

L'analyse des SNP TAT, M242 et RPS4Y a été réalisée en une seule réaction PCR suivie d'une extension d'amorce simultanée. Le marqueur M242 se trouve dans une région du chromosome Y qui présente de fortes homologies avec une partie de la séquence du chromosome X. Ainsi, malgré le respect des règles de base pour le dessin des amorces, leurs

sites d'hybridations sont relativement éloignés du marqueur ciblé et le fragment amplifié est de longueur plus importante que souhaité (143 pb).

La nature des séquences bordant les trois marqueurs a permis le dessin d'amorces de taille différentes mais de température de fusion relativement proche, facilitant leur analyse simultanée. Les tests de sensibilité effectués sur des échantillons d'ADN masculin non dégradé ont permis de déterminer le seuil de PCR de ~25 pg d'ADN génomique nécessaires à l'obtention du génotype complet des 3 marqueurs (~4 copies du chromosome Y). En deçà, le passage graduel de l'amplification préférentielle des fragments de taille réduite conduit à l'absence d'amplification. Ce seuil a été déterminé par observation sous UV, en présence de SybrGreen et après séparation électrophorétique sur gel d'agarose des produits PCR ainsi que par détection par spectrométrie de masse des produits recherchés au point expérimental final.

III.4.2. TAT, M242 ET RPS4Y DANS LE CONTEXTE DES ECHANTILLONS

L'analyse des marqueurs SNP du chromosome Y a concerné des spécimens masculins de la nécropole de la vallée d'Egyin Gol [Keyser-Tracqui et al., 2003]. Six échantillons dégradés et faiblement concentrés étaient disponibles. Trois échantillons ont permis l'obtention de profils : deux complets (19EG et 112EG) et un partiel (70EG). L'échantillon 70EG, qui contenait de l'ADN fortement dégradé et en faible quantité (30 pg au total), a permis l'analyse du fragment cible le plus court (RPS4Y) seulement. Les trois profils distincts révèlent l'appartenance de ces spécimens à trois clades NRY majeurs (Tableau 19).

Tableau 19 : Profils obtenus sur les marqueurs M242, TAT et RPS4Y à partir de 3 spécimens anciens de la Nécropole de la vallée d'Egyin Gol et les haplogroupes auxquels ils appartiennent. La nomenclature des tombes et secteurs se réfère à l'article de Keyser-Tracqui et al., 2003.

Echantillon	Tombe	Secteur	M242	TAT	RPS4Y	Haplogroupe
19EG	25A	A	C	C	C	N3
112EG	50	B	T	T	C	Q
70EG	81	C	/	/	T	C

IV. Discussion

En médecine légale et en anthropologie, les circonstances à l'occasion desquelles se pose le problème de l'identification individuelle et du contrôle de filiation sont nombreuses. Aujourd'hui, la méthode de référence employée dans les domaines de l'identification et de la filiation génétique repose sur l'exploitation du polymorphisme des microsatellites. Ces séquences permettent de distinguer les individus qu'ils soient apparentés ou non [Cotton et al., 2000]. L'analyse de microsatellites autosomaux permet d'obtenir des résultats fiables, reproductibles et possédant un pouvoir de discrimination très élevé à partir de nombreux types d'échantillons [Buckleton et al., 2005]. Méthode bien ancrée dans les laboratoires, elle est désormais uniformisée et à la base du système d'identification que ce soit dans le cadre des connaissances des fréquences alléliques et des taux de mutation ou dans celui de la construction de bases de données nationales de profils génétiques.

IV.1. Etude de séquences de plus en plus courtes

Les microsatellites étant des polymorphismes de longueur de séquence, la taille des fragments produits pour leur détection (entre 100 et 450 paires de bases) reste un inconvénient majeur [Gill, 2002]. En effet, leur étude s'avère difficile à partir d'échantillons ayant subi une dégradation par l'action de facteurs physiques, tels que les rayonnements ultraviolets, l'humidité, la chaleur, ou de la contamination, bactérienne ou fongique. L'absence d'amplification de fragments de haut poids moléculaire aboutit à un profil partiel qui possède un pouvoir discriminant réduit [Golenberg et al., 1996]. L'ADN subissant une dégradation rapide par l'influence de différents facteurs dès la mort de la cellule, ce type d'échantillon est très fréquent en médecine légale et anthropologie moléculaire. Les catastrophes de masse, naturelles, accidentelles ou terroristes, représentent des exemples de situations dans lesquelles les restes des victimes peuvent varier d'un prélèvement intact à un échantillon sévèrement dégradé [Budowle et al., 2005].

L'investigation de l'ADN mitochondrial est moins sensible à la dégradation du fait de l'existence d'un très grand nombre de copies par cellule et de la structure circulaire du chondrome [Buckleton et al., 2005]. Néanmoins, cette analyse est beaucoup moins discriminante que celle des marqueurs nucléaires autosomaux car elle permet d'identifier une lignée maternelle et non un individu [Budowle et al., 2003].

Ces observations renforcent l'intérêt lié au développement de méthodes d'identification génétique basées sur l'analyse de fragments d'ADN plus courts augmentant ainsi la diversité des échantillons exploitables. En effet, la probabilité de rencontrer dans un échantillon un fragment d'intérêt intact est inversement proportionnelle à sa longueur.

L'objectif de notre travail a été la sélection d'un ensemble de SNP autosomaux répondant aux critères des marqueurs pour l'identification génétique. Une différence ponctuelle de séquence entre les copies du gène de l'amélogénine portées par les gonosomes a été choisie en tant que marqueur de détermination du sexe biologique d'un échantillon. Nous avons choisi d'étudier ces marqueurs par amplification simultanée des fragments les contenant, suivie de la discrimination allélique par extension allèle spécifique d'amorce dont les produits sont détectés par spectrométrie de masse MALDI-TOF. L'évaluation des marqueurs et de la méthode d'analyse a été faite sur un échantillon de 100 sujets non apparentés de la population occidentale et sur 24 trios père biologique présumé - enfant - mère.

IV.1.1. UTILISATION DES SNP

IV.1.1.1. Caractéristiques intéressantes des SNP en général

Les polymorphismes ponctuels de séquence représentent le type de variation le plus courant du génome humain avec une fréquence moyenne d'apparition de un SNP pour 1 kpb [Li et al., 1999]. En raison de leur nombre et de leur distribution, leur analyse représente un intérêt grandissant pour de multiples applications telles la génétique médicale, la pharmacogénétique ou les études d'associations de maladies.

Leur caractère biallélique permet une expression du résultat de détection simple (du type oui/non) compatible avec une grande diversité de plateformes d'analyse et un fort potentiel d'automatisation [Wise et al., 2003 ; Sobrino et al., 2005].

IV.1.1.2. Avantages des SNP par rapport aux STR

Les avantages de l'analyse des marqueurs SNP par rapport à celle des marqueurs traditionnels, permettent d'entrevoir un avenir pour ces mêmes applications.

En effet, la taille des fragments d'intérêt est inférieure à celle des fragments ciblés lors de l'analyse STR. Par définition, un SNP correspond à une seule base mais sa détection spécifique dépend de ses séquences bordantes.

De plus, le taux de mutation de ces marqueurs est d'environ 10^{-8} [Nachman et Crowell, 2000 ; Kondrashov, 2003] contre 10^{-3} pour les marqueurs STR [Brinkmann et al., 1998]. Ainsi, ces marqueurs portent l'information relative à des périodes de temps plus longues en comparaison des microsatellites ou des minisatellites [de Knijff, 2000 ; Jobling, 2001 ; Mountain et al., 2002 ; Jobling et al., 1999 ; Brion et al., 1999] puisque les allèles seront transmis sans altération sur un plus grand nombre de générations. Pour la même raison, ces marqueurs sont des outils très intéressants en termes de tests de parenté [Reich et al., 2002 ; Thomson et al., 2000].

Enfin, contrairement à l'analyse des STR, celle des SNP n'étant pas basée sur l'étude de la taille de fragments composés de séquences répétées, elle n'est pas confrontée au phénomène de "bégayement" de la polymérase lors de l'amplification et de ce fait ne souffre pas des difficultés d'interprétation pouvant être induites par la présence de *stutters*.

IV.1.1.3. Inconvénients des SNP par rapport aux STR

Derrière les avantages conférés aux SNP par leur caractère biallélique se cachent néanmoins des écueils. En effet, considérés individuellement, les SNP contiennent une information génétique très inférieure à celle accessible par les marqueurs multialléliques de type microsatellites. En effet, il est nécessaire d'analyser 4 à 5 marqueurs SNP pour égaler l'information contenue dans un microsatellite [Chakraborty et al., 1999 ; Coquoz, 2003]. Des évaluations statistiques plus précises montrent qu'en excluant la possibilité d'allèles nuls, l'analyse d'un nombre relativement réduit (50) de marqueurs bialléliques serait suffisante pour atteindre les valeurs de rapports de vraisemblance obtenues grâce à l'étude de 12 marqueurs microsatellites si leur fréquences alléliques sont comprises entre 0,2 et 0,8 [Gill, 2001].

IV.1.2. DETECTION PAR SPECTROMETRIE DE MASSE MALDI-TOF

L'analyse de marqueurs bialléliques est compatible avec une grande variété de plateformes technologiques et présente un fort potentiel d'automatisation [Sobrinho et al., 2005]. L'avantage principal de la spectrométrie de masse réside dans la mesure d'une propriété physique intrinsèque des analytes. Contrairement aux autres méthodes basées sur une analyse indirecte reposant sur l'interprétation d'un signal fluorescent ou radioactif émis par un indicateur, il s'agit de la détermination directe de la masse des molécules. Cette caractéristique permet (1) des optimisations et des résolutions de problèmes simplifiées par rapport aux méthodes indirectes (2) une sensibilité des résultats aux variations des conditions réactionnelles diminuée et (3) un avantage économique non négligeable dû à l'absence d'indicateurs.

L'application de l'étude des SNP à l'identification génétique nécessite l'analyse simultanée d'un grand nombre de marqueurs afin d'obtenir un pouvoir de discrimination significatif. La technologie MALDI - ionisation dite douce - est un procédé très rapide qui facilite la production d'ions de grandes biomolécules intactes, principalement monochargés et en phase gazeuse. La génération actuelle des instruments MALDI est capable d'enregistrer un spectre en moins d'une seconde. Son couplage à l'analyse du temps de vol est adapté à l'étude de mélanges moléculaires complexes car elle permet une fenêtre de détection de masses très large. En termes de génotypage de SNP, cela confère une forte capacité de multiplexage dont le degré ne peut être atteint par la plupart des autres méthodes d'analyse. La résolution des instruments actuels permettrait l'analyse simultanée de plusieurs dizaines de SNP à condition que ce même degré soit atteint dans les étapes précédant la détection [Mengel-Jorgensen et al., 2004]. En pratique, cela permettrait un débit d'analyse largement supérieur à 30000 génotypes SNP par jour.

L'application d'une méthode basée sur l'analyse de fragments assistée de fluorescence permettrait l'utilisation de l'équipement nécessaire à l'analyse des STR et éviterait ainsi un investissement de départ lourd. En électrophorèse capillaire, les amorces étendues sont séparées selon la couleur et la taille. Dans le but de réduire le risque de fausse annotation à un niveau négligeable, la séparation d'amorces étendues par des ddNTP marquées du même chromophore nécessite une différence de longueur d'au moins quatre nucléotides [Sanchez et al., 2003]. Dans ce contexte, le seul moyen pour augmenter le nombre de SNP analysés dans une seule expérience est d'augmenter la longueur des amorces. Leur analyse en grand nombre nécessite l'atteinte de tailles d'environ 100 nucléotides. Cependant, outre la diminution de la

qualité de tels fragments, la synthèse de tels oligonucléotides est difficile et coûteuse. En spectrométrie de masse, il est possible de distinguer sans ambiguïté deux à trois oligonucléotides de même longueur puisque leur masse est aussi dépendante de leur composition.

Enfin, la même plateforme peut aussi être employée pour l'analyse de méthylation de l'ADN [Tost et al., 2003, Schatz et al., 2004], de profils d'expression génique [Ding et Cantor, 2003], ainsi que pour la protéomique [Aebersold et Mann, 2003], ce qui fait d'un spectromètre de masse MALDI-TOF l'un des outils les plus polyvalents.

IV.2. La problématique au niveau de chaque étape technique

IV.2.1. LA PCR

La faible capacité de discrimination apportée par les marqueurs bialléliques en comparaison de celle obtenue à partir de marqueurs microsatellites pourrait en partie être compensée par leur analyse en grand nombre. Il est évident qu'il est impossible d'analyser 51 polymorphismes par le biais de réactions individuelles lorsqu'il s'agit d'étudier des traces biologiques, que ce soit dans le contexte de prélèvements sur une scène de crime, sur un lieu de catastrophe de masse ou sur un site de fouilles archéologiques. Outre la limite quantitative de matrice, un tel nombre de réaction et de manipulations multiplie les risques de contamination et le niveau de complexité de l'analyse. La difficulté réside dans l'amplification simultanée de manière robuste et équilibrée d'un tel nombre de fragments [Pastinen et al., 2000].

Dans cette étude, l'amplification des fragments contenant les 51 polymorphismes d'intérêt a été optimisée en 5 réactions PCR. Selon nos premiers résultats, l'obtention d'un profil complet sur 51 marqueurs est aisée à partir de 900 pg d'ADN matrice (180 pg d'ADN par réaction). Cela représente la limite inférieure de ce que l'on nomme la quantité optimale d'ADN matrice. Des profils complets ont été obtenus à partir de quantités inférieures (500 pg d'ADN matrice au total, 100 pg par réaction) mais leur interprétation dans le cadre d'analyses de routine nécessite la définition de directives précises. En effet, lorsque la quantité d'ADN cible

diminue, les différences d'intensité augmentent, entre les produits correspondant aux différents SNP ainsi qu'entre les produits correspondant aux allèles d'un même SNP, et les résultats deviennent déséquilibrés. A l'image de celles établies pour l'analyse des microsatellites [Whitaker et al., 2001], les directives doivent être établies par la répétition des expériences à grande échelle et leur comparaison avec des "profils sur ADN optimal".

Dans cette étude, le succès des réactions d'amplification a été limité en premier lieu par la quantité d'ADN matrice disponible avant sa qualité. L'application d'un protocole d'extension d'amorce discriminant simultanément les allèles d'un nombre de loci plus important, couplée à un haut degré de multiplexage lors de la PCR (1 à 2 réactions d'amplification), permettraient certainement de réduire la limite inférieure de la quantité d'ADN matrice nécessaire. Cependant, l'augmentation importante du degré de multiplexage des réactions d'amplification diminue leur sensibilité et peut engendrer ainsi une perte d'informations. En effet, la sensibilité d'une réaction ciblant 15 à 16 loci ne pourra probablement pas être atteinte lors d'une réaction visant à amplifier plus de 50 marqueurs.

La fragmentation diminue l'efficacité d'amplification, même lorsque la quantité d'ADN matrice n'est pas limitante [Dixon et al., 2005]. Tant que l'analyse est basée sur une étape de PCR, sa sensibilité en est directement dépendante. Ainsi, quelque soit la technique de discrimination allélique ou de détection de ses produits, elle sera toujours dépendante de la faisabilité et de l'efficacité de la réaction PCR. De ce fait, l'utilisation des marqueurs SNP n'apporte en l'état actuel pas d'avancées en ce qui concerne la quantité d'ADN matrice minimale.

Les avancées technologiques et scientifiques ouvrent une nouvelle porte aux SNP par le développement de méthodes de discrimination allélique ne nécessitant pas de pré-amplification des fragments les contenant [Pastinen et al., 2000 ; De la Vega et al., 2005 ; Nie et al., 2005]. Celles-ci doivent être au minimum aussi sensibles que la PCR et permettraient la mise en place d'analyses en un seul tube, réduisant les étapes de purification et de ce fait le risque de contamination. De même, l'élimination de cette étape pourrait encore réduire la taille des fragments d'ADN d'intérêt en ne ciblant que les séquences adjacentes au marqueur.

Lors de l'analyse d'échantillons pauvres et dégradés, les biologistes sont confrontés au problème de la limite d'informativité de l'ADN dégradé et en très faible nombre de copies. En effet, lorsqu'un échantillon est riche en ADN dégradé, statistiquement les chances de trouver

et d'amplifier les fragments courts contenant les SNP sont meilleures que celles de trouver un STR (taille de fragment cible supérieure). Néanmoins, dans le cas d'un échantillon pauvre et dégradé, les chances de trouver les SNP dans des parties intactes sont tout aussi faibles que celles de trouver un STR intact car les coupures ne se feront pas de manière ordonnée autour des marqueurs (pour les 51 marqueurs). Ainsi, pour de tels échantillons, l'analyse des marqueurs SNP ne représente pas le moyen absolu d'identification mais permet d'étudier des matrices dégradées en quantités relativement importantes.

IV.2.2. PEX & MALDI-TOF

La spectrométrie de masse MALDI-TOF, méthode directe, très sensible et très rapide, est compatible avec toutes les méthodes de discrimination allélique décrites à ce jour. Celles-ci sont basées sur quatre principes : l'hybridation, la ligature, le clivage et l'extension d'amorce. Nous avons choisi la discrimination allélique simultanée de plusieurs SNP par extension d'amorce. De part la spécificité d'hybridation des amorces et celle de l'incorporation des nucléotides allèles spécifiques, la technique d'extension d'amorce présente un moyen très efficace de discrimination allélique. Basée sur l'amplification linéaire des entités à détecter, augmentant ainsi la sensibilité globale de la méthode, elle répond aux besoins qualitatifs relatifs à l'identification génétique. La combinaison technologique retenue concurrence la majorité des autres méthodes en termes de robustesse, précision, reproductibilité et sensibilité lorsqu'elles sont directement comparées [Tost et Gut, 2005 ; Le Hellard et al., 2002 ; Lessig et al., 2005].

Nous avons choisi d'appliquer la discrimination à son état le plus simple, c'est à dire pour l'analyse directe d'entités issues de la réaction d'extension d'amorce en phase liquide, sans modificateur de poids et sans digestion enzymatique. Cette approche permet une orientation de la méthode vers toute adaptation ou modification aidant la détection et convient à l'étape de sélection de marqueurs dans le processus de développement d'une technologie. Cependant, lors de l'analyse, ce choix restreint le nombre de marqueurs qui peuvent être étudiés simultanément, réduit la gamme de taille des fragments cibles et ne permet la détection de l'apparition d'un nouvel allèle que si les nucléotides nécessaires sont présents dans le milieu réactionnel [Kostrzewa et al., 2002 ; Sauer et al., 2000a ; Sauer et al., 2000b]. D'une part le

multiplexage nécessite des propriétés similaires des amorces (des températures de fusion proches restreignent la diversité de longueur) et d'autre part les mélanges de dNTP et ddNTP doivent être spécifiques des polymorphismes cibles. En moyenne, la différence de masse ainsi obtenue entre les différents produits relatifs à un SNP est de 300 Da [Buetow et al., 2001 ; Petkovski et al., 2005]. Dans cette étude, les produits attendus lors de l'analyse simultanée de plusieurs SNP sont séparés d'au moins 60 Da permettant de larges fenêtres de lecture de chaque pic sans chevauchement des fenêtres d'identification lors de l'attribution des masses. La distribution des masses observées suit la loi normale et les déviations standard sont inférieures à 4 Da. Afin de simplifier la lecture et surtout par manque d'automatisation de cette étape, la même fenêtre de lecture est appliquée pour tous les produits. Ainsi, au moins 99,7% des pics de chaque produit seront compris dans l'intervalle de 12 Da de part et d'autre de la masse attendue ce qui correspond à 3 déviations standard maximales ($m \pm 3\sigma$). En considérant les produits les plus proches en masse (analytes AB2T - C et SP02 - G) et en appliquant la fenêtre de lecture de masse de 24 Da, l'intervalle le plus petit les séparant serait de 36 Da. En conséquence, le risque de fausse annotation de masse est quasi nul.

Ces résultats, ainsi que toutes les attributions de masses dans cette étude, ont été obtenus grâce à la fonction d'annotation "centroid" qui ne prend en compte que les 10% supérieurs de chaque pic. Outre une plus faible différence entre les masses attendues et observées, cette fonction permet de s'affranchir d'une partie du bruit de fond qui peut être engendré par des résidus salins.

Des différences de hauteurs (40%) de hauteur entre les pics correspondant aux deux allèles d'un génotype hétérozygote sont classiquement observés. Le même ordre de grandeur de différences a également été observé entre deux dépôts consécutifs. Cela peut s'expliquer par les variations engendrées lors de la préparation du dépôt sur la cible. En effet, plusieurs facteurs restent non contrôlables dans cette étape. D'une part, la nécessité de la parfaite homogénéisation de l'échantillon est soulignée. D'autre part, la cristallisation de la matrice ne se fait pas de manière homogène et les cristaux formés peuvent avoir un contenu en ADN différent. Il est donc indispensable d'ioniser par tir au laser de nombreux points d'un même dépôt. Enfin, la cristallisation de la matrice est fortement influencée par la qualité de l'environnement dans lequel elle a lieu. Une variation du taux d'humidité ou du flux d'air se répercute dans la qualité du dépôt. Ainsi, il est apparu nécessaire de préparer tous les dépôts en double, lors de l'analyse manuelle, afin de minimiser la probabilité de perte d'information

due au dépôt. Le nombre de dépôt recommandé par le fournisseur lors de l'analyse automatique est de 4 pour chaque analyse.

L'optimisation des dépôts au travers d'une cristallisation plus homogène permettrait l'utilisation de la spectrométrie de masse MALDI-TOF de manière quantitative. En effet, contrairement aux peptides et aux protéines, l'ADN est une molécule "monotone" dont les différentes séquences affichent peu ou pas de différences d'efficacité de désorption. Dans le cadre de l'identification génétique par les SNP, l'aspect quantitatif de la méthode de détection joue un rôle très important dans l'investigation de mélanges d'ADN de plusieurs individus. En effet, contrairement à l'analyse de systèmes STR, l'étude de marqueurs bialléliques à partir d'échantillons contenant un mélange d'ADN ne permet pas d'attribuer les allèles de manière qualitative. La détection des deux allèles à un locus donné peut correspondre à plusieurs cas de figures. Une différence d'intensité des produits correspondant aux deux allèles peut indiquer alors leur différence quantitative de représentation dans l'échantillon. Dans une certaine mesure cette différence peut permettre de distinguer les génotypes générant le profil du mélange. Cependant, la distinction entre un ADN hétérozygote et un mélange d'ADN hétérozygotes ou de génotypes homozygotes opposés en proportions égales, reste impossible. De même, l'analyse d'un mélange d'ADN en proportions fortement déséquilibrées (tels que dans les prélèvements effectués suite à des agressions sexuelles) s'avère très difficile car les variations d'intensité peuvent être attribuées à des facteurs expérimentaux. Enfin, certaines aberrations génétiques, telles les trisomies ou les duplications de séquence, ainsi que les artéfacts techniques pourraient donner lieu à de forts déséquilibres quantitatifs des produits allèle spécifiques et mener à de fausses interprétations des profils.

L'utilisation d'une méthode de détection de sensibilité subfemtolaire, non spécifique de l'ADN présente aussi des inconvénients. Des composants de tampons de réaction des enzymes, tels les détergents, les stabilisants, le glycérol et surtout les sels, peuvent sévèrement interférer avec les spectres [Guo, 1999]. Ainsi, un pic correspondant à un produit présent en de très faibles quantités peut être confondu avec des pics du bruit de fond. Dans le cas de mélanges d'ADN inégaux, cette situation est rendue encore plus complexe par la présence de pics de forte intensité correspondant au profil de l'ADN majoritaire. Cette difficulté est réduite lors de la détection de fluorescence puisque les produits non marqués (ne correspondant pas à de l'ADN) ne sont pas détectés. Afin de réduire la présence de pics parasites dans les spectres, il est indispensable de procéder à des étapes de purification rigoureuse. Plusieurs fournisseurs

proposent des formats de purification différents. L'une de ces solutions est la technique "GOOD Assay" [Sauer et al., 2000a, 2000b] qui utilise des amorces modifiées et évite ainsi toute étape de purification.

L'analyse de SNP par miniséquençage suivi de la détection par spectrométrie de masse MALDI-TOF développée dans cette étude s'avère efficace, fiable, rapide, peu coûteuse, sensible, spécifique et génère des résultats reproductibles. Cependant, l'analyse d'un nombre de loci aussi important nécessite l'augmentation du degré de multiplexage tant au niveau de la pré-amplification (afin de diminuer la quantité d'ADN matrice nécessaire à l'analyse), qu'au niveau de la discrimination allélique (afin d'en diminuer la complexité expérimentale et le coût).

Un niveau supérieur de multiplexage dans cette deuxième étape pourrait être réalisé par la modification de deux paramètres. L'utilisation de ddNTP seuls et non mélangés à des dNTP simplifierait grandement la composition des milieux réactionnels. Cela peut être atteint par la sélection d'un nouvel ensemble de SNP à l'exclusion des polymorphismes A/T dont la distinction des génotypes AA, TT et AT est possible mais difficile puisqu'il s'agit de distinguer des entités de plusieurs kilodaltons (kDa) ne différant que de 9 Da [Haff et Smirnov, 1997a]. Cependant, selon la précision de la détermination de la masse des analytes calculée, il serait nécessaire de restreindre les polymorphismes à A/C, T/G et G/C. C'est seulement dans ces cas là que la différence de masse permet d'avoir des intervalles de confiance non chevauchants, comprenant 99,7% des pics. Afin de pallier à cette limitation, il est possible d'utiliser des ddNTP portant des masses ajoutées différentes, permettant de réguler la différence de masse de manière précise [Fei et al., 1998 ; Mengel-Jorgensen et al., 2004].

La corrélation entre les écart-types des distributions des masses et la masse moyenne est positive. Il serait intéressant d'exploiter la capacité de très haute résolution des analytes de masse inférieure. L'utilisation d'amorces clivables permettrait leur extension dans les mêmes conditions réactionnelles mais leurs produits après clivage auraient une masse inférieure. Les amorces contenant un lien photolabile (photo linker) permettent le clivage des produits d'extension par rayonnement UV [Wenzel et al., 2003 ; Vallone et al., 2005]. La digestion enzymatique à la ribonucléase peut être effectuée de manière spécifique sur les oligonucléotides contenant un ribonucléotide dans leur séquence [Mengel-Jorgensen et al., 2005].

La combinaison de produits de basse masse (produits pouvant être clivés) et de masse moyenne (amorces non modifiées et éventuellement distinguées par ajout de queues de séquences non humaines à l'extrémité 3' des amorces [Ross et al., 1998 ; Kim et al., 2003]) permet l'exploitation d'une fenêtre de détection plus large et par conséquent aboutit à l'analyse simultanée d'un grand nombre de SNP.

IV.3. SNP : "to be or not to be"

Un ensemble de 50 SNP autosomaux répondant aux critères des marqueurs pour l'identification génétique a été sélectionné. Une différence ponctuelle de séquence entre les gènes de l'amélogénine portés par les chromosomes sexuels a été choisie en tant que marqueur de détermination du sexe biologique d'un échantillon.

La sélection théorique est basée sur les informations disponibles dans les bases de données publiques qui aujourd'hui répertorient près de 10 millions de SNP. Grâce à cet outil, des marqueurs de position génomique connue, qui ne sont ni exoniques ni rapportés comme liés à un caractère phénotypique et présentant des fréquences alléliques de 0,15 à 0,85 dans au moins 3 populations différentes, ont pu être choisis.

L'étape expérimentale a permis d'exclure des marqueurs (tel le marqueur rs1321158) dont le voisinage proche n'était pas exempt de variabilité. En effet, la présence de polymorphismes dans les séquences bordantes pourrait d'une part limiter voire empêcher l'hybridation des amorces de discrimination allélique et d'autre part rendre impossible l'étude des mêmes marqueurs par d'autres techniques d'analyse. La validation expérimentale comprend aussi la comparaison des profils obtenus par séquençage et par l'extension d'amorce suivie de spectrométrie de masse sur fragments plus courts. Par cette comparaison de fragments amplifiés par des couples d'amorces différents il est possible de s'assurer de l'absence d'allèles nuls. Aucune divergence de génotypes n'a été observée lors de la comparaison des deux approches.

IV.3.1. IDENTIFICATION GENETIQUE

IV.3.1.1. Evaluation des loci

Un aspect important de l'utilisation des marqueurs bialléliques réside dans leur tendance à être spécifiques d'une population ; un marqueur qui est polymorphe dans une population peut ne pas l'être dans une autre. Cette tendance, néanmoins observée, est moindre pour les marqueurs microsatellites qui sont généralement polymorphes à travers diverses populations. Dans les deux cas (SNP et STR), la plupart des génotypes se retrouvent dans toutes les populations humaines. Cependant, leur fréquence peut varier d'une population à l'autre. Cela signifie que les marqueurs SNP doivent être générés spécifiquement pour la population étudiée et que des études détaillées de très grands échantillons de populations doivent précéder leur généralisation.

Lors de cette étude, nous avons observé des divergences plus ou moins importantes entre les fréquences alléliques de référence (base de données dbSNP) et celles obtenues sur l'échantillon de population française. Ces écarts peuvent s'expliquer par la taille réduite de la population étudiée (ceci concerne autant la population objet de ce travail que celle ayant permis l'estimation des fréquences de référence) et soulignent les différences de distribution alléliques. L'exemple le plus prononcé de ces différences est celui du marqueur 27 (rs936250) dont les fréquences alléliques observées (G - C : 0,23 - 0,77) s'éloignent de manière très importante des fréquences rapportées sur la collection de 30 individus caucasiens sans lien de parenté de CELERA (G - C : 0,707 - 0,293) où celles-ci sont quasiment inversées.

Les distributions alléliques population spécifique de certains marqueurs SNP pourraient être exploitées comme indicateurs de population d'origine d'un échantillon à l'image des marqueurs du chromosome Y [Jobling, 2001 ; Phillips et al., 2004 ; Brion et al., 2005]. Dans le cadre judiciaire, cette possibilité apparaît très intéressante en l'absence de profil de référence.

Les paramètres tels que l'hétérozygotie (ou diversité allélique), le pouvoir de discrimination et la probabilité d'identité ont été calculés pour chaque locus, selon les formules détaillées dans le paragraphe II.11.1. Ces résultats permettent de constater que les tests sont hautement discriminatifs puisque le pouvoir de discrimination (Pd) de chacun des marqueurs est supérieur à 0,5.

Le déséquilibre apparent d'un marqueur dans une population par rapport à la loi de Hardy-Weinberg indique un événement relativement récent, telle une mutation, une migration voire un effet de pression de sélection, qui empêche l'application des lois biostatistiques sur sa transmission. De même, un déséquilibre apparent par rapport à la loi de Hardy-Weinberg peut être induit par la taille faible ou une organisation en sous-populations de l'échantillon. Ainsi, les loci caractérisés par un déséquilibre dans l'hérédité ne peuvent être utilisés en tant que marqueurs pour l'identification génétique. Les loci rs868432 et rs754, cités dans ce travail à titre illustratif, sont des exemples de marqueurs n'ayant pas franchi l'étape de la sélection statistique. Les 50 SNP sélectionnés ne présentant pas d'écart significatif des effectifs théoriques, la transmission de leurs allèles étant conformément à la relation de Hardy Weinberg est admise.

Le critère non explicité par les résultats de l'évaluation individuelle de chaque marqueur est l'estimation du déséquilibre de liaison. Son évaluation par des tests exacts est indispensable malgré la distance de sélection minimale entre 2 marqueurs fixée à 4 Mpb. En effet, une telle liaison peut s'étendre sur des régions de plusieurs millions de paires de bases [Collins et al., 2001] et varie grandement entre les femmes et les hommes [Broman et al., 1998]. Toutefois, si une moyenne devait être utilisée, la distance de 1 cM sur la carte génétique correspond à environ 1 Mpb sur la carte physique [Petes, 2001]. Le déséquilibre de liaison de deux marqueurs se traduit par leur transmission non indépendante d'une génération à l'autre et leur analyse réduit l'information qui en découle voire revient à l'analyse de l'un des deux loci seulement.

Au vu des résultats obtenus les 50 marqueurs sélectionnés sont tous polymorphes, à l'équilibre de Hardy-Weinberg et ne présentent pas de déséquilibre de liaison dans la population étudiée. L'identification génétique, reposant sur l'étude de cet ensemble de SNP, s'avère efficace au point qu'elle permet d'atteindre une probabilité d'identité de $3,75 \cdot 10^{-21}$. Ce paramètre exprime la probabilité d'avoir deux génotypes identiques en analysant deux individus pris au hasard dans la population considérée. La valeur de ce paramètre, obtenue grâce aux kits d'analyse des

marqueurs STR utilisés aujourd'hui en routine, est comprise entre 10^{-21} et 10^{-16} selon le nombre de loci étudiés [Buckleton et al., 2005]. Ainsi, les pouvoirs discriminants obtenus, par l'usage de "kits STR" d'une part et grâce à l'analyse de l'ensemble de SNP décrit dans cette étude d'autre part, sont équivalents.

IV.3.1.2. Etude de mélanges

Parmi les marqueurs sélectionnés, 90% présentent des fréquences alléliques comprises entre 0,35 et 0,65. Dans le cadre d'un calcul d'identité entre deux génotypes et de l'étude d'une trace biologique simple, les SNP dont les fréquences alléliques sont identiques (0,5/0,5) sont les plus informatifs. En effet, en considérant un ensemble de 50 SNP de distributions alléliques identiques, les nombres de génotypes homozygotes et hétérozygotes observés seront respectivement $50(a^2+b^2)$ et $100ab$, si a et b représentent les fréquences des allèles possibles à chaque locus. Les rapports de vraisemblance, opposant les hypothèses d'identité et de non-identité connaissant le profil obtenu, seront calculés selon la formule $LR_{50} = \left(\frac{1}{a^2}\right)^{50a^2} \left(\frac{1}{2ab}\right)^{100ab} \left(\frac{1}{b^2}\right)^{50b^2}$. La courbe représentant les LR en fonction des distributions alléliques sera symétrique par rapport aux fréquences 0,5/0,5. En effet, le rapport de vraisemblance obtenu grâce à 50 marqueurs de ce type atteint la valeur de $3,78 \cdot 10^{22}$, alors qu'il diminue jusqu'à atteindre $8,24 \cdot 10^{16}$ lors de l'analyse de 50 SNP dont les fréquences alléliques sont 0,2 et 0,8.

Dans les cas plus complexes de traces comportant l'ADN de plus d'une personne, le pouvoir discriminant des SNP chute et le nombre de marqueurs à analyser simultanément augmente de manière vertigineuse. L'interprétation de tels profils devient un défi puisque quelque soit le nombre de contributeurs, le nombre maximal d'allèles observés à chaque locus est de deux alors que les marqueurs STR permettent, dans le meilleur des cas, d'observer deux allèles par contributeur. Dans ce cas, l'analyse de 50 marqueurs dont les fréquences sont situées entre 0,2 et 0,8, permettrait d'atteindre des LR compris entre environ 10^4 et 10^6 . L'information contenue dans un SNP en fonction de ses fréquences alléliques suit une tendance inverse à la logique de l'interprétation d'échantillons simples. A l'approche des fréquences alléliques de 0,5, les LR augmentent dans l'étude d'un seul ADN, mais diminuent dans le cadre de l'analyse d'un mélange d'ADN de plusieurs individus et atteignent alors leur minimum. Ainsi les marqueurs dont les répartitions de fréquences alléliques sont moins homogènes (fréquence de l'allèle le

plus représenté comprise entre 0,65 et 0,80) se révèlent plus informatifs lors de l'analyse d'échantillons complexes sans pénaliser le pouvoir discriminatoire du lot lors de l'analyse d'ADN d'un seul contributeur. Malgré cela, la réduction de l'interprétation de profils complexes à la détection d'allèles peu représentés et l'interprétation quantitative des signaux obtenus pour des marqueurs bialléliques (paragraphe IV.2.2.) n'est pas envisageable.

IV.3.2. ETUDE DES PATERNITES

De façon à vérifier la fiabilité de filiations testées au travers des SNP sélectionnés, des tests de paternité ont été réalisés à partir d'échantillons déjà typés par d'autres méthodes.

Les probabilités soutenant l'hypothèse de paternité, obtenues pour les cas de paternité confirmés, atteignent aisément les valeurs obtenues grâce à la filiation par les marqueurs STR (valeurs supérieures à 99,999%). Les investigations ont aussi permis d'établir avec certitude une relation oncle - neveu. Ainsi, les marqueurs SNP s'avèrent être aussi des outils puissants dans l'établissement de liens de parenté. L'étude des non-paternités a permis d'exclure irréfutablement les pères présumés grâce à 8 ou 9 loci présentant des génotypes incompatibles avec l'héritage d'allèles. En théorie, considérant le faible taux de mutation des marqueurs SNP (10^{-9} à 10^{-8} mutation par génération) et supposant l'absence d'allèles nuls, l'incompatibilité entre les génotypes sur un seul locus serait nécessaire et suffisante pour exclure un père présumé. Dans cette optique, les SNP sont considérés comme marqueurs plus puissants que les STR (ou autres marqueurs constitués de séquences répétées) par certains auteurs [Carracedo et Sánchez-Diz, 2005 ; Sobrino et al., 2005].

Cependant, ceci n'est vrai que si l'on suppose l'absence d'allèles nuls. En se rapportant à des systèmes de pouvoirs discriminants équivalents, un exemple de kit commercial d'analyse des STR à 15 loci autosomaux d'une part et le lot de SNP sélectionné dans cette étude d'autre part, les marqueurs bialléliques présentent un inconvénient nouveau.

Les systèmes STR et SNP étant tous deux basés sur la PCR, les fréquences d'apparition d'allèles nuls ($f(0)$) peuvent être supposées égales et comprises entre 10^{-3} et $5 \cdot 10^{-3}$ [Clayton et al., 2004 ; Ballard et al., 2005, Leibel et al., 2003]. La probabilité, P_{locus} , d'observer une telle incompatibilité à un locus donné dans le cas de l'analyse d'un trio - mère, enfant et père biologique – est calculée en fonction des fréquences alléliques (p et q) du locus considéré.

$$P_{\text{locus}} = p f(0) (1-p-f(0)) + q f(0) (1-q-f(0))$$

$$P_1 = 1 - \prod_{i=1}^{50} (1 - P_{\text{locus}})$$

La probabilité, P_1 , d'observer au moins une telle incompatibilité due à un allèle nul lors de l'analyse de l'ensemble des marqueurs sélectionnés ($n=50$) dans cette étude est comprise entre 0,02336 et 0,11066 selon la valeur de $f(0)$ employée (10^{-3} ou $5 \cdot 10^{-3}$). La probabilité, P_2 , d'en observer deux est alors comprise entre 0,0003 et 0,0062. Respectivement, pour un système de marqueurs STR à 15 loci autosomaux, la valeur de P_1 est comprise entre 0,013 et 0,063 et celle de P_2 entre 0 et 0,0019 selon la fréquence d'apparition d'allèles nuls considérée.

Ainsi, l'observation d'exclusions apparentes dues à des allèles nuls sur 1 et 2 loci se produirait respectivement 2 et 3 fois plus souvent lors de l'analyse de SNP dans le cas où la possibilité de mutation n'est pas prise en compte. En considérant le taux de mutation des STR (ajout ou suppression d'une répétition) et une fréquence d'allèles nuls de 10^{-3} , la fréquence de "fausses" exclusions (qu'elles soient dues à une mutation ou à un allèle nul) est de 2,73% et 0,1‰ pour respectivement 1 et 2 loci. Dans le cas des SNP, ces valeurs sont respectivement 2,46% et 0,3‰. Il apparaît que la fréquence d'allèles nuls lors de l'étude d'un système SNP est équivalente à celle observée lors de l'étude d'un système STR additionnée du taux de mutation des microsatellites.

De plus, la fréquence d'apparition d'allèles nuls ainsi calculée reste sous-évaluée car l'étude de SNP implique une étape supplémentaire par rapport à celle des STR. En effet, les marqueurs STR sont analysés directement en tant que produits d'amplification PCR alors que, la majorité des méthodologies d'analyse des SNP requiert une réaction de discrimination allélique. Celle-ci repose sur l'hybridation d'une sonde ou d'une amorce. Ainsi, il est nécessaire de prendre en considération la possibilité d'erreur d'incorporation des nucléotides commise par la Taq polymérase lors des premiers cycles d'amplification dans la région d'hybridation voire même au niveau du polymorphisme. Lors de l'analyse de STR seules les erreurs de la polymérase produisant des "stutters" seront discernables alors qu'une mésincorporation ne pourra être observée et n'influencera pas le résultat de l'analyse.

IV.3.3. DETERMINATION DU SEXE

La détermination moléculaire du sexe des individus repose sur l'identification de séquences caractéristiques des chromosomes X et Y. La méthode d'analyse de fragments employée en routine repose sur une délétion de 6 pb du 1^{er} intron du gène de l'amélogénine porté par le chromosome X, en dehors des régions recombinantes [Sullivan et al., 1993]. Actuellement, les kits commerciaux d'amplification des STR autosomaux incluent systématiquement l'amplification de cette portion du gène de l'amélogénine. La faible taille des fragments analysés (106 et 112 pb) permet leur étude à partir d'échantillons dégradés et en faible quantité. Cependant, dans le cadre de l'identification génétique par SNP, le changement technologique ne permet pas l'étude de cette délétion. Afin de combiner à l'analyse de SNP un marqueur de détermination du sexe, une différence ponctuelle de séquence située dans la même région du génome et portée par les gonosomes a été sélectionnée (AmelSD). Les résultats obtenus pour cette détermination moléculaire du sexe n'ont montré aucune divergence par rapport au sexe phénotypique. L'étude d'une telle position par rapport à celle d'une délétion permet de générer un produit spécifique des deux chromosomes sexuels. De ce fait, elle permet d'une part un contrôle de la réaction et d'autre part la détection d'une éventuelle composante masculine dans le cadre où l'échantillon contiendrait un mélange de plusieurs individus.

Dans certains cas, il peut être observé un défaut d'amplification lors de l'analyse de fragment du gène de l'amélogénine du chromosome Y. Il est certain que de telles particularités seront rencontrées dans le cadre de l'étude du marqueur AmelSD. Ceci peut être dû à un défaut d'amplification (par exemple lors de la dégradation excessive de l'ADN cible) dans le cas des échantillons d'ADN anciens. Par ailleurs, 0,018% des hommes présentent un profil génétique féminin [Roffey et al., 2000 ; Steinlechner et al., 2002]. Certaines particularités génétiques telles des délétions de la région cible, des mutations dans les sites de fixation des amorces ou des phénomènes d'aneuploidie [Jobling, 2000 ; Brown, 2001] peuvent aussi engendrer un profil XX.

Bien que l'étude du marqueur AmelSD se soit révélée aussi fiable que celle basée sur l'analyse la délétion décrite ci-dessus, il serait intéressant d'inclure dans les tests une position nucléotidique du gène SRY (*Sex-determining Region on the Y chromosome*). En effet, chez les mammifères, la détermination du sexe est contrôlée par ce gène situé dans la partie terminale du bras court du chromosome Y, à proximité de la région pseudo-autosomique. Il joue un rôle dominant dans le déterminisme du sexe mâle puisque, un individu porteur de ce

gène sera de sexe masculin indépendamment du nombre de chromosomes X qu'il possède [Berta et al., 1990]. Le problème de non concordance entre le sexe génétique et le sexe phénotypique d'un individu peut être contourné par l'amplification d'une région spécifique du chromosome Y, tels des microsatellites ou un fragment de 93 pb du gène SRY (Santos et al., 1998). L'inclusion d'un marqueur du gène SRY dans l'ensemble de SNP décrit permettrait de réduire considérablement le taux de "fausses" déterminations du sexe génétique.

IV.3.4. LES SNP DU CHROMOSOME X

Au regard des caractéristiques d'hérédité du chromosome X, son analyse en criminalistique joue un rôle mineur. En effet, le pouvoir discriminant des marqueurs portés par ce chromosome est moins intéressant chez l'homme (hémizygote) que chez la femme (dizygote) qui hérite d'un haplotype X inchangé de son père et d'un chromosome X recombiné de sa mère.

Le génotypage de 120 hommes caucasoïdes sans lien de parenté sur ces 5 marqueurs a généré 29 profils différents, alors que 23 combinaisons alléliques différentes ont été observées lors de l'étude de 28 ADN féminins sans liens de parenté.

En dehors de leur taux de mutation très faible, les SNP partagent les caractéristiques de tous les marqueurs situés sur le chromosome X. De manière générale, leur pouvoir informatif est inférieur à celui des marqueurs autosomiaux mais vraisemblablement supérieur à celui des marqueurs du chromosome Y.

Malgré cela, leur application spécifique à certains cas de filiation peut apporter l'information que les marqueurs autosomiaux ne contiennent pas pour cause de recombinaison méiotique. Par exemple, les marqueurs du chromosome X peuvent présenter un pouvoir d'exclusion supérieur lorsque les pères présumés alternatifs sont en proche parenté (par exemple père et fils qui ne partagent pas d'allèles des marqueurs du chromosome X) [Szibor et al., 2003]. De même, les marqueurs du chromosome X, s'avèrent très utiles dans les tests de relation mère-fils en l'absence du père ou encore père-fille en l'absence de la mère. D'autres investigations, telles l'étude de paternité par l'analyse des chromosomes X de deux demi-sœurs, ou de la grand-mère paternelle et la petite-fille, peuvent être facilitées.

IV.4. *SNP versus STR*

Malgré les avantages cités, l'analyse des SNP présente des limites.

Les données existantes, collectées dans les banques de données nationales ou au niveau des laboratoires, sont toutes basées sur les marqueurs STR. Le retypage des SNP sur des échantillons existants engendrerait un coût très élevé. De plus, cette ré-analyse ne serait pas envisageable. En effet, contrairement aux prélèvements de question (traces), les nouvelles directives de la réglementation française n'imposent plus la conservation des échantillons issus de prélèvements de référence destinés à l'alimentation du fichier national des empreintes génétiques. Bien qu'applicable dans le cadre de l'identification de victimes [Budowle, 2005], l'utilisation de marqueurs différents des STR rendrait la consultation de bases de données au niveau international obsolète à court ou moyen terme.

L'adoption d'une méthode basée sur la spectrométrie de masse MALDI-TOF permettrait une augmentation de la capacité et réduirait le coût des analyses. Pourtant les investissements de départ pour l'équipement doivent être pris en considération. Contrairement aux analyseurs génétiques basés sur l'électrophorèse capillaire, la spectrométrie de masse MALDI-TOF n'étant pas aussi largement implantée dans les laboratoires d'identification génétique, les coûts engendrés par l'acquisition de l'appareillage sont, à ce jour, élevés. L'impact de l'investissement pourrait être réduit par l'utilisation de la même plateforme pour des types d'analyses différents puisque la technologie bénéficie d'un spectre d'applications très large et diversifié [Lessig et al., 2005].

L'observation des résultats d'électrophorèse lors de l'analyse des marqueurs STR est déjà substantiellement automatisée, mais la lecture finale du résultat est réalisée et vérifiée manuellement [Coquoz, 2003 ; Buckelton et al., 2005]. Le nombre d'observations nécessaires par échantillon étant trop élevé, l'interprétation manuelle du résultat de l'investigation de marqueurs SNP ne peut être envisagée. L'automatisation doit concerner la lecture et l'interprétation du résultat. Par exemple, l'existence d'un mélange d'ADN de plusieurs individus ne pouvant être formellement démontrée ou exclue, l'interprétation devra en tenir compte pour tout échantillon dont la nature rend cette hypothèse vraisemblable. De même, l'analyse des SNP étant en premier lieu dédiée aux échantillons dégradés, elle sera malgré tout confrontée aux problèmes des profils partiels. Des indicateurs de qualité de la matrice pourraient être inclus, tel le ciblage de séquences non polymorphes du génome. Aussi des

paramètres de correction en découlant pourraient faciliter dans une certaine mesure l'interprétation des profils. Les calculs engendrés par la nécessité de grandes précautions, exigent le passage à une démarche purement informatique et statistique. Par le biais d'une collaboration avec la société Bruker Daltonics GmbH (Brème, Allemagne), nous avons initié le développement d'outils bioinformatiques adaptés à l'identification génétique, permettant la lecture, l'exploitation et l'interprétation automatisées des données SNP obtenues.

Les efforts dans le développement de technologies, plateformes et de la biochimie permettant un très haut débit d'analyse, ces progrès sont adaptés à l'industrie pharmaceutique où un nombre relativement faible de marqueurs est étudié sur un très grand nombre d'échantillons. En criminalistique et en paléanthropologie, il s'agit d'étudier un panel important de marqueurs sur de faibles quantités d'échantillons souvent dégradés. La méthode développée dans cette étude est économiquement très avantageuse puisqu'elle ne nécessite pas l'utilisation de marquage fluorescent. Cependant, l'utilisation d'amorces et de ddNTP non modifiées, diminue la capacité indispensable de haut degré de multiplexage des réactions. Ainsi, la difficulté principale réside dans la mise au point de réactions visant à amplifier un grand nombre de fragments simultanément. De même, l'augmentation du degré de multiplexage de réactions d'amplification diminue leur sensibilité. Le développement de réactions multiplexes larges permettrait certes de réduire la limite de sensibilité atteinte dans ce travail (500 pg), mais ne dépassera pas celle des kits STR reposant sur l'amplification d'une quinzaine de fragments.

Le caractère biallélique des SNP est limitant lors de l'interprétation de mélanges d'ADN de plusieurs individus. Un mélange analysé par SNP peut être caractérisé par un excès de loci hétérozygotes, mais cela ne permet pas la résolution des profils génétiques spécifiques des contributeurs. De même, une contamination sera difficile voire impossible à détecter. L'interprétation des mélanges est clairement la principale faiblesse des SNP.

Enfin, il faut noter que la limite des analyses de STR en multiplex face aux échantillons fortement dégradés pourrait être partiellement corrigée. Des stratégies sont envisagées consistant à utiliser des amorces particulièrement proches de la séquence répétitive pour obtenir des produits d'amplification de petite taille (inférieure à 130 pb) [Wiegand et Kleiber, 2001 ; Tsukada et al., 2002 ; Butler et al., 2003 ; Coble et Butler, 2005]. L'avantage majeur de ces STR plus courts ou mini-STR est leur compatibilité avec les banques de données construites jusqu'à présent. Néanmoins, certains amplicons ne pouvant être réduits de manière

satisfaisante, seront remplacés. De plus, ces produits plus courts sont compatibles avec des techniques alternatives d'analyse telle la spectrométrie de masse en temps de vol [Butler et al., 1998]. Toutefois, les SNP étant des polymorphismes ponctuelles, les fragments cibles peuvent être de taille inférieures à cette limite (60 à 80 pb).

Le remplacement des STR par des SNP pour l'identification génétique n'est pas envisageable dans un futur proche [Gill, 2004]. En France, il devra se confronter aux modifications de la réglementation en vigueur où les marqueurs pour l'identification génétique sont fixés. Néanmoins, l'analyse des SNP peut compléter celle des STR que ce soit dans le cadre d'échantillons ayant subi une dégradation importante ou dans celui des études de paternités ne permettant pas le calcul de valeur statistiques suffisantes. Dans le cadre d'une expertise, l'identification génétique donne des éléments de réponse voire une orientation à l'enquêteur. La réunion des résultats de l'investigation d'un certain nombre de SNP (même très inférieur à 50) avec un profil partiel obtenu à partir de marqueurs STR peut permettre d'améliorer les pouvoirs discriminants nécessaires à la résolution de cas difficiles.

IV.5. Les SNP dans les laboratoires d'identification génétique

L'utilisation des SNP en tant que marqueurs pour l'identification ou la filiation en médecine légale ne peut être écartée et leur implantation pourrait être très rapide pour certaines applications plus spécifiques.

Le polymorphisme de l'ADN mitochondrial, étudié pour identifier les lignées maternelles, correspond à des SNP. L'investigation est concentrée sur les régions hypervariables I et II (HVI et HVII) de la région contrôle et fournit un pouvoir de discrimination limité dans le contexte de la criminalistique et une information insuffisante dans les études de l'évolution [Quintáns et al., 2004]. Le séquençage de l'ADN mitochondrial dans son intégralité n'étant pas envisageable, l'analyse d'informations supplémentaires contenues dans la région codante semble être la stratégie de choix pour pallier cet inconvénient [Quintáns et al., 2004 ; Coble et al., 2005 ; Niederstatter et al., 2005 ; Divne et Allen, 2005 ; Hall et al., 2005 ; Budowle et al., 2005 ; Kline et al., 2005].

L'analyse de séquences d'ADN déterminant des caractères morphologiques, tels la couleur des yeux [Frudakis et al., 2003 ; Sturm et Frudakis, 2004] ou encore la pigmentation de la peau [Anno et al., 2005 ; Bonilla et al., 2005] présente un intérêt implicite dans le cadre de la recherche d'un suspect. Cependant, cette utilisation des marqueurs moléculaires soulève de nombreuses questions d'éthique. En outre, il est nécessaire de prendre garde au fait que les "populations" décrites, dont la définition repose sur l'ascendance géographique, pourraient être interprétées comme des "races" (un concept surtout social) et considérées à tort comme des concepts précis et significatifs fondés sur la biologie. De même, selon la législation française, les profils ADN ne doivent pas contenir de données dont la divulgation pourrait être problématique et doivent de ce fait porter sur des régions non-codantes. Cette restriction exclut les informations sur le sexe de l'individu.

Enfin, le faible taux de mutation des SNP permet leur utilisation dans la reconstruction de l'histoire humaine et la dispersion de l'homme anatomiquement moderne. Cette caractéristique est d'un intérêt particulier lorsque les marqueurs sont haplotypiques. En effet, les marqueurs de la partie NRY sont hérités comme un seul locus permettant la reconstruction de lignées paternelles sur des milliers d'années. Les combinaisons alléliques changent uniquement au travers de l'accumulation de mutations parallèles et récurrentes. Au regard de leur taux de

mutation respectifs, les SNP qu'il porte se réfèrent à des événements temporellement plus lointains que les STR. Ainsi, les combinaisons de SNP permettent la définition d'haplogroupes du chromosome Y, informatifs sur la diversité d'une population. Les combinaisons de STR permettent la définition des haplotypes, dont l'étude est utilisée pour l'investigation d'évènements démographiques plus récents [de Knijff et al., 1997 ; Forster et al., 2000 ; Zhivotovsky et al., 2004]. Outre l'étude de la composante masculine dans les échantillons d'ADN de plusieurs contributeurs, les SNP du chromosome Y sont des marqueurs de choix lorsqu'il s'agit d'élucider l'origine géographique d'un spécimen [Kayser et al., 2001 ; Flores et al., 2004 ; Jobling, 2001 ; Underhill et al., 2001 ; Jobling et Tyler-Smith, 2003 ; Mengel-Jorgensen et al., 2004 ; 2005].

IV.6. Etude des échantillons anciens

Dans ce travail, nous avons étudié trois SNP du chromosome Y sur des échantillons d'ADN masculin anciens obtenus à partir de restes osseux découverts dans la nécropole de la vallée d'Egyin Gol en Mongolie du nord. Cette nécropole relève de la culture Xiongnu, tribu Turco-Mongole de guerriers nomades contre lesquels les Chinois ont élevé la Grande Muraille. Au 3^{ème} siècle avant Jésus Christ, ils ont créé le premier empire dirigé par des nomades de l'Asie centrale. Ils ont régné sur le territoire s'étendant depuis le lac Baïkal au nord, jusqu'au désert de Gobi au sud et depuis l'ouest de la Manchourie jusqu'aux Pamirs à l'ouest. Du temps de la dynastie des Han (206 avant Jésus Christ - 220 après Jésus Christ), la Chine a élargi ses frontières au dépend de l'empire Xiongnu [Marx, 2000].

Selon la datation au carbone 14 le site d'Egyin Gol a été utilisé du 3^{ème} siècle avant Jésus Christ au 2nd siècle après Jésus Christ, c'est-à-dire durant toute la période Xiongnu. Les tombes sont organisées en quatre secteurs désignés "A", "B", "C" et "D", des deux côtés d'une faible dépression de la vallée (figure 31). Le site se serait étendu depuis le secteur A vers les secteurs B et C, avec certaines tombes du centre du secteur B qui semblent antérieures à l'ensemble B/C.

L'environnement froid de la Mongolie du Nord permet la conservation de tombes à l'état gelé. Ces conditions climatiques étant favorables à une bonne préservation de l'ADN [Burger et al., 1999 ; Leonard et al., 2000], les restes osseux excavés ont pu être étudiés au travers des

marqueurs STR biparentaux et du chromosome Y ainsi que du polymorphisme de l'ADN mitochondrial [Keyser-Tracqui et al., 2003].

Cette étude a permis de mettre en évidence des relations de proche parenté entre plusieurs des spécimens inhumés dans ce site et a fourni des informations supplémentaires sur l'organisation de la nécropole et les pratiques funéraires des Xiongnu. Notamment, les haplotypes trouvés pour des spécimens retrouvés dans un groupe de tombes relativement isolées, suggèrent des contacts entre les populations d'Europe et d'Asie, antérieurs au développement de la culture Xiongnu ainsi que l'émergence d'une composante Turque dans la tribu Xiongnu parallèle au vieillissement du site funéraire [Keyser-Tracqui et al., 2003]. Afin d'approfondir l'exploration de cette diversité suggérée par l'analyse des STR, nous avons étudié les marqueurs bialléliques TAT, M242 et RPS4Y sur des spécimens de cette nécropole [Petkovski et al., sous presse].

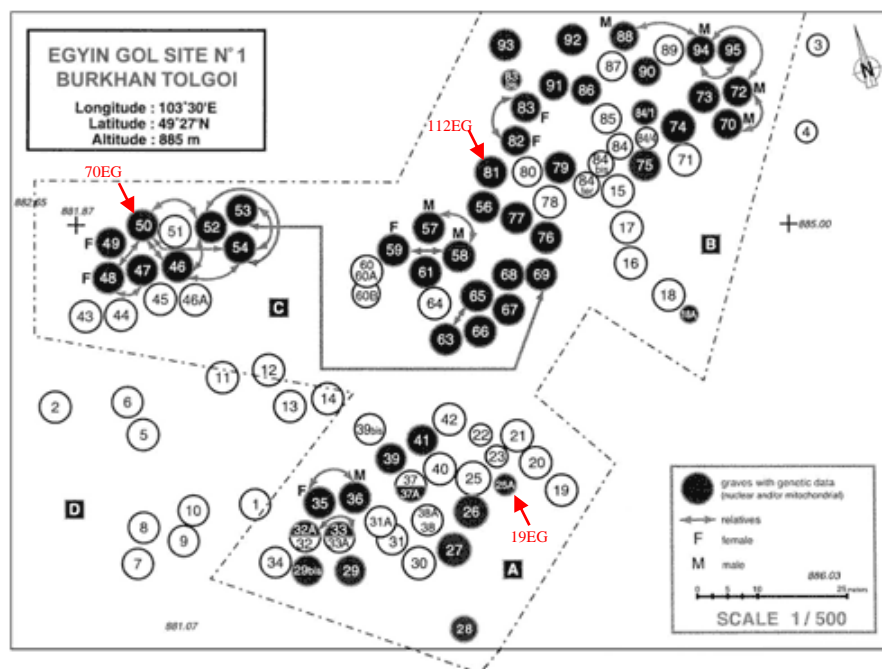


Figure 31 : Carte de la nécropole d'Egyin Gol. Les tombes sont matérialisées par des cercles. Les lettres A, B, C et D désignent les 4 secteurs distingués par Keyser-Tracqui et al. (2003) et délimités de pointillés. Les flèches rouges désignent les tombes étudiées par les marqueurs SNP dans cette étude.

Les trois spécimens étudiés appartiennent chacun à un secteur différent de la nécropole, défini grâce à la datation au carbone 14 (Figure 31). La tombe du spécimen 19EG est située dans le

secteur le plus ancien de la nécropole. Les tombes des individus 112EG et 70EG se trouvent dans les secteurs B et C respectivement.

L'allèle TAT-C, définissant l'haplogroupe N3, est observé à de très hautes fréquences chez les Yakoutes [Zerjal et al., 1997 ; Pakendorf et al., 2002 ; Karafet et al., 1999 ; Lell et al., 2002]. L'haplogroupe Q, défini par l'allèle M242-T, est observé sur tous les chromosomes Y amérindiens ne portant pas la mutation au locus RPS4Y [Seielstad et al., 2003]. De même, il est observé à des fréquences non négligeables dans les populations de Sibérie et d'Asie centrale [Underhill et al., 2000 ; Tambets et al., 2004]. La mutation M242_(C→T) est fréquente dans les populations Eurasiennes. Etant apparue après les mutations M45 et M74 mais antérieurement à la mutation M3, elle est antérieure à la première migration vers l'Amérique. Elle serait originaire d'Asie centrale et aurait migré par la région Altai / Baïkal de l'Eurasie du Nord vers l'Amérique. La mutation RPS4Y_{711(C→T)}, définissant l'haplogroupe C, est restreinte à l'Asie de l'Est et à l'Amérique [Bergen et al., 1998 ; 1999 ; Lell et al., 2002].

Parmi les individus masculins de cette nécropole seuls 2 appartiennent à l'haplogroupe N3 (TAT_{T→C}), l'un étant exhumé du secteur A (19EG) et l'autre, dont l'haplogroupe a été déterminé lors d'une autre étude, du centre du secteur B (tombe 69, figure 31) [Keyser-Tracqui et al., 2004]. Ainsi, dans cette nécropole, la présence de cet haplogroupe est antérieure à celle des 2 autres haplogroupes qui ont été retrouvés dans les secteurs plus récents du site. Le spécimen 70EG (secteur C), a été exhumé d'un regroupement de tombes d'individus génétiquement liés. Ces individus présentent le même haplotype Y et certains partagent la même séquence d'ADNmt. Ces lignées uniparentales ont été retrouvées chez des individus Turques. L'appartenance de l'individu 70EG à l'haplogroupe C, confirme et précise les contacts suggérés entre les populations d'Europe et d'Asie, antérieurs au développement de la culture Xiongnu. Malheureusement, dans le cadre de l'étude d'échantillons anciens, où la quantité de matrice disponible est très limitée, il n'est pas toujours possible de mettre en œuvre toutes les analyses. Le faible nombre d'échantillons analysés par ces marqueurs ne permet pas de conclusions certaines. Cependant, ils démontrent l'hétérogénéité de la tribu Xiongnu, suggérée par les résultats de l'analyse de marqueurs STR. Cette diversité présente un lien avec l'expansion temporelle de la nécropole. Au vu du poids de l'information que peut apporter l'approche SNP, il sera intéressant de l'exploiter lors de l'étude d'autres populations anciennes.

V. Conclusions

Les polymorphismes ponctuels de séquence constituent le type de variation le plus répandu du génome humain. Leur étude à des fins d'identification génétique et de filiation dans les domaines de la médecine légale et de l'anthropologie moléculaire, souffrant tous deux de limitations communes, présente donc un réel intérêt. En effet, les SNP autosomaux permettent, à l'image des STR autosomaux, d'atteindre un haut pouvoir discriminant mais, comme l'ADNmt, à partir d'échantillons dégradés par l'analyse de fragments d'ADN courts.

L'objectif du travail présenté était la sélection et la validation d'un lot de marqueurs SNP pour l'identification et la filiation génétique humaines et le développement d'une méthode basée sur la PCR permettant une analyse sensible, rapide, peu coûteuse et applicable à une grande variété d'échantillons. La définition de cet ensemble de marqueurs est un travail original dépassant ceux décrits dans la littérature tant d'un point de vue quantitatif menant à un pouvoir de discrimination élevé que d'un point de vue qualitatif en ne considérant que des marqueurs non codants. En effet, la législation française en vigueur exclut les régions codantes de l'ADN des études d'identification génétique. Les marqueurs sélectionnés lors de ce travail sont bialléliques, phénotypiquement neutres et de fréquences alléliques comprises entre 0,2 et 0,8.

Nous avons choisi d'étudier ces marqueurs par amplification simultanée des fragments les contenant, suivie de l'extension allèle spécifique d'amorce dont les produits sont détectés par spectrométrie de masse MALDI-TOF. Cette approche est directe, très sensible, très rapide et permet l'analyse simultanée d'un grand nombre de molécules.

Le génotypage de ces marqueurs sur un échantillon de la population française a permis de fournir des informations sur leur distribution allélique, de valider l'ensemble des SNP sélectionnés comme outil pour la filiation et l'identification génétique et de développer une méthode d'analyse fiable et rapide générant des résultats reproductibles.

En effet, selon nos résultats, ces marqueurs ne présentent pas de déséquilibre de liaison et permettent d'atteindre des pouvoirs de discrimination semblables à ceux obtenus par analyse de marqueurs STR dans le cadre de filiation et d'identification génétiques.

Bien que performante, la technique mise au point au cours de ce travail de thèse doit faire l'objet d'améliorations. D'après les résultats obtenus, le lot de marqueurs sélectionnés convient aux applications de l'identification génétique, puisqu'il permet d'atteindre le pouvoir discriminant des kits STR utilisés dans le cadre d'identité ou de filiation. La poursuite de ce travail consistera, d'une part, à compléter les données obtenues en augmentant considérablement la taille de l'échantillon étudié puis en élargissant les investigations à d'autres populations. D'autre part, elle s'articulera autour des améliorations techniques de la

méthode. En effet, il s'avère indispensable d'augmenter le degré de multiplexage des réactions d'amplification et de discrimination allélique. Une autre alternative consisterait en une discrimination allélique sensible et permettant une amplification du produit suffisante à sa détection, sans passer par l'étape de préamplification. La résolution de la spectrométrie de masse MALDI-TOF permettrait la détection simultanée de plusieurs dizaines d'analytes si la réaction de discrimination allélique engendre des produits couvrant un large éventail de tailles et donc de masses.

Le caractère biallélique des SNP et le pouvoir discriminatoire nécessaire à leur application aux domaines de l'identification génétique, impliquent l'analyse simultanée d'un grand nombre de ces marqueurs. Cette limitation est principalement due à la complexité de l'équilibre entre la quantité d'ADN matrice et le nombre de marqueurs à analyser par le biais de la technique PCR. De ce fait, l'avantage de l'étude des SNP sur celle des STR, lorsqu'il s'agit des marqueurs autosomaux, est manifeste lorsqu'il s'agit d'étudier les échantillons contenant de l'ADN dégradé dont le contenu en matrice n'est pas limitant. Par exemple, dans le cadre de l'identification de victimes de catastrophes de masse, la dégradation du matériel génétique favorise l'analyse de SNP, qui concerne des fragments courts, au détriment des marqueurs STR. De même, la combinaison d'un profil partiel sur des marqueurs STR avec le résultat de l'investigation de marqueurs SNP permettrait de repousser la limite d'informativité de l'ADN dégradé et en très faible nombre de copies. Ainsi, l'analyse des marqueurs bialléliques validés dans ce travail représente un outil complémentaire à l'investigation de microsatellites dans la manifestation de la vérité.

Le travail, objet de ce mémoire de thèse, a été réalisé dans le cadre d'une convention CIFRE (Convention Industrielle de Formation par la Recherche) dont l'objectif est d'allier recherches en laboratoire et applications industrielles. La définition d'un protocole de routine permettra pour des échantillons "modernes", d'introduire la technique d'identification basée sur les marqueurs SNP au niveau de laboratoires d'identification génétique.

De plus, en ciblant les SNP adéquats, la méthode développée dans ce travail intéresse de nombreux secteurs tels que la recherche biologique (et notamment la génomique fonctionnelle), la recherche pharmaceutique, le génotypage, le diagnostic ou les contrôles agro-alimentaires et industriels. Outre la criminalistique et la paléoanthropologie, où la quantité et la qualité de l'échantillon sont limitantes, la paléopathologie représente une discipline pour laquelle l'analyse des SNP par MALDI-TOF semble prometteuse.

Cette méthodologie touche différents marchés à forts potentiels économiques. Le spectromètre de masse MALDI-TOF peut être couplé à un robot préparateur permettant de traiter simultanément jusqu'à 384 analyses et l'acquisition du spectre ne dure que quelques secondes par échantillon. Ces analyses rapides et massivement parallèles produisent de nombreuses données, qui croissent exponentiellement. L'exploitation des résultats au moyen de programmes informatiques n'en est qu'à ses débuts. De nombreux développements sont en cours au niveau des logiciels d'analyse et de représentation afin d'extraire de ces données le maximum de sens biologique.

Le projet de génomique fonctionnelle devra permettre d'obtenir une connaissance plus fine de la fonction de ces séquences d'ADN. La spectrométrie de masse MALDI-TOF constitue un outil d'analyse fonctionnelle [Sauer et al., 2005] et de reséquençage [Fu et al., 1998 ; Krebs et al., 2003 ; Stanssens et al., 2004], permettant d'étudier à grande échelle des séquences déjà connues. C'est donc l'outil adapté pour la communauté scientifique travaillant en génomique pour établir des bases de données contenant des informations structurelles et fonctionnelles. La spectrométrie de masse MALDI-TOF joue ainsi un grand rôle dans le décryptage des SNP. Le potentiel de cette technique est important et la masse de données produite est énorme.

Le secteur de la pharmacogénomique étudie essentiellement l'expression des gènes et sa relation à l'efficacité d'un traitement médicamenteux. Il s'agit de déterminer la carte personnalisée des SNP du patient. L'accès libre à la connaissance et à l'utilisation des SNP est considéré comme primordial [Hoehe et al., 2003 ; Cashman et al., 2001 ; Lattard et al., 2003 ; Drysdale et al., 2000 ; Kim et al., 2003 ; Tost et al., 2002].

Une autre application de l'analyse des polymorphismes et de la détection banalisée de ceux-ci pourrait entraîner la commercialisation de systèmes de diagnostic clinique déterminant les prédispositions d'un patient à différentes maladies génétiques. Une fois le SNP ou le groupe de SNP identifié comme marqueur associé à une maladie il peut être utilisé pour le diagnostic [Sauer, 2005].

Les applications du génotypage de SNP ne sont pas restreintes à l'Homme. Les mêmes essais sont utilisés chez les animaux et les plantes. Par exemple, chez les animaux de ferme, ils sont utilisés pour la détermination de facteurs de risque génétique ou l'identification et la traçabilité animale [Laegreid et al., 2002 ; Kim et al., 2003 ; Nonneman et al., 2003 ; Heaton et al., 2001].

Enfin, les secteurs de la défense et de l'environnement envisagent aussi diverses applications pour la détection rapide et à faible coût d'éléments organiques, notamment des agents pathogènes dilués dans l'environnement [von Wintzingerode et al., 2002]. Les applications se

trouvent alors dans la protection contre la guerre bactériologique et chimique, dans la protection antiterroriste en milieu urbain, et dans la protection écologique.

Les SNP ont un avenir prometteur et la spectrométrie de masse MALDI-TOF présente un partenaire économiquement intéressant dans leur détection et leur exploitation. Ce sujet ciblé sur le domaine médico-légal ouvre vers de nombreuses applications dans les débats de sécurité et santé publiques.

VI. Références bibliographiques

- Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422(6928):198-207.
- Agah A, Aghajan M, Mashayekhi F, Amini S, Davis RW, Plummer JD et al. A multi-enzyme model for Pyrosequencing. *Nucleic Acids Res.* 2004;32(21):e166.
- Akane A, Seki S, Shiono H, Nakamura H, Hasegawa M, Kagawa M et al. Sex determination of forensic samples by dual PCR amplification of an X-Y homologous gene. *Forensic Sci. Int.* 1992;52(2):143-8.
- Amarger V, Gauguier D, Yerle M, Apiou F, Pinton P, Giraudeau F et al. Analysis of distribution in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* 1998;52(1):62-71.
- Ananthaswamy HN, Pierceall WE. Molecular mechanisms of ultraviolet radiation carcinogenesis. *Photochem. Photobiol.* 1990;52(6):1119-36.
- Anderson RR, Parrish JA. The optics of human skin. *J. Invest. Dermatol.* 1981;77(1):13-9.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981;290(5806):457-65.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 1999;23(2):147.
- Anno S, Kudo S, Hamasaki K, Matsumura K, Yamamoto T, Ogata K. Techniques of SNP genotyping and remote sensing applied for elucidation of the contribution of polygene and environmental factors to inter-individual variation in skin pigmentation phenotype. *J. Physiol. Anthropol. Appl. Human Sci.* 2005;24(4):483-6.
- Asicioglu F, Oguz-Savran F, Ozbek U. Mutation rate at commonly used forensic STR loci: paternity testing experience. *Dis; Markers.* 2004;20(6):313-5.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. (Eds.), *Current Protocols in Molecular Biology*, Wiley, New York, 2002.
- Axelsson J, Hoberg A-M, Waterson C, Myatt P, Shield GL, Varney J et al. Improved Reproducibility and Increased Signal Intensity in Matrix-assisted Laser Desorption/Ionization as a Result of Electrospray Sample Preparation. *Rapid Commun. Mass Spectrom.* 1997;11:209-13.
- Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res.* 2000;28(2):e8.

- Ballard DJ, Phillips C, Wright G, Thacker CR, Robson C, Revoir AP, Court DS. A study of mutation rates and the characterisation of intermediate, null and duplicated alleles for 13 Y chromosome STRs. *Forensic Sci. Int.* 2005;155(1):65-70.
- Baner J, Nilsson M, Mendel-Hartvig M, Landegren U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* 1998;26(22):5073-8.
- Bar W, Kratzer A, Machler M, Schmid W. Post-mortem stability of DNA. *Forensic Sci. Int.* 1988;39(1):59-70.
- Barany F. The ligase chain reaction in PCR world. *PCR Methods Appl.* 1991;1(1):5-16.
- Baron H, Fung S, Aydin A, Bahring S, Luft FC, Schuster H. Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. *Nat. Biotechnol.* 1996;14(10):1279-82.
- Beavis RC, Chait BT. Cinnamic acid derivatives as matrices for ultraviolet laser desorption Mass Spectrometry of proteins. *Rapid Commun. Mass Spectrom.* 1989;3:432-5.
- Beavis RC, Chaudhary T, Chait BT. Alpha – Cyano – 4 - hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* 1992;27:156-158.
- Beckman JS, Weber JL. Survey of human and rat microsatellites. *Genomics* 1992;12(4):627-31.
- Benjamin WH Jr, Smith KR, Waites KB. Ligase chain reaction. *Methods Mol. Biol.* 2003;226:135-50.
- Bennetts LE, Aitken RJ. A comparative study of oxidative DNA damage in mammalian spermatozoa. *Mol. Reprod. Dev.* 2005;71(1):77-87.
- Bergen AW, Pratt M, Mehlman PT, Goldman D. Evolution of RPS4Y. *Mol. Biol. Evol.* 1998;15(11):1412-9.
- Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD et al. An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* 1999;63(Pt 1):63-80.
- Berger B, Holzl G, Oberacher H, Niederstatter H, Huber CG, Parson W. Single nucleotide polymorphism genotyping by on-line liquid chromatography-mass spectrometry in forensic science of the Y-chromosomal locus M9. *J. Chromatogr. B* 2002;782(1-2):89-97.
- Berta P, Hawkins JR, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. Genetic evidence equating SRY and the testis-determining factor. *Nature* 1990;348(6300):448-50.

- Birus I, Marcikic M, Lauc D, Dzijan S, Lauc G. How high should paternity index be for reliable identification of war victims by DNA typing? *Croat. Med. J.* 2003;44(3):322-6.
- Boerwinkle E, Xiong WJ, Fourest E, Chan L. Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region. *Proc. Natl. Acad. Sci. USA* 1989;86(1):212-6.
- Bonilla C, Boxill LA, Donald SA, Williams T, Sylvester N, Parra EJ et al. The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Hum. Genet.* 2005;116(5):402-6.
- Borer PN, Dengler B, Tinoco I Jr, Uhlenbeck OC. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* 1974;86(4):843-53.
- Børsting C, Sanchez JJ, Morling N. Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray. *Int. J. Legal Med.* 2004;118(2):75-82.
- Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 1980;32(3):314-331.
- Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, Wallace DC. MITOMAP: a human mitochondrial genome database-2004 update. *Nucleic Acids Res.* 2005;33 Database Issue:D611-3.
- Braun A, Little DP, Koster H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin. Chem.* 1997;43(7):1151-8.
- Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. *Theor. Popul. Biol.* 2003;63(3):173-8.
- Breslauer KJ, Frank R, Blocker H, Marky LA. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* 1986;83(11):3746-50.
- Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* 1998;62(6):1408-15.
- Brion M, Cao R, Salas A, Lareu MV, Carracedo A. New method to measure minisatellite variant repeat variation in population genetic studies. *Am. J. Hum. Biol.* 2002;14(4):421-8.

- Brion M, Dupuy BM, Heinrich M, Hohoff C, Hoste B, Ludes B et al. A collaborative study of the EDNAP group regarding Y-chromosome binary polymorphism analysis. *Forensic Sci. Int.* 2005;153(2-3):103-8.
- Brion M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C et al. Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis* 2005;26(23):4411-20
- Brion M, Sobrino B, Blanco-Verea A, Lareu MV, Carracedo A. Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int. J. Legal. Med.* 2005;119(1):10-5.
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* 1998;63(3):861-9.
- Brookes AJ. The essence of SNPs. *Gene* 1999;234(2):177-86.
- Brown KA. Identifying the sex of human remains by ancient DNA analysis. *Anc. Biomol.* 2001;3:215-25.
- Brown RS, Lennon JJ. Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Anal. Chem.* 1995;67(13):1998-2003.
- Brown TA, Brown KA. Ancient DNA: using molecular biology to explore the past. *Bioessays.* 1994;16(10):719-26.
- Buckleton J, Triggs CM, Walsh SJ. *Forensic DNA Evidence Interpretation.* CRC Press, 2005.
- Budimlija ZM, Prinz MK, Zelson-Mundorff A, Wiersema J, Bartelink E, MacKinnon G et al. World Trade Center human identification project: experiences with individual body identification cases. *Croat. Med. J.* 2003;44(3):259-63.
- Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu. Rev. Genomics Hum. Genet.* 2003;4:119-41.
- Budowle B, Bieber FR, Eisenberg AJ. Forensic aspects of mass disasters: strategic considerations for DNA-based human identification. *Leg. Med. (Tokyo).* 2005;7(4):230-43.
- Budowle B, Gyllensten U, Chakraborty R, Allen M. Forensic analysis of the mitochondrial coding region and association to disease. *Int. J. Legal. Med.* 2005;119(5):314-5.
- Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians,

- Hispanics, Bahamians, Jamaicans and Trinidadians. *J. Forensic Sci.* 1999;44(6):1277–86.
- Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL. Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci. Int.* 1999;103(1):23-35.
- Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2001;98(2):581-4.
- Burger J, Hummel S, Hermann B, Henke W. DNA preservation: a microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 1999;20(8):1722-8.
- Butler JM, Levin BC. Forensic applications of mitochondrial DNA. *Trends Biotechnol.* 1998;16(4):158-62.
- Butler JM, Li J, Shaler TA, Monforte JA, Becker CH. Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *Int. J. Legal Med.* 1998;112(1):45–9.
- Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.* 2003;48(5):1054-64.
- Butler JM. *Forensic DNA typing: biology and technology behind STR markers.* London: Academic Press, 2001.
- Cai H, White PS, Torney D, Deshpande A, Wang Z, Keller RA et al. Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. *Genomics* 2000;66(2):135-43.
- Camire R, Ann Denchy R, Day GA 3rd, Lanzano P, Sheth S, Brown S. Prenatal diagnosis of factor X deficiency using a combination of direct mutation detection and linkage analysis with an intragenic single nucleotide polymorphism. *Prenat. Diagn.* 2003;23(6):457-60.
- Carracedo A, Sanchez-Diz P. Forensic DNA-typing technologies: a review. *Methods Mol. Biol.* 2005;297:1-12.
- Carvalho CM, Fujisawa M, Shirakawa T, Gotoh A, Kamidono S, Freitas Paulo T et al. Lack of association between Y chromosome haplogroups and male infertility in Japanese men. *Am. J. Med. Genet. A* 2003;116(2):152-8.

- Cashman JR, Zhang J, Leushner J, Braun A. Population distribution of human flavin-containing monooxygenase form 3: gene polymorphisms. *Drug Metab. Dispos.* 2001;29:1629–37.
- Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 1999;20(8):1682-96.
- Chaurand P, Luetzenkirchen F, Spengler B. Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* 1999;10(2):91-103.
- Chen J, Iannone MA, Li MS, Taylor JD, Rivers P, Nelsen AJ et al. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* 2000;10(4):549-57.
- Chen X, Kwok PY. Homogeneous genotyping assays for single nucleotide polymorphisms with fluorescence resonance energy transfer detection. *Genet. Anal.* 1999;14(5-6):157-63.
- Chen X, Levine L, Kwok PY. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* 1999;9(5):492-8.
- Chen X, Livak KJ, Kwok PY. A homogeneous ligase-mediated DNA diagnostic test. *Genome Res.* 1998;8(5):549-56.
- Chiapparino E, Lee D, Donini P. Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. *Genome* 2004;47(2):414-20.
- Ching A, Rafalski A. Rapid genetic mapping of ESTs using SNP pyrosequencing and indel analysis. *Cell. Mol. Biol. Lett.* 2002;7(2B):803-10.
- Clawson ML, Heaton MP, Chitko-McKown CG, Fox JM, Smith TP, Snelling WM et al. Beta-2-microglobulin haplotypes in U.S. beef cattle and association with failure of passive transfer in newborn calves. *Mamm. Genome* 2004;15(3):227-36.
- Clayton TM, Hill SM, Denton LA, Watson SK, Urquhart AJ. Primer binding site mutations affecting the typing of STR loci contained within the AMPFISTR SGM Plus kit. *Forensic Sci. Int.* 2004;139(2-3):255-9.
- Clegg RM. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.* 1992;211:353-88.
- Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA. *J. Forensic Sci.* 2005;50(1):43-53.

- Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int. J. Legal Med.* 2004;118(3):137-46.
- Coble MD, Vallone PM, Just RS, Diegoli TM, Smith BC, Parsons TJ. Effective strategies for forensic analysis in the mitochondrial DNA coding region. *Int. J. Legal Med.* 2005; disponible sur internet.
- Collins A, Ennis S, Taillon-Miller P, Kwok PY, Morton NE. Allelic association with SNPs: metrics, populations, and the linkage disequilibrium map. *Hum. Mutat.* 2001;17(4):255-62.
- Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 1998;8(12):1229-31.
- Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* 2004 Jun;12(6):495-504.
- Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* 1983;80(1):278-82.
- Cooper A, Poinar HN. Ancient DNA: do it right or not at all. *Science* 2000 Aug 18;289(5482):1139.
- Cooper A, Rambaut A, Macaulay V, Willerslev E, Hansen AJ, Stringer C. Human origins and ancient human DNA. *Science* 2001;292(5522):1655-6.
- Coquoz R. Preuve par l'ADN. La génétique au service de la justice. Presses polytechniques et universitaires romandes. Collection sciences forensiques, 2003.
- Cotton EA, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP et al. Validation of the AMPFISTR SGM plus system for use in forensic casework. *Forensic Sci. Int.* 2000;112(2-3):151-61.
- Croes EA, Alizadeh BZ, Bertoli-Avella AM, Rademaker T, Vergeer-Drop J, Dermaut B et al. Polymorphisms in the prion protein gene and in the doppel gene increase susceptibility for Creutzfeldt-Jakob disease. *Eur. J. Hum. Genet.* 2004;12(5):389-94.
- Crubezy E, Ludes B, Poveda JD, Clayton J, Crouau-Roy B, Montagnon D. Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C. R. Acad. Sci. III* 1998;321(11):941-51.

- de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C et al. Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Legal Med.* 1997;110(3):134-49.
- de Knijff P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* 2000;67(5):1055-61.
- De la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. *Mutat. Res.* 2005;573(1-2):111-35.
- de Laat A, van Tilburg M, van der Leun JC, van Vloten WA, de Gruijl FR. Cell cycle kinetics following UVA irradiation in comparison to UVB and UVC irradiation. *Photochem. Photobiol.* 1996;63(4):492-7.
- Delahunty C, Ankener W, Deng Q, Eng J, Nickerson DA. Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay. *Am. J. Hum. Genet.* 1996;58(6):1239-46.
- Detlaff-Kakol A, Pawlowski R. First Polish DNA "manhunt"-an application of Y-chromosome STRs. *Int. J. Legal Med.* 2002;116(5):289-91.
- Di Gaetano C, Crobu F, Guarrera S, Polidoro S, Gasparini M, Underhill PA et al. The TDI-FP assay in human Y chromosome SNP haplotyping. *Genet. Test.* 2004;8(4):400-3.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* 1994;91(8):3166-70.
- Ding C, Cantor CR. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS. *Proc. Natl. Acad. Sci. USA* 2003;100(6):3059-64.
- Divne AM, Allen M. A DNA microarray system for forensic SNP analysis. *Forensic Sci. Int.* 2005;154(2-3):111-21.
- Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci. Int.* 2005;154(1):62-77.
- Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K et al. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc. Natl. Acad. Sci. USA* 2000;97(19):10483-8.

- Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA et al. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* 1993;365(6446):566-8.
- Eglinton G, Logan GA. Molecular preservation. *Philos. Trans. R. Soc. Lond. Ser B* 1991;333 (1268) 315-27.
- Fei Z, Ono T, Smith LM. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res.* 1998;26(11):2827-8.
- Fei Z, Smith LM. Analysis of single nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2000;14(11):950-9.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989;246(4926):64-71.
- Fenselau C. MALDI MS and strategies for protein analysis. *Anal. Chem.* 1997;69:661A-5A.
- Filon D, Faerman M, Smith P, Oppenheim A. Sequence analysis reveals a beta-thalassaemia mutation in the DNA of skeletal remains from the archaeological site of Akhziv, Israel. *Nat. Genet.* 1995;9(4):365-8.
- Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, Perez JA et al. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur. J. Hum. Genet.* 2004;12(10):855-63.
- Fors L, Lieder KW, Vavra SH, Kwiatkowski RW. Large-scale SNP scoring from unamplified genomic DNA. *Pharmacogenomics* 2000;1(2):219-29.
- Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* 2000;67(1):182-96.
- Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R et al. Jefferson fathered slave's last child. *Nature* 1998;396(6706):27-8.
- Freeman SE, Hacham H, Gange RW, Maytum DJ, Sutherland JC, Sutherland BM. Wavelength dependence of pyrimidine dimer formation in DNA of human skin irradiated in situ with ultraviolet light. *Proc. Natl. Acad. Sci. USA* 1989;86(14):5605-9.
- Frudakis T, Venkateswarlu K, Thomas MJ, Gaskin Z, Ginjupalli S, Gunturi S et al. A classifier for the SNP-based inference of ancestry. *J. Forensic Sci.* 2003a;48(4):771-82.

- Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, Ginjupalli S et al. Sequences associated with human iris pigmentation. *Genetics* 2003b;165(4):2071-83.
- Fu DJ, Tang K, Braun A, Reuter D, Darnhofer-Demar B, Little DP et al. Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 1998;16(4):381-4.
- Furey TS, Haussler D. Integration of the cytogenetic map with the draft human genome sequence. *Hum. Mol. Genet.* 2003;12(9):1037-44.
- Gill P, Jeffreys A J, Werrett DJ. Forensic application of DNA 'fingerprints'. *Nature* 1985;318:577-9.
- Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA data-bases- joint considerations of the DNA working group of the ENFSI and the SWGDAM. *Sci. Justice.* 2004;44(1):51-3.
- Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int. J. Legal Med.* 2001;114(4-5):204-10.
- Gill P. Role of short tandem repeat DNA in forensic casework in the UK--past, present, and future perspectives. *Biotechniques* 2002;32(2):366-8.
- Giusti A, Baird M, Pasquale S, Balazs I, Glassberg J. Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J. Forensic Sci.* 1986;31(2):409-17.
- Glückmann M., Karas M. The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization. *J. Mass Spectrom.* 1999;34:467-77.
- Golenberg EM, Bickel A, Weihs P. Effect of highly fragmented DNA on PCR. *Nucleic Acids Res.* 1996;24(24):5026-33.
- Griffin TJ, Hall JG, Prudent JR, Smith LM. Direct genetic analysis by matrix-assisted laser desorption / ionization mass spectrometry. *Proc. Natl. Acad. Sci. USA* 1999;96(11):6301-6.
- Griffin TJ, Tang W, Smith LM. Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 1997;15(13):1368-72.
- Griffiths R, Double MC, Orr K, Dawson RJ. A DNA test to sex most birds. *Mol. Ecol.* 1998;7(8):1071-5.
- Grossman PD, Bloch W, Brinson E, Chang CC, Eggerding FA, Fung S et al. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res.* 1994;22(21):4527-34.

- Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, Kotzbach R. High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision. *Electrophoresis* 2003;24(7-8):1159-65.
- Grzybowski T. Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis* 2000;21(3):548-53.
- Guo B. Mass Spectrometry in DNA analysis. *Anal. Chem.* 1999;71(12):333R-7R.
- Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992;48(2):361-72.
- Gut IG. Automation in genotyping of single nucleotide polymorphisms. *Hum. Mutat.* 2001 Jun;17(6):475-92.
- Haff LA, Smirnov IP. Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res.* 1997;25(18):3749-50.
- Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res.* 1997;7(4):378-88.
- Hagelberg E, Sykes B, Hedges R. Ancient bone DNA amplified. *Nature* 1989;342(6249):485.
- Hagelberg E, Thomas MG, Cook CE Jr, Sher AV, Baryshnikov GF, Lister AM. DNA from ancient mammoth bones. *Nature* 1994;370(6488):333-4.
- Hall JG, Eis PS, Law SM, Reynaldo LP, Prudent JR, Marshall DJ et al. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl. Acad. Sci. USA* 2000;97(15):8272-7.
- Hall TA, Budowle B, Jiang Y, Blyn L, Eshoo M, Sannes-Lowery KA et al. Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. *Anal. Biochem.* 2005;344(1):53-69.
- Hammer MF. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.* 1994;11(5):749-61.
- Handt O, Krings M, Ward RH, Paabo S. The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.* 1996;59(2):368-76.
- Hanni C, Laudet V, Sakka M, Begue A, Stehelin D. Amplification of mitochondrial DNA fragments from ancient human teeth and bones. *C R Acad. Sci. III* 1990;310(9):365-70.
- Harris H, Hopkinson DA. *Handbook of enzyme electrophoresis in human genetics.* North-Holland Pub. Co., Amsterdam / New York, 1976.

- Hauge XY, Litt M. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.* 1993;2(4):411-5.
- Heaton MP, Chitko-McKnown CG, Grosse WM, Keele JW, Keen JE, Laegreid WW. Interleukin-8 haplotype structure from nucleotide sequence variation in commercial populations of U.S. beef cattle. *Mamm. Genome* 2001;12:219– 26.
- Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 2003;270(1512):313-21.
- Helmuth R, Fildes N, Blake E, Luce MC, Chimera J, Madej R et al. HLA-DQ alpha allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. *Am. J. Hum. Genet.* 1990;47(3):515-23.
- Hemminki K, Forsti A, Lorenzo Bermejo J. Single nucleotide polymorphisms (SNPs) are inherited from parents and they measure heritable events. *J. Carcinog.* 2005;4(1):2.
- Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* 1997;23(3):504-11.
- Henke J, Henke L. Mutation rate in human microsatellites. *Am. J. Hum. Genet.* 1999;64(5):1473-4.
- Hillenkamp F. Laser induced ion formation from organic solids, in *Ion formation from organic solids*, A. Benninghoven, Editor. 1983, Springer Verlag: Berlin. p. 190-205.
- Hoehe MR, Timmermann B, Lehrach H. Human inter-individual DNA sequence variation in candidate genes, drug targets, the importance of haplotypes and pharmacogenomics. *Curr. Pharm. Biotechnol.* 2003;4:351– 78.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S. Ancient DNA. *Nat. Rev. Genet.* 2001;2(5):353-9.
- Holland MM, Cave CA, Holland CA, Bille TW. Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks. *Croat. Med. J.* 2003;44(3):264-72.
- Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Sci. Rev.* 1999;11: 21–49.
- Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'->3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 1991;88(16):7276-80.
- Holliday R, Grigg GW. DNA methylation and mutation. *Mutat. Res.* 1993;285(1):61-7.

- Horai S, Hasayaka K, Murayama K, Wate N, Koike H, Nakai N. DNA amplification from ancient human skeletal remains and their sequence analysis. *Proc. Japan. Acad.* 1989;SerB(65):229-33.
- Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 1996;24(7):1304-7.
- Hoss M, Paabo S, Vereshchagin NK. Mammoth DNA sequences. *Nature* 1994;370(6488):333.
- Hoss M, Paabo S. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res.* 1993;21(16):3913-4.
- Hsu CM, Huang NE, Tsai LC, Kao LG, Chao CH, Linacre A, Lee JC. Identification of victims of the 1998 Taoyuan Airbus crash accident using DNA analysis. *Int. J. Legal Med.* 1999;113(1):43-6.
- Hsu TM, Law SM, Duan S, Neri BP, Kwok PY. Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clin. Chem.* 2001;47(8):1373-7.
- Huffine E, Crews J, Kennedy B, Bomberger K, Zinbo A. Mass identification of persons missing from the break-up of the former Yugoslavia: structure, function and role of the ICMP. *Croat. Med. J.* 2001;42(3):271-5.
- Hussein MR. Ultraviolet radiation and skin cancer: molecular mechanisms. *J. Cutan. Pathol.* 2005;32(3):191-205.
- Iannone MA, Taylor JD, Chen J, Li MS, Rivers P, Slentz-Kesler KA, Weiner MP. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* 2000;39(2):131-40.
- Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Imabayashi K et al. A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci. Int.* 2004;144(1):45-57.
- International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* 2001;409(6822):934-41.
- Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat. Genet.* 1996;12(4):417-20.

- Jaenicke-Despres V, Buckler ES, Smith BD, Gilbert MT, Cooper A, Doebley J, Paabo S. Early allelic selection in maize as revealed by ancient DNA. *Science* 2003;302(5648):1206-8.
- Jeffreys AJ, Brookfield JF, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* 1985c;317(6040):818-9.
- Jeffreys AJ, Murray J, Neumann R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell.* 1998;2(2):267-73.
- Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JA. Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.* 1994;6(2):136-45.
- Jeffreys AJ, Turner M, Debenham P. The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *Am. J. Hum. Genet.* 1991;48(5):824-40.
- Jeffreys AJ, Wilson V, Neumann R, Keyte J. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* 1988;16(23):10953-71.
- Jeffreys AJ, Wilson V, Thein SL, Weatherall DJ, Ponder BA. DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* 1986;39(1):11-24.
- Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature* 1985a;314:67-73.
- Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature* 1985b;316:76-79.
- Jobling M. 2000. The Y chromosome as a forensic tool: progress and prospects for the millennium. http://www.forensic.gov.uk/forensic/conference/papers/y_chromosome.
- Jobling MA, Heyer E, Dieltjes P, de Knijff P. Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1. *Hum. Mol. Genet.* 1999;8(11):2117-20.
- Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* 1997;110(3):118-24.
- Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 2003;4(8):598-612.

- Jobling MA. In the name of the father: surnames and genetics. *Trends Genet.* 2001;17(6):353-7.
- Jobling MA. Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci. Int.* 2001;118(2-3):158-62.
- Juhasz P, Roskey MT, Smirnov IP, Haff LA, Vestal ML, Martin SA. Applications of delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry to oligonucleotide analysis. *Anal. Chem.* 1996;68(6):941-6.
- Jurinke C, van den Boom D, Jacob A, Tang K, Worl R, Koster H. Analysis of ligase chain reaction products via matrix-assisted laser desorption/ionization time-of-flight-mass spectrometry. *Anal. Biochem.* 1996;237(2):174-81.
- Kaestle FA, Horsburgh KA. Ancient DNA in anthropology: methods, applications, and ethics. *Am. J. Phys. Anthropol.* 2002;Suppl 35:92-130.
- Kalinina O, Lebedeva I, Brown J, Silver J. Nanoliter scale PCR with TaqMan detection. *Nucleic Acids Res.* 1997;25(10):1999-2004.
- Kammerer S, Roth RB, Hoyal CR, Reneland R, Marnellos G, Kiechle M et al. Association of the NuMA region on chromosome 11q13 with breast cancer susceptibility. *Proc. Natl. Acad. Sci. USA* 2005;102(6):2004-9.
- Kanofsky JR, Sima P. Singlet oxygen production from the reactions of ozone with biological molecules. *J Biol Chem.* 1991;266(14):9039-42.
- Kanter E, Baird M, Shaler R, Balazs I. Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains. *J. Forensic Sci.* 1986;31(2):403-8.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS et al. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 2001;69(3):615-28.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J et al. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* 1999;64(3):817-31.
- Karas M, Bachmann D, Bahr U, Hullenkamp F. Matrix-Assisted Laser Desorption of Non-Volatile Compounds. *Int. J. Mass Spectrom. Ion Processes.* 1987;78:53-68.
- Karas M, Gluckmann M, Schafer J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* 2000;35(1):1-12.

- Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular mass exceeding 10,000 Daltons. *Anal. Chem.* 1988;60:2299-301.
- Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Stoneking M. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* 2001;68(1):173-190.
- Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* 1997;110(3):125-33, 141-9.
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S et al. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 2000;66(5):1580-8.
- Kayser M, Sajantila A. Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci. Int.* 2001;118(2-3):116-21.
- Ke Y, Su B, Song X, Lu D, Chen L, Li H et al. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 2001;292(5519):1151-3.
- Kelman Z, Moran L. Degradation of ancient DNA. *Curr Biol.* 1996;6(3):223.
- Keyser-Tracqui C, Blandin P, Ricaut FX, Petkovski E, Crubézy E, Ludes B. Does the Tat polymorphism originate in northern Mongolia? *Progress in forensic genetics 10; International Congress Series 1261 (2004) 325– 327.*
- Keyser-Tracqui C, Crubezy E, Ludes B. Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. *Am. J. Hum. Genet.* 2003;73(2):247-60.
- Kidd LC, Paltoo DN, Wang S, Chen W, Akereyeni F, Isaacs W et al. Sequence variation within the 5' regulatory regions of the vitamin D binding protein and receptor genes and prostate cancer risk. *Prostate* 2005;64(3):272-82.
- Kim JG, Nonneman D, Rohrer GA, Vallet JL, Christenson RK. Linkage mapping of a SNP in the porcine MADH1 gene to a region of chromosome 8 that contains QTL for uterine capacity. *Anim. Genet.* 2003;34:310–1.
- Kim S, Shi S, Bonome T, Ulz ME, Edwards JR, Fodstad H et al. Multiplex genotyping of the human beta2-adrenergic receptor gene using solid-phase capturable dideoxynucleotides and mass spectrometry. *Anal. Biochem.* 2003;316(2):251-8.

- Kline MC, Vallone PM, Redman JW, Duewer DL, Calloway CD, Butler JM. Mitochondrial DNA typing screens with control region and coding region SNPs. *J. Forensic Sci.* 2005;50(2):377-85.
- Knochenmuss R, Zenobi R. MALDI ionization: the role of in-plume processes. *Chem. Rev.* 2003;103(2):441-52.
- Knowlton RG, Brown VA, Braman JC, Barker D, Schumm JW, Murray C et al. Use of highly polymorphic DNA probes for genotypic analysis following bone marrow transplantation. *Blood* 1986;68(2):378-85.
- Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC. MITOMAP: a human mitochondrial genome database. *Nucleic Acids Res.* 1996;24(1):177-9.
- Komar D. Lessons from Srebrenica: the contributions and limitations of physical anthropology in identifying victims of war crimes. *J. Forensic Sci.* 2003;48(4):713-6.
- Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 2003;21(1):12-27.
- Kostrikis LG, Tyagi S, Mhlanga MM, Ho DD, Kramer FR. Spectral genotyping of human alleles. *Science* 1998;279(5354):1228-9.
- Kostrzewa M, Elssner T, Fahr K, Wenzel T. 50th ASMS Conference on Mass Spectrometry, Orlando, FL, 2002; <http://www.inmerged.com/aspfolder/ASMSarcresults.asp>.
- Koyama H, Iwasa M, Tsuchimochi T, Maeno Y, Isobe I, Matsumoto T, Nagao M. Utility of Y-STR haplotype and mtDNA sequence in personal identification of human remains. *Am. J. Forensic Med. Pathol.* 2002;23(2):181-5.
- Krebs S, Medugorac I, Seichter D, Forster M. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res.* 2003;31(7):e37.
- Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ et al. Validation of a 16-locus fluorescent multiplex system. *J. Forensic Sci.* 2002;47(4):773-85.
- Kusmann M, Nordhoff E, Rabek-Nielsen H, Haebel S, Rossel-Larsen M, Jakobsen L et al. Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation techniques Designed for Various Peptide and Protein Analytes. *J. Mass Spectrom.*, 1997;32:593-601.
- Kwok PY, Deng Q, Zakeri H, Taylor SL, Nickerson DA. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 1996;31(1):123-6.
- Kwok PY. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2001;2:235-58.

- Kwok PY. SNP genotyping with fluorescence polarization detection. *Hum. Mutat.* 2002;19(4):315-23.
- Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* 1990;18(4):999-1005.
- Ladika S. DNA forensics. Laying ghosts to rest in Bosnia. *Science* 2001;293(5534):1422-3.
- Ladika S. South Asia tsunami. DNA helps identify missing in the tsunami zone. *Science* 2005;307(5709):504.
- Laegreid WW, Heaton MP, Keen JE, Grosse WM, Chitko-McKown CG, Smith TP et al. Association of bovine neonatal Fc receptor alpha-chain gene (FCGRT) haplotypes with serum IgG concentration in newborn calves. *Mamm. Genome.* 2002;13(12):704-10.
- Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, Aula P, Sajantila A. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* 1999;7(4):447-58.
- Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. *Science* 1988;241(4869):1077-80.
- Landegren U, Samiotaki M, Nilsson M, Malmgren H, Kwiatkowski M. Detecting genes with ligases. *Methods.* 1996;9(1):84-90.
- Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralbl. Bakteriol.* 1900;27:357-62.
- Lattard V, Zhang J, Tran Q, Furnes B, Schlenk D, Cashman JR. Two new polymorphisms of the FMO3 gene in Caucasian and African-American populations: comparative genetic and functional studies. *Drug. Metab. Dispos.* 2003;31:854-60.
- Lavebratt C, Sengul S, Jansson M, Schalling M. Pyrosequencing-based SNP allele frequency estimation in DNA pools. *Hum. Mutat.* 2004;23(1):92-7.
- Lazaruk K, Wallin J, Holt C, Nguyen T, Walsh PS. Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. *Forensic Sci. Int.* 2001;119(1):1-10.
- Le Hellard S, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, Morris SW et al. SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res.* 2002;30(15):e74.

- Leclair B, Fregeau CJ, Bowen KL, Fourney RM. Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. *J. Forensic Sci.* 2004;49(5):939-53.
- Ledford M, Friedman KD, Hessner MJ, Moehlenkamp C, Williams TM, Larson RS. A multi-site study for detection of the factor V (Leiden) mutation from genomic DNA using a homogeneous invader microtiter plate fluorescence resonance energy transfer (FRET) assay. *J. Mol. Diagn.* 2000;2(2):97-104.
- Lee LG, Livak KJ, Mullah B, Graham RJ, Vinayak RS, Woudenberg TM. Seven-color, homogeneous detection of six PCR products. *Biotechniques* 1999;27(2):342-9.
- Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G et al. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. *Forensic Sci. Int.* 2003;133(3):220-7.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG et al. The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* 2002;70(1):192-206.
- Leonard JA, Wayne RK, Cooper A. Population genetics of ice age brown bears. *Proc. Natl. Acad. Sci. USA* 2000;97(4):1651-4.
- Leopoldino AM, Pena SD. The mutational spectrum of human autosomal tetranucleotide microsatellites. *Hum. Mutat.* 2003;21(1):71-9.
- Lessig R, Zoledziewska M, Fahr K, Edelmann J, Kostrzewa M, Dobosz T, Kleemann WJ. Y-SNP-genotyping - a new approach in forensic analysis. *Forensic Sci. Int.* 2005;154(2-3):128-36.
- Leushner J, Chiu NH. Automated mass spectrometry: a revolutionary technology for clinical diagnostics. *Mol. Diagn.* 2000;5(4):341-8.
- Levene H. On a matching problem arising in genetics. *Ann. Math. Stat.* 1949;20:91-4.
- Leykin I, Hao K, Cheng J, Meyer N, Pollak MR, Smith RJ et al. Comparative linkage analysis and visualization of high-density oligonucleotide SNP array data. *BMC Genet.* 2005;6(1):7.
- Li J, Butler JM, Tan Y, Lin H, Royer S, Ohler L et al. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 1999;20(6):1258-65.
- Liljedahl U, Fredriksson M, Dahlgren A, Syvanen AC. Detecting imbalanced expression of SNP alleles by minisequencing on microarrays. *BMC Biotechnol.* 2004;4(1):24.

- Lindahl T. Instability and decay of the primary structure of DNA. *Nature* 1993;362(6422):709-15.
- Lins AM, Micka KA, Sprecher CJ, Taylor JA, Bacher JW, Rabbach DR et al. Development and population study of an eight-locus short tandem repeat (STR) multiplex system. *J. Forensic Sci.* 1998;43(6):1168-80.
- Litt M, Hauge X, Sharma V. Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *Biotechniques* 1993;15(2):280-4.
- Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 1989;44(3):397-401.
- Little DP, Braun A, Darnhofer-Demar B, Koster H. Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry. *Eur. J. Clin. Chem. Clin. Biochem.* 1997;35(7):545-8.
- Liu YH, Bai J, Zhu Y, Liang X, Siemieniak D, Venta PJ, Lubman DM. Rapid screening of genetic polymorphisms using buccal cell DNA with detection by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 1995;9(9):735-43.
- Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.* 1995;4(6):357-62.
- Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet. Anal.* 1999;14(5-6):143-9.
- Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 1998;19(3):225-32.
- Ludes B, Mangin P. (1992) *Les empreintes génétiques en médecine légale* (Coll. G2). Editions Médicales Internationales.
- Ludes B, Tracqui A, Pfitzinger H, Kintz P, Levy F, Disteldorf M et al. Medico-legal investigations of the Airbus, A320 crash upon Mount Ste-Odile, France. *J. Forensic Sci.* 1994;39(5):1147-52.
- Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* 1999;17(3):292-6.

- Lyamichev V, Neri B. Invader assay for SNP genotyping. *Methods Mol Biol.* 2003;212:229-40.
- Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM et al. The validation of short tandem repeat (STR) loci for use in forensic casework. *Int. J. Legal Med.* 1994;107(2):77-89.
- Mamyrin BA, Shmikk DV. The linear mass reflectron. *Sov. Phys. JETP*, 1979. 49: p. 762.
- Manen JF, Bouby L, Dalnoki O, Marinval P, Turgay M, Schlumbaum A. Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J. Archaeo. Sci.* 2003;30:721-9.
- Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. *J. Clin. Lab. Anal.* 2002;16(1):47-51.
- Marota I, Basile C, Ubaldi M, Rollo F. DNA decay rate in papyri and human remains from Egyptian archaeological sites. *Am. J. Phys. Anthropol.* 2002;117(4):310-8.
- Marras SA, Kramer FR, Tyagi S. Multiplex detection of single-nucleotide variations using molecular beacons. *Genet. Anal.* 1999;14(5-6):151-6.
- Marth GT, Czabarka E, Murvai J, Sherry ST. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 2004;166(1):351-72.
- Martinez-Cruz B, Godoy JA, Negro JJ. Population genetics after fragmentation: the case of the endangered Spanish imperial eagle. *Mol. Ecol.* 2004;13(8):2243-55
- Marx I (2000) The Xiongnu culture - third century BCE. <http://www.silk-road.com/toc/index.html>
- McCarthy JJ, Hilfiker R. The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat. Biotechnol.* 2000;18(5):505-8.
- Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L et al. Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res.* 2000;10(3):330-43.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome SNPs. *Anal. Chem.* 2004;76(20):6039-45.
- Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. Typing of multiple single-nucleotide polymorphisms using ribonuclease cleavage of DNA/RNA chimeric single-base extension primers and detection by MALDI-TOF mass spectrometry. *Anal. Chem.* 2005;77(16):5229-35.

- Mitchell D, Willerslev E, Hansen A. Damage and repair of ancient DNA. *Mutat. Res.* 2005;571(1-2):265-76.
- Mornhinweg E, Luckenbach C, Fimmers R, Ritter H. D3S1358: sequence analysis and gene frequency in a German population. *Forensic Sci. Int.* 1998;95(2):173-8.
- Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA. SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* 2002;12:1766-72.
- Mourant AE, Kopec AC, Domaniewska-Sobczak K. The distribution of the human blood groups and other polymorphisms; Oxford Univ. Press, London. 1976.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold. Spring Harb. Symp. Quant. Biol.* 1986;51(Pt 1):263-73.
- Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 1987;155:335-50.
- Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000;156(1):297-304.
- Nakahori Y, Hamano K, Iwaya M, Nakagome Y. Sex identification by polymerase chain reaction using X-Y homologous primer. *Am. J. Med. Genet.* 1991;39(4):472-3.
- Nakahori Y, Takenaka O, Nakagome Y. A human X-Y homologous region encodes "amelogenin". *Genomics* 1991;9(2):264-9.
- Nakamura Y, Carlson M, Krapcho K, Kanamori M, White R. New approach for isolation of VNTR markers. *Am. J. Hum. Genet.* 1988;43(6):854-9.
- Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987;235(4796):1616-22.
- Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 1979;76(10):5269-73.
- Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 1989;17(7):2503-16.
- Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.* 1998;19(3):233-40.

- Nie B, Shortreed MR, Smith LM. Scoring single-nucleotide polymorphisms at the single-molecule level by counting individual DNA cleavage events on surfaces. *Anal. Chem.* 2005;77(20):6594-600.
- Niederstatter H, Coble MD, Grubwieser P, Parsons TJ, Parson W. Characterization of mtDNA SNP typing and mixture ratio assessment with simultaneous real-time PCR quantification of both allelic states. *Int. J. Legal Med.* 2005; disponible sur internet.
- Nielsen PE. PNA Technology. *Mol; Biotechnol.* 2004;26(3):233-48.
- Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 2000;154(2):931-42.
- Nielsen R. Population genetic analysis of ascertained SNP data. *Hum. Genomics* 2004;1(3):218-24.
- Nilsson M, Krejci K, Koch J, Kwiatkowski M, Gustavsson P, Landegren U. Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. *Nat. Genet.* 1997;16(3):252-5.
- Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 1994;265(5181):2085-8.
- Nonneman DJ, Rohrer GA. Comparative mapping of a region on chromosome 10 containing QTL for reproduction in swine. *Anim. Genet.* 2003;34:42-6.
- Olaisen B, Stenersen M, Mevag B. Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. *Nat. Genet.* 1997;15(4):402-5.
- Oota H, Saitou N, Matsushita T, Ueda S. A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. *Am. J. Phys. Anthropol.* 1995;98(2):133-45.
- Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N et al. Genetic analyses from ancient DNA. *Annu. Rev. Genet.* 2004;38:645-79.
- Paabo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. USA* 1989;86(6):1939-43.
- Pakendorf B, Morar B, Tarskaia LA, Kayser M, Soodyall H, Rodewald A, Stoneking M. Y-chromosomal evidence for a strong reduction in male population size of Yakuts. *Hum. Genet.* 2002;110(2):198-200.
- Panzer S, Kuhl DP, Caskey CT. Unstable triplet repeat sequences: a source of cancer mutations? *Stem Cells.* 1995;13(2):146-57.

- Parson W, Brandstatter A, Alonso A, Brandt N, Brinkmann B, Carracedo A et al. The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci. Int.* 2004;139(2-3):215-26.
- Parson W, Niederstätter H, Brandstätter A, Berger B. Improved specificity of Y-STR typing in DNA mixture samples. *Int. J. Legal Med.* 2003;117:109–114.
- Parson W, Pegoraro K, Niederstätter H, Foger M, Steinlechner M. Species identification by means of the cytochrome b gene. *Int. J. Legal Med.* 2000;114(1-2):23-8.
- Parsons BL, Heflich RH. Genotypic selection methods for the direct analysis of point mutations. *Mutat. Res.* 1997;387(2):97-121.
- Pascal O. Empreintes génétiques : pourquoi et pour qui ? *Med. et Droit* 1998;32:1-6.
- Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* 2000;10(7):1031-42.
- Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 1988;85(8):2444-8.
- Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA* 1994;91(11):5022-6.
- Petes TD. Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* 2001;2(5):360-9.
- Petkovski E, Keyser-Tracqui C, Crubézy E, Hienne R, Ludes B. MALDI-TOF MS analysis of Y-SNPs in ancient samples. *Progress in Forensic Genetics 11; ICS 2006; sous presse.*
- Petkovski E, Keyser-Tracqui C, Hienne R, Ludes B. SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. *J. Forensic Sci.* 2005;50(3):535-41.
- Phillips C, Lareu M, Salas A, Fondevilla M, Berniel Lee G, Carracedo A et al. Population specific single nucleotides polymorphism. *Progress in Forensic Genetics 10; ICS 2004;1261:233-5.*
- Pieles U, Zurcher W, Schär M, Moser HE. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry : a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* 1993;21(14): 3191-6.
- Poinar HN, Hoss M, Bada JL, Paabo S. Amino acid racemization and the preservation of ancient DNA. *Science* 1996;272(5263):864-6.

- Posthumus MA, Kistemaker PG, Meuzelaar HLC, Ten Noever de Brauw MC. Laser desorption-mass spectrometry of polar nonvolatile bio-organic molecules. *Anal. Chem.* 1978;50(7):985-91.
- Pourzand C, Cerutti P. Genotypic mutation analysis by RFLP/PCR. *Mutat. Res.* 1993;288(1):113-21.
- Primorac D, Andelinovic S, Definis-Gojanovic M, Drmic I, Rezic B, Baden MM et al. Identification of war victims from mass graves in Croatia, Bosnia, and Herzegovina by use of standard forensic methods and DNA typing. *J. Forensic Sci.* 1996;41(5):891-4.
- Pusch W, Wurmbach JH, Thiele H, Kostrzewa M. MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 2002;3(4):537-48.
- Quaife R, Wong LF, Tan SY, Chua WY, Lim SS, Hammersley CJ, Yeo HL. QF-PCR-based prenatal detection of aneuploidy in a southeast Asian population. *Prenat. Diagn.* 2004;24(6):407-13.
- Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci. Int.* 2004;140(2-3):251-7.
- Rao KV, Stevens PW, Hall JG, Lyamichev V, Neri BP, Kelso DM. Genotyping SNPs directly from genomic DNA by invasive cleavage reaction on microspheres. *Nucleic Acids Res.* 2003;31(11):e66.
- Raoult D, Aboudharam G, Crubezy E, Larrouy G, Ludes B, Drancourt M. Molecular identification by "suicide PCR" of *Yersinia pestis* as the agent of medieval black death. *Proc. Natl. Acad. Sci. USA* 2000;97(23):12800-3.
- Raymond M, Rousset F. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Heredity* 1995;86:248-9.
- Read CM, Baldwin JP, Crane-Robinson C. Structure of subnucleosomal particles. Tetrameric (H3/H4)₂ 146 base pair DNA and hexameric (H3/H4)₂(H2A/H2B)₁ 146 base pair DNA complexes. *Biochemistry* 1985;24(16):4435-50.
- Reed FA, Kontanis EJ, Kennedy KA, Aquadro CF. Brief communication: ancient DNA prospects from Sri Lankan highland dry caves support an emerging global pattern. *Am. J. Phys. Anthropol.* 2003;121(2):112-6.
- Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM et al. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.* 2002;32(1):135-42.

- Ricaud FX, Keyser-Tracqui C, Bourgeois J, Crubezy E, Ludes B. Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. *Hum. Biol.* 2004;76(1):109-25.
- Richards M, Macaulay V. The mitochondrial gene tree comes of age. *Am. J. Hum. Genet.* 2001;68(6):1315-20.
- Roewer L, Epplen JT. Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci. Int.* 1992;53:163.
- Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglia A et al. A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci. Int.* 2000;114(1):31-43.
- Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A et al. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci. Int.* 2001;118(2-3):106-13.
- Roffey PE, Eckhoff CI, Kuhl JL. A rare mutation in the amelogenin gene and its potential investigative ramifications. *J. Forensic Sci.* 2000;45(5):1016-9.
- Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R. Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. *Int. J. Legal Med.* 2001;115(1):12-5.
- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 1996;242(1):84-9.
- Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 2001;11(1):3-11.
- Ross P, Hall L, Smirnov I, Haff L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 1998;16(13):1347-51.
- Ross PL, Lee K, Belgrader P. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal. Chem.* 1997;69(20):4197-202.
- Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics.* 1988;3(4):352-60.
- Ryan D, Nuccie B, Arvan D. Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. *Mol. Diagn.* 1999;4(2):135-44.

- Rychlik W, Rhoads RE. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* 1989;17(21):8543-51.
- Rychlik W, Spencer WJ, Rhoads RE. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* 1990;18(21):6409-12.
- Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 1986;324(6093):163-6.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;239(4839):487-91.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA. Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230(4732):1350-4.
- Samiotaki M, Kwiatkowski M, Parik J, Landegren U. Dual-color detection of DNA sequence variants by ligase-mediated analysis. *Genomics* 1994;20(2):238-42.
- Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N. Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs. *Forensic Sci. Int.* 2003;137(1):74-84.
- Sanchez JJ, Borsting C, Morling N. Typing of Y chromosome SNPs with multiplex PCR methods. *Methods Mol. Biol.* 2005;297:209-28.
- Santos FR, Epplen JT, Pena SD. Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. *EXS* 1993;67:261-5.
- Santos FR, Pandya A, Tyler-Smith C. Reliability of DNA-based sex tests. *Nat. Genet.* 1998;18(2):103.
- Sauer S, Lange BM, Gobom J, Nyarsik L, Seitz H, Lehrach H. Miniaturization in functional genomics and proteomics. *Nat. Rev. Genet.* 2005;6(6):465-76.
- Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res.* 2000a;28(5):e13.
- Sauer S, Lechner D, Berlin K, Plancon C, Heuermann A, Lehrach H, Gut IG. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res.* 2000b;28(23):e100.

- Sauer S. Typing of single nucleotide polymorphisms by MALDI mass spectrometry: Principles and diagnostic applications. *Clin. Chim. Acta* 2006;363(1-2):95-105.
- Schatz P, Dietrich D, Schuster M. Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. *Nucleic Acids Res.* 2004;32(21):e167.
- Schraml E, Daxberger H, Watzinger F, Lion T. Quantitative analysis of chimerism after allogeneic stem cell transplantation by PCR amplification of microsatellite markers and capillary electrophoresis with fluorescence detection: the Vienna experience. *Leukemia* 2003;17(1):224-7.
- Schultes T, Hummel S, Herrmann B. Amplification of Y-chromosomal STRs from ancient skeletal material. *Hum. Genet.* 1999;104(2):164-6.
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am. J. Hum. Genet.* 2003;73(3):700-5.
- Shahgholi M, Garcia BA, Chiu NH, Heaney PJ, Tang K. Sugar additives for MALDI matrices improve signal allowing the smallest nucleotide change (A:T) in a DNA sequence to be resolved. *Nucleic Acids Res.* 2001 ;29(19):e91.
- Shawky RM, el-Aleem KA, Rifaat MM, el-Naggat RL, Marzouk GM. Rapid carrier screening using short tandem repeats in the phenylalanine hydroxylase gene. *East Mediterr. Health J.* 2002;8(1):49-54.
- Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. *Hum. Mutat.* 1996;7(4):346-54.
- Sibille I, Duverneuil C, Lorin de la Grandmaison G, Guerrouache K, Teissiere F, Durigon M, de Mazancourt P. Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa. *Forensic Sci. Int.* 2002;125(2-3):212-6.
- Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD et al. Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes* 2004;53(4):1141-9.
- Sinden RR, Potaman VN, Oussatcheva EA, Pearson CE, Lyubchenko YL, Shlyakhtenko LS. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci.* 2002;27(Suppl 1):53.
- Smith JC, Anwar R, Riley J, Jenner D, Markham AF, Jeffreys AJ. Highly polymorphic minisatellite sequences: allele frequencies and mutation rates for five locus-specific probes in a Caucasian population. *J. Forensic Sci. Soc.* 1990;30(1):19-32.

- Sobrinho B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. Int.* 2005;154(2-3):181-194.
- Sommer R, Tautz D. Minimal homology requirements for PCR primers. *Nucleic Acids Res.* 1989;17(16):6749.
- Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 1975;98(3):503-17.
- Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N et al. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res.* 2004;14(1):126-33.
- Steinlechner M, Berger B, Niederstatter H, Parson W. Rare failures in the amelogenin sex test. *Int. J. Legal Med.* 2002;116(2):117-20.
- Stomakhin AA, Vasiliskov VA, Timofeev E, Schulga D, Cotter RJ, Mirzabekov AD. DNA sequence analysis by hybridization with oligonucleotide microchips: MALDI mass spectrometry identification of 5mers contiguously stacked to microchip oligonucleotides. *Nucleic Acids Res.* 2000 ;28(5):1193-8.
- Strupat K, Karas M, Hillenkamp F. 2,5-Hydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* 1992;27:156-8.
- Sturm RA, Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. *Trends Genet.* 2004;20(8):327-32.
- Suggs SV, Hirose T, Miyake EH, Kawashima MJ, Johnson KI, Wallace RB. In D.D. Brown (ed.), *ICN-UCLA Symp. Dev. Biol. Using Purified Genes*. Academic Press Inc., New York, 1981;23:683-93.
- Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* 1993;15:636-41.
- Suzuki M, Muto S, Hara K, Ozeki T, Yamada Y, Kadowaki T et al. Single-nucleotide polymorphisms in the 17beta-hydroxysteroid dehydrogenase genes might predict the risk of side-effects of estramustine phosphate sodium in prostate cancer patients. *Int. J. Urol.* 2005;12(2):166-72.
- Sykes B, Irvén C. Surnames and the Y chromosome. *Am. J. Hum. Genet.* 2000;66(4):1417-9.
- Syvanen AC. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* 1999;13(1):1-10.

- Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D. Use of X-linked markers for forensic purposes. *Int. J. Legal. Med.* 2003;117(2):67-74.
- Szibor R, Lautsch S, Plate I, Bender K, Krause D. Population genetic data of the STR HumD3S1358 in two regions of Germany. *Int. J. Legal Med.* 1998;111(3):160-1.
- Taillon-Miller P, Kwok PY. A high-density single-nucleotide polymorphism map of Xq25-q28. *Genomics* 2000;65(3):195-202.
- Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL et al. The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* 2004;74(4):661-82.
- Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T. Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry *Rapid Commun. Mass Spectrom.* 1988;2:151-2.
- Tang K, Allman SL, Chen CH, Chang LY, Schell M. Matrix-assisted Laser Desorption/Ionization of restriction Enzyme-Digested DNA. *Rapid Commun. Mass Spectrom.* 1994;8:183-6.
- Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 1989;17(16):6463-71.
- Tautz D. Notes on the definition and nomenclature of tandemly repetitive DNA sequences. *EXS* 1993;67:21-8.
- Thomas WK, Paabo S. DNA sequences from old tissue remains. *Methods Enzymol.* 1993;224:406-19.
- Thomson JA, Pilotti V, Stevens P, Ayres KL, Debenham PG. Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. *Forensic Sci. Int.* 1999;100(1-2):1-16.
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* 2000;97(13):7360-5.
- Thorisson GA, Stein LD. The SNP Consortium website: past, present and future. *Nucleic Acids Res.* 2003;31(1):124-7.
- Tost J, Brandt O, Boussicault F, Derbala D, Caloustian C, Lechner D, Gut IG. Molecular haplotyping at high throughput. *Nucleic Acids Res.* 2002;30(19):e96.
- Tost J, Gut IG. Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin. Biochem.* 2005;38(4):335-50.

- Tost J, Gut IG. Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.* 2002;21(6):388-418.
- Tost J, Schatz P, Schuster M, Berlin K, Gut IG. Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* 2003;31(9):e50.
- Trumme T, Herrmann B, Hummel S. Genetics in genealogical research--reconstruction of a family tree by means of Y-haplotyping. *Anthropol. Anz.* 2004;62(4):379-86.
- Tsukada K, Takayanagi K, Asamura H, Ota M, Fukushima H. Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. *Legal Med.* 2002;4:239-45.
- Tully G, Bar W, Brinkmann B, Carracedo A, Gill P, Morling N et al. Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci. Int.* 2001;124(1):83-91.
- Tully G, Barritt SM, Bender K, Brignon E, Capelli C, Dimo-Simonin N et al. Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. *Forensic Sci. Int.* 2004;140(1):1-11.
- Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* 1998;16(1):49-53.
- Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 1996;14(3):303-8.
- Tyagi S, Marras SA, Kramer FR. Wavelength-shifting molecular beacons. *Nat. Biotechnol.* 2000;18(11):1191-6.
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 1997;7(10):996-1005.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 2001;65(Pt 1):43-62.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH et al. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 2000;26(3):358-61.
- Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Legal Med.* 1994;107(1):13-20.

- Vallone PM, Fahr K, Kostrzewa M. Genotyping SNPs using a UV-photocleavable oligonucleotide in MALDI-TOF MS. *Methods Mol. Biol.* 2005;297:169-78.
- Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ. A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int. J. Legal Med.* 2004;118(3):147-57.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al. The sequence of the human genome. *Science* 2001;291(5507):1304-51.
- Vergnaud G. Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* 1989;17(19):7623-30.
- von Wintzingerode F, Bocker S, Schlotelburg C, Chiu NH, Storm N, Jurinke C et al. Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. *Proc. Natl. Acad. Sci. USA* 2002 ;99(10):7039-44.
- Vorm O, Roepstorff P, Mann M. Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* 1994;66(19): 3281-7.
- Vouillamoz J, Maigre D, Meredith CP. Microsatellite analysis of ancient alpine grape cultivars: pedigree reconstruction of *Vitis vinifera* L. 'Cornalin du Valais'. *Theor. Appl. Genet.* 2003;107(3):448-54.
- Walker JA, Hedges DJ, Perodeau BP, Landry KE, Stoilova N, Laborde ME et al. Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: application in human identification. *Anal. Biochem.* 2005;337(1):89-97.
- Wallace DC. Mitochondrial DNA mutations and neuromuscular disease. *Trends Genet.* 1989;5(1):9-13.
- Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res.* 1979;6(11):3543-57.
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998;280(5366):1077-82.
- Watanobe T, Ishiguro N, Okumura N, Nakano M, Matsui A, Hongo H, Ushiro H. Ancient mitochondrial DNA reveals the origin of *Sus scrofa* from Rebun Island, Japan. *J. Mol. Evol.* 2001;52(3):281-9.

- Weber JL, Wong C. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 1993;2(8):1123-8.
- Wenzel T, Ellsner T, Fahr K, Bimmler J, Richter S, Thomas I, Kostrzewa M. Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides. *Nucleosides Nucleotides Nucleic Acids.* 2003;22(5-8):1579-81.
- Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. *Biotechniques* 1995;18(4):670-7.
- Whitaker JP, Cotton EA, Gill P. A comparison of the characteristics of profiles produced with the AMPFISTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci. Int.* 2001;123(2-3):215-23.
- Whitfield LS, Sulston JE, Goodfellow PN. Sequence variation of the human Y chromosome. *Nature* 1995;378(6555):379-80.
- Whittall RM, Li L. High-Resolution Matrix-Assisted Laser Desorption/Ionization in a Linear Time-of-Flight Mass Spectrometer. *Anal. Chem.* 1995; 67(13):1950-4.
- Wiegand P, Kleiber M. Less is more-length reduction of STR amplicons using redesigned primers. *Int. J. Legal. Med.* 2001;114(4-5):285-7.
- Wilkins Stevens P, Hall JG, Lyamichev V, Neri BP, Lu M, Wang L et al. Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions. *Nucleic Acids Res.* 2001;29(16):e77
- Wise CA, Paris M, Morar B, Wang W, Kalaydjieva L, Bittles AH. A standard protocol for single nucleotide primer extension in the human genome using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003;17(11):1195-202.
- Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ. The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* 1997;22(1):176-81.
- Wong KK, Tsang YT, Shen J, Cheng RS, Chang YM, Man TK, Lau CC. Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA. *Nucleic Acids Res.* 2004;32(9):e69.
- Wong Z, Wilson V, Jeffreys AJ, Thein SL. Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* 1986;14(11):4605-16.

- Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* 1987;51(Pt 4):269-88.
- Wu KJ, Shaler TA, Becker CH. Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. *Anal. Chem.* 1994;66:1637-45.
- Wu KJ, Steding A, Becker CH. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun. Mass Spectrom.* 1993;7(2):142-6.
- Wyman AR, White R. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* 1980;77(11):6754-8.
- Xiang F, Beavis RC. Growing Protein-doped Sinapinic Acid Crystals for Laser desorption: An alternative Preparation Method for Difficult Samples. *Org. Mass Spectrom.* 1993;28:1424-9.
- Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 2002;12(2):339-48.
- Yakes FM, Van Houten B. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc. Natl. Acad. Sci. USA* 1997;94(2):514-9.
- Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* 2001;29(17):e88-8.
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR et al. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* 1997;60(5):1174-83.
- Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* 2003;312:207-13.
- Zhigilei LV, Leveugle E, Garrison BJ, Yingling YG, Zeifman MI. Computer simulations of laser ablation of molecular substrates. *Chem. Rev.* 2003;103(2):321-48.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T et al. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 2004;74(1):50-61.
- Zhou L, Deng H, Deng Q, Zhao S. A mixed matrix of 3-hydroxypicolinic acid and pyrazinecarboxylic acid for matrix-assisted laser desorption/ionization time-of-flight

- mass spectrometry of oligodeoxynucleotides. *Rapid Commun. Mass Spectrom.* 2004;18(7):787-94.
- Zhu L, Parr GR, Fitzgerald MC, Nelson CM, Smith LM. Oligodeoxynucleotide Fragmentation in MALDI/TOF Mass Spectrometry Using 355-nm Radiation. *J. Am. Chem. Soc.* 1995;117:6048-56.
- Zink A, Haas CJ, Reischl U, Szeimies U, Nerlich AG. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J. Med. Microbiol.* 2001;50(4):355-66.
- Zink AR, Grabner W, Nerlich AG. Molecular identification of human tuberculosis in recent and historic bone tissue samples: The role of molecular techniques for the study of historic tuberculosis. *Am. J. Phys. Anthropol.* 2005;126(1):32-47.

VII. Annexes

Annexe 1 : Exemple de consentement éclairé signé par toute personne ayant accepté que son ADN fasse partie des échantillons analysés dans cette étude.

Institut
de médecine
légale et de
médecine sociale



Professeur Bertrand Ludes
Directeur

Docteur Antoine Tracqui
MCU-PH

Docteur Jean-Sébastien Raul
MCU-PH

Docteur Annie Gérard
PH

Faculté de médecine
Laboratoire d'analyses
toxicologiques
11 rue Humann
F-67085 Strasbourg Cedex
Tél. : (33) 03 90 24 33 63
Fax : (33) 03 90 24 33 62
Laboratoire : (33) 03 90 24 33 53
www-ulpmed.u-strasbg.fr
institut.iml@iml-ulp.u-strasbg.fr
Urgences médico-judiciaires
24h/24h
Tél. Centre 15

CONSENTEMENT ECLAIRE

Je soussigné(e), accepte que soit effectué sur ma personne un prélèvement de cellules buccales par cytobrosse à des fins d'études statistiques sur les fréquences alléliques de marqueurs génétiques situés dans des régions non codantes du génome.

J'ai bien été informé(e) que l'objectif exclusif des ces études consiste en la recherche de marqueurs génétiques pour l'identification. Il ne sera procédé à aucune investigation sans recueil de mon consentement écrit.

Ces recherches seront pratiquées conformément à la législation en vigueur et notamment en vertu de la loi n° 94-653 du 29 juillet 1994 et des articles 226-25 à 226-30 du Code Pénal et des articles 16-10) 16-12 di Code Civil.

J'atteste avoir parfaitement compris ces renseignements qui m'ont été apportés de façon claire et intelligible.

.....
Nom et Prénom

.....
Date

.....
Signature

Annexe 2 (1 sur 4) : Résumé des loci étudiés en tant que marqueurs pour l'identification génétique.

Le tableau comprend le numéro de la séquence de référence du SNP dans la base de données Single Nucleotide Polymorphism database (dbSNP rs) du National Center for Biotechnology Information (NCBI), le code allégeant la dénomination attribué au laboratoire (labcode), leurs position sur la carte cytogénétique génétique (CytPos), leurs positions génomiques absolues, les allèles observés de chaque polymorphisme, les fréquences alléliques moyennes (telles que disponibles en avril 2005 dans la banque de données, f(allèles)) et les taux moyens d'hétérozygotie (Hmoy).

dbSNP rs	Labcode	CytPos	Position génomique	Allèles	f(allèles) (NCBI)	Hmoy (NCBI)
rs930548	SP01	1p13,2	chr1 : 112219606	G / C	0,544 / 0,456	0,496 +/- 0,044
rs991191	SP02	1p31,3	chr1 : 62233326	A / G	0,562 / 0,438	0,492 +/- 0,062
rs2013526	Or05	1q24,1	chr1 : 163631465	C / T	0,531 / 0,469	0,498 +/- 0,031
rs726427	SP03	1q25,3	chr1 : 181397250	C / A	0,550 / 0,450	0,495 +/- 0,050
rs1410996	SP04	1q31,3	chr1 : 193428590	C / T	0,662 / 0,338	0,448 +/- 0,153
rs586726	SP05	1q43	chr1 : 232961831	T / C	0,552 / 0,448	0,495 +/- 0,052
rs1019264	Or11	2p16,1	chr2 : 59924453	C / T	0,563 / 0,437	0,492 +/- 0,063
rs714855	SP06	2p22,1	chr2 : 39680134	G / C	0,553 / 0,447	0,494 +/- 0,053
rs1020636	Or09	2p22,3	chr2 : 33097912	T / C	0,577 / 0,423	0,488 +/- 0,076
rs1516898	SP07	2p24,2	chr2 : 18433537	A / T	0,709 / 0,291	0,412 +/- 0,190
rs222	WI1121	2q11,2	chr2 : 98606390	T / C	0,758 / 0,242	0,367 +/- 0,221
rs223	WI1960	2q21,2	chr2 : 133206388	A / T	0,573 / 0,427	0,489 +/- 0,072
rs967782	SP08	2q23,3	chr2 : 153587047	A / G	0,640 / 0,360	0,461 +/- 0,134
rs207908	SP09	2q35	chr2 : 216841453	A / T	0,551 / 0,449	0,495 +/- 0,051
rs958383	SP10	3p13	chr3 : 74019454	G / A	0,564 / 0,436	0,492 +/- 0,063
rs428311	SP11	3p24,1	chr3 : 27956513	A / T	0,610 / 0,390	0,476 +/- 0,107
rs227	WI1341	3p26,1	chr3 : 7881607	G / A	0,676 / 0,324	0,438 +/- 0,165
rs228	WI1837	3p26,1	chr3 : 5575129	C / T	0,743 / 0,257	0,382 +/- 0,212
rs1551995	SP12	3p26,3	chr3 : 1375096	C / T	0,602 / 0,398	0,479 +/- 0,100
rs909078	SP13	3q13,33	chr3 : 121846997	T / A	0,591 / 0,409	0,483 +/- 0,090
rs986909	SP14	3q21,2	chr3 : 125766591	A / G	0,620 / 0,380	0,471 +/- 0,117
rs229	WI562	3q22,3	chr3 : 139125169	T / C	0,786 / 0,214	0,337 +/- 0,234
rs1371896	SP15	3q26,1	chr3 : 167477381	T / C	0,562 / 0,438	0,492 +/- 0,062
rs936476	SP16	3q26,31	chr3 : 176327242	G / A	0,547 / 0,453	0,496 +/- 0,046
rs225	WI1327	3q29	chr3 : 196059587	C / G	0,580 / 0,420	0,487 +/- 0,079
rs348495	SP17	4p13	chr4 : 45025370	T / C	0,619 / 0,381	0,472 +/- 0,116
rs17044	WI1888	4p13	chr4 : 41923407	A / G	0,767 / 0,233	0,358 +/- 0,226
rs231	WI945	4p16,1	chr4 : 9882117	G / C	0,649 / 0,351	0,456 +/- 0,142
rs873924	SP18	4p16,3	chr4 : 4440150	G / A	0,541 / 0,459	0,497 +/- 0,041
rs763443	SP19	4q22,1	chr4 : 91177139	C / T	0,579 / 0,421	0,488 +/- 0,078
rs1365491	SP20	4q23	chr4 : 101104432	G / A	0,671 / 0,329	0,442 +/- 0,160
rs279536	SP21	4q26	chr4 : 116245328	C / T	0,579 / 0,421	0,488 +/- 0,078
rs1541735	SP22	4q31,21	chr4 : 142203879	A / G	0,654 / 0,346	0,453 +/- 0,146
rs230	WI1803	4q34,1	chr4 : 173271203	G / A	0,549 / 0,451	0,495 +/- 0,049
rs1037102	SP23	5p13,3	chr5 : 33845060	C / A	0,552 / 0,448	0,495 +/- 0,051

Annexe 2 (2 sur 4) : Résumé des loci étudiés en tant que marqueurs pour l'identification génétique.

dbSNP rs	Labcode	CytPos	Position génomique	Allèles	f(allèles) (NCBI)	Hmoy (NCBI)
rs1394641	SP24	5p15,31	chr5 : 8560134	A / G	0,661 / 0,339	0,448 +/- 0,152
rs428538	SP25	5q21,2	chr5 : 104053606	C / T	0,698 / 0,302	0,422 +/- 0,181
rs724784	AB1D	5q21,3	chr5 : 109081234	A / C	0,592 / 0,408	0,483 +/- 0,090
rs1371370	SP26	5q23,2	chr5 : 124654291	T / C	0,620 / 0,380	0,471 +/- 0,116
rs889012	AB1C	5q31,1	chr5 : 135582615	C / A	0,54 / 0,46	0,497 +/- 0,039
rs3792774	AB1B	5q33,1	chr5 : 150483570	A / G	0,550 / 0,450	0,495 +/- 0,050
rs878953	SP27	5q34	chr5 : 163859070	G / A	0,568 / 0,432	0,491 +/- 0,067
rs2242209	AB1A	5q35,3	chr5 : 179982224	T / C	0,556 / 0,444	0,494 +/- 0,055
rs614570	AB1I	6p12,2	chr6 : 52427210	C / T	0,574 / 0,426	0,489 +/- 0,073
rs833069	SP28	6p21,1	chr6 : 43850557	A / G	0,526 / 0,474	0,499 +/- 0,026
rs910170	AB1H	6p21,2	chr6 : 39140393	C / T	0,592 / 0,408	0,483 +/- 0,090
rs1075665	Or24	6q15	chr6 : 88423321	T / C	0,599 / 0,401	0,480 +/- 0,09
rs713053	SP29	6q26	chr6 : 162541674	G / A	0,649 / 0,351	0,456 +/- 0,142
rs1578260	SP30	6q27	chr6 : 164671460	G / A	0,507 / 0,493	0,500 +/- 0,007
rs997556	Or29	7p12,1	chr7 : 51738954	T / C	0,627 / 0,373	0,468 +/- 0,123
rs953572	SP31	7p12,3	chr7 : 47785601	T / C	0,575 / 0,425	0,489 +/- 0,074
rs1072292	Or28	7p21,1	chr7 : 15291850	T / C	0,543 / 0,457	0,496 +/- 0,043
rs721015	SP32	7q21,2	chr7 : 91569907	T / C	0,547 / 0,453	0,496 +/- 0,047
rs2260	WI1126	7q21,3	chr7 : 96009108	T / C	0,585 / 0,415	0,486 +/- 0,083
rs234	WI1732	7q22,3	chr7 : 105155086	C / T	0,523 / 0,477	0,499 +/- 0,023
rs1468656	SP33	7q31,32	chr7 : 122969456	C / G	0,526 / 0,474	0,499 +/- 0,026
rs725	WI681	7q34	chr7 : 140350113	A / G	0,662 / 0,338	0,448 +/- 0,153
rs875588	SP34	7q36,1	chr7 : 150606478	T / C	0,514 / 0,486	0,500 +/- 0,014
rs1002058	SP35	7q36,3	chr7 : 157065310	G / A	0,691 / 0,309	0,427 +/- 0,176
rs1116085	SP36	8p21,3	chr8 : 22450865	T / C	0,510 / 0,490	0,500 +/- 0,010
rs734701	Or32	8p23,1	chr8 : 6375655	T / C	0,656 / 0,344	0,451 +/- 0,149
rs1371766	SP37	8p23,3	chr8 : 1989945	G / T	0,584 / 0,416	0,486 +/- 0,083
rs919023	Or31	8q12,1	chr8 : 57724593	T / C	0,624 / 0,376	0,469 +/- 0,120
rs1507079	SP38	8q12,1	chr8 : 59270104	A / G	0,537 / 0,463	0,497 +/- 0,037
rs235	WI1305	8q21,12	chr8 : 78713968	C / T	0,543 / 0,457	0,496 +/- 0,043
rs892503	Or33	8q22,3	chr8 : 103619190	C / T	0,617 / 0,383	0,473 +/- 0,113
rs874746	SP39	8q24,3	chr8 : 140722490	A / C	0,605 / 0,395	0,478 +/- 0,102
rs1034043	SP40	9q21,33	chr9 : 85114701	G / A	0,617 / 0,383	0,473 +/- 0,114
rs946054	SP41	9q31,1	chr9 : 100123570	T / G	0,582 / 0,418	0,487 +/- 0,081
rs240	WI2032	9q31,3	chr9 : 109558585	C / G	0,599 / 0,401	0,480 +/- 0,097
rs237	WI1306	9q33,1	chr9 : 119503239	A / G	0,703 / 0,297	0,417 +/- 0,186
rs878680	SP42	9q33,3	chr9 : 124166231	C / T	0,697 / 0,303	0,423 +/- 0,181
rs943623	SP43	9q34,2	chr9 : 133296925	C / A	0,512 / 0,488	0,500 +/- 0,012
rs936250	SP44	9q34,3	chr9 : 138247513	G / C	0,634 / 0,366	0,464 +/- 0,129
rs927099	SP45	10p11,22	chr10 : 33565842	G / A	0,644 / 0,356	0,459 +/- 0,138
rs997750	Or36	10p12,1	chr10 : 27959937	C / T	0,500 / 0,500	0,5 +/- 0,000
rs241	WI1349	10q22,3	chr10 : 78114462	A / C	0,540 / 0,460	0,497 +/- 0,040

Annexe 2 (3 sur 4) : Résumé des loci étudiés en tant que marqueurs pour l'identification génétique.

dbSNP rs	Labcode	CytPos	Position génomique	Allèles	f(allèles) (NCBI)	Hmoy (NCBI)
rs766398	SP46	10q23,32	chr10 : 93864089	A / G	0,565 / 0,435	0,492 +/- 0,064
rs243	WI2054	10q24,1	chr10 : 98230372	T / C	0,539 / 0,461	0,497 +/- 0,039
rs1377006	SP48	10q25,1	chr10 : 106882129	A / C	0,539 / 0,461	0,497 +/- 0,039
rs740600	SP49	10q25,3	chr10 : 118641872	G / A	0,642 / 0,358	0,459 +/- 0,137
rs1316543	SP50	10q26,11	chr10 : 121194650	G / A	0,771 / 0,229	0,353 +/- 0,228
rs363225	SP51	10q26,11	chr10 : 119014492	C / T	0,569 / 0,431	0,490 +/- 0,069
rs729999	Or39	11p15,1	chr11 : 19934294	C / T	0,585 / 0,415	0,486 +/- 0,083
rs728639	SP52	11p15,1	chr11 : 17905113	A / G	0,635 / 0,365	0,463 +/- 0,130
rs451041	AB2R	11p15,4	chr11 : 3017301	G / A	0,604 / 0,396	0,479 +/- 0,101
rs2301698	AB2B	11p15,5	chr11 : 2394001	G / T	0,564 / 0,436	0,492 +/- 0,064
rs86392	SP53	11p15,5	chr11 : 2683761	T / C	0,537 / 0,463	0,497 +/- 0,037
rs544021	AB2Q	11q13,2	chr11 : 66072237	C / G	0,541 / 0,459	0,497 +/- 0,041
rs586459	SP54	11q13,3	chr11 : 69170555	G / C	0,545 / 0,455	0,496 +/- 0,045
rs882937	AB1M	11q21	chr11 : 93548520	A / G	0,502 / 0,498	0,500 +/- 0,002
rs693172	SP55	11q21	chr11:95276266	C / T	0,607 / 0,393	0,477 +/- 0,104
rs1021290	Or41	11q22,3	chr11 : 103347752	T / C	0,599 / 0,401	0,481 +/- 0,097
rs1013582	SP56	11q23,3	chr11 : 114740307	A / T	0,559 / 0,441	0,493 +/- 0,059
rs1793286	AB2J	11q25	chr11 : 132084148	A / T	0,563 / 0,437	0,492 +/- 0,063
rs959566	Or43	12p11,22	chr12 : 30160004	T / C	0,683 / 0,317	0,433 +/- 0,170
rs1485421	SP57	12p11,22	chr12 : 30275862	C / T	0,697 / 0,303	0,423 +/- 0,181
rs956864	SP58	12q13,12	chr12 : 48576290	T / C	0,635 / 0,365	0,463 +/- 0,130
rs2123	WI563	12q14,1	chr12 : 61333877	A / G	0,33 / 0,67	0,438 +/- 0,115
rs958644	SP59	12q15	chr12 : 69383756	C / A	0,642 / 0,358	0,460 +/- 0,136
rs811	WI597	12q21,33	chr12 : 89278405	A / G	0,750 / 0,250	0,375 +/- 0,217
rs4502027	WI1731	12q22	chr12 : 91477210	A / G	0,88 / 0,12	0,265 +/- 0,244
rs730013	Or44	12q23,2	chr12 : 100652449	T / C	0,562 / 0,438	0,492 +/- 0,061
rs1025256	SP60	12q24,21	chr12 : 113067877	A / G	0,566 / 0,434	0,491 +/- 0,065
rs730249	Or45	13q13,3	chr13 : 38999740	C / T	0,558 / 0,442	0,493 +/- 0,057
rs1924609	AB2H	13q14,3	chr13 : 51410519	C / T	0,513 / 0,487	0,500 +/- 0,013
rs1010297	SP61	13q21,33	chr13 : 70637202	C / T	0,700 / 0,300	0,420 +/- 0,183
rs719223	SP62	13q31,3	chr13 : 93532937	G / A	0,787 / 0,213	0,336 +/- 0,235
rs1336659	SP63	13q33,1	chr13: 101750260	A / G	0,576 / 0,424	0,488 +/- 0,075
rs729549	Or48	13q33,3	chr13:108213189	C / T	0,580 / 0,420	0,487 +/- 0,079
rs911621	Or49	14q22,2	chr14 : 54195466	C / T	0,516 / 0,484	0,499 +/- 0,016
rs757655	SP64	14q24,2	chr14 : 72174748	A / G	0,737 / 0,263	0,387 +/- 0,209
rs734656	Or50	14q31,3	chr14 : 83737776	T / C	0,669 / 0,331	0,443 +/- 0,159
rs765034	SP65	14q32,2	chr14 : 97285394	A / G	0,557 / 0,443	0,493 +/- 0,057
rs999842	Or51	15q11,2	chr15 : 20551713	C / T	0,642 / 0,358	0,460 +/- 0,136
rs780	WI1147	15q12	chr15 : 23246132	G / A	0,775 / 0,225	0,349 +/- 0,230
rs588739	SP66	15q15,1	chr15 : 38172380	G / C	0,526 / 0,474	0,499 +/- 0,026
rs877228	Or53	15q22,2	chr15 : 58863883	C / T	0,549 / 0,451	0,495 +/- 0,049
rs767453	SP67	15q23	chr15 : 66873370	A / T	0,508 / 0,492	0,500 +/- 0,008

Annexe 2 (4 sur 4) : Résumé des loci étudiés en tant que marqueurs pour l'identification génétique.

dbSNP rs	Labcode	CytPos	Position génomique	Allèles	f(allèles) (NCBI)	Hmoy (NCBI)
rs1550436	SP68	15q24,1	chr15 : 72008210	T / C	0,564 / 0,436	0,492 +/- 0,064
rs893613	SP69	15q26,1	chr15 : 88196981	C / A	0,570 / 0,430	0,490 +/- 0,069
rs765287	SP70	16p13,2	chr16 : 9783883	C / T	0,612 / 0,388	0,475 +/- 0,109
rs1519976	SP71	16q24,1	chr16 : 84503176	A / C	0,599 / 0,401	0,480 +/- 0,097
rs231026	SP72	17p12	chr17 : 15097654	C / A	0,580 / 0,420	0,487 +/- 0,079
rs887260	SP73	17p13,1	chr17 : 9812713	T / C	0,503 / 0,497	0,500 +/- 0,003
rs2073867	AB2D	17q11,2	chr17 : 23703087	G / C	0,621 / 0,379	0,471 +/- 0,117
rs727206	Or54	17q12	chr17 : 28942222	C / T	0,536 / 0,464	0,497 +/- 0,036
rs1367957	SP74	17q12	chr17 : 29822312	A / G	0,519 / 0,481	0,499 +/- 0,019
rs903107	AB2N	17q25,1	chr17 : 68573529	A / G	0,622 / 0,378	0,470 +/- 0,118
rs878906	AB2O	17q25,1	chr17 : 70280929	A / G	0,571 / 0,429	0,490 +/- 0,070
rs820129	AB2P	17q25,1	chr17 : 71179702	A / G	0,511 / 0,489	0,500 +/- 0,011
rs868432	Or57	17q25,3	chr17 : 76492330	C / T	0,569 / 0,431	0,490 +/- 0,069
rs965593	SP75	18q11,2	chr18 : 20539763	G / A	0,659 / 0,341	0,449 +/- 0,151
rs339942	SP76	18q11,2	chr18 : 20540488	C / T	0,569 / 0,431	0,490 +/- 0,068
rs276922	AB2T	18q12,1	chr18 : 26858793	A / C	0,572 / 0,428	0,490 +/- 0,072
rs1565784	SP77	18q12,1	chr18 : 29612489	G / T	0,619 / 0,381	0,472 +/- 0,115
rs948438	SP78	18q12,2	chr18 : 33105936	A / G	0,552 / 0,448	0,494 +/- 0,052
rs1144098	SP79	18q21,1	chr18 : 48380783	A / G	0,590 / 0,410	0,484 +/- 0,088
rs768058	SP80	18q21,33	chr18 : 59107756	A / G	0,475 / 0,525	0,314 +/- 0,242
rs1000329	Or61	19p13,11	chr19 : 16310517	T / C	0,683 / 0,317	0,433 +/- 0,171
rs1465701	SP81	19p13,2	chr19 : 10539842	G / A	0,562 / 0,438	0,492 +/- 0,062
rs312926	SP82	19p13,3	chr19 : 3226315	G / A	0,531 / 0,469	0,498 +/- 0,031
rs238417	SP83	19q13,32	chr19 : 50548308	G / C	0,619 / 0,381	0,472 +/- 0,115
rs392891	SP84	19q13,32	chr19 : 53326131	A / T	0,541 / 0,459	0,497 +/- 0,041
rs2290775	AB2C	19q13,33	chr19 : 55067054	C / T	0,625 / 0,375	0,469 +/- 0,121
rs1674139	AB2I	19q13,33	chr19 : 54963314	C / T	0,552 / 0,448	0,495 +/- 0,052
rs1003204	Or62	20p11,21	chr20 : 23478035	C / T	0,549 / 0,451	0,495 +/- 0,049
rs1321158	SP85	20p12,1	chr20 : 12582137	A / G	0,637 / 0,363	0,463 +/- 0,131
rs1476718	SP86	20q13,2	chr20 : 51961640	G / C	0,573 / 0,427	0,489 +/- 0,072
rs754	WI1325	20q13,33	chr20 : 58102499	A / G	0,767 / 0,233	0,356 +/- 0,143
rs964224	SP87	21q21,3	chr21 : 27545948	T / G	0,576 / 0,424	0,489 +/- 0,075
rs762173	SP88	21q22,11	chr21 : 32754546	A / G	0,619 / 0,381	0,472 +/- 0,116
rs1543754	SP89	21q22,13	chr21 : 38015977	C / G	0,540 / 0,460	0,497 +/- 0,040
rs474656	SP90	21q22,2	chr21 : 39907990	G / C	0,523 / 0,477	0,499 +/- 0,023
rs228047	AB1T	21q22,3	chr21 : 42825058	C / T	0,753 / 0,247	0,371 +/- 0,219
rs696880	SP91	22q11,21	chr22 : 18609754	A / G	0,621 / 0,379	0,471 +/- 0,118
rs1540299	AB2G	22q12,3	chr22 : 32232518	A / G	0,500 / 0,500	0,500 +/- 0,000
rs738518	Or68	22q12,3	chr22 : 35444300	T / C	0,783 / 0,217	0,339 +/- 0,233
rs737612	SP92	22q13,2	chr22 : 41971914	A / G	0,648 / 0,352	0,456 +/- 0,141

Annexe 3 (1 sur 4) : Récapitulatif des amorces utilisées pour l'étude des loci sélectionnés en tant que marqueurs pour l'identification génétique et la taille des fragments en paires de bases (pb) qu'elles permettent d'amplifier. Les séquences sont décrites de leur extrémité 5' vers leur extrémité 3'. Les amorces sont nommées selon le code de laboratoire du SNP contenu dans le fragment qu'elles permettent d'amplifier et leur sens d'hybridation (s pour sens et as pour antisens).

>AB1C-s	GTA CCC AGA CAC AGC CCA AAC C	
>AB1C-as	TCA TCT GAG CAT CAA AAC AAA GG	100 pb
>AB1D-s	CTT TCT CCT TAA TCT TAG C	
>AB1D-as	AGT CCT TCC ATG ATG CCC	70 pb
>AB1H-s	TGT GAG GCT GAT AGA GGC	
>AB1H-as	TAA GTT CCC CTA CAG AGA G	79 pb
>AB1M-s	CTG TGG GGA GAA GGA ATG	
>AB1M-as	TTA TGG AAG CAG CAG AAA CTC	103 pb
>AB1T-s	TCC TCA GAA CTC TTT GTG C	
>AB1T-as	CCT TCT CCT TGA TGT TGG	114 pb
>AB2H-s	TCC TTC CCC CCT TTT CTG	
>AB2H-as	ACC TGC CTG TCA CAT TCG	80 pb
>AB2I-s	GGA AGA GTT GTC GTG ACG ATG G	
>AB2I-as	AGC CAT TCC CAG GAG TTG AGA G	64 pb
>AB2J-s	GAG AAG AGA AAT GAC AAC TGG	
>AB2J-as	ATG GGT CAC GGA GTT TCA C	91 pb
>AB2P-s	GCA ACT GTG TGG ACA TTT TCT GG	
>AB2P-as	TGG AAA TCC AAT GTG TCT CAG G	100 pb
>AB2Q-s	GAT GAC AGG AGA TGA CCC C	
>AB2Q-as	GCC GTG CTT GGA GAC AGG	109 pb
>AB2T-s	AGT GAC AAA TTG GGA ATT GCT G	
>AB2T-as	AAT CAG AAA ATT ATG CTG GAG AGA AC	99 pb
>AmelSD-s	CAT GGG TGG ATG GCT GC	
>AmelSD-as	CCA CTG GGA TGT GGT GAT G	95 pb

Annexe 3 (2 sur 4) : Récapitulatif des amorces et la taille des fragments qu'elles permettent d'amplifier

>Or05-s	CTA CAA GTT CAA ATA TAC TAA AC	
>Or05-as	TGT GAC ATA GAG CCA GAT TC	61 pb
>Or09-s	AGG CTG GTC TGG CAA TCT GG	
>Or09-as	TGC ACG TGG AAC AGT GAA GTA CC	91 pb
>Or11-s	CTG AAA GTA GAA CCA ATC AAG	
>Or11-as	CCT GAA AAT GTA GCC TGG G	86 pb
>Or29-s	CAG AGT GGA GAG TCA CAA AAT GG	
>Or29-as	GCT GCA CAA AAT GGA GGG AAT C	79 pb
>Or33-s	GCT GCT GAG GAA GAG TGC	
>Or33-as	GGT TGG GAT GGC ATT TGG	71 pb
>Or36-s	GAA TTT TGT GTA GGC AGC C	
>Or36-as	AGA GGT AGA TAG AAT TGA ATA G	57 pb
>Or54-s	GAT GGC AGT CAC CCT CAG	
>Or54-as	GTC TAT TTC CTA ACA CCT CTG	74 pb
>Or62-s	GTG TGT TGT CCA ACC TTC C	
>Or62-as	TGT CCC TGT GTG TAG TGG	82 pb
>Or68-s	ATG CCG CTG TGA GAC AAA G	
>Or68-as	TGT CAT TCA GAG AAG GAA GC	68 pb
>SP02-s	ATA TCT TAA TGC TTC ACT TGG	
>SP02-as	TTA ATC CAT ACC AGA GTG CG	137 pb
>SP12-s	CTT GAC TCT ATT CTG ACC G	
>SP12-as	CTA GAG CAT AAG TGA CCA C	95 pb
>SP19-s	GGA AGC CTT CTC AGG TGA AGT TAG G	
>SP19-as	GAA CAG GGG AAT TCA AAG AGA CC	71 pb
>SP21-s	TTG GCT GTC AGA GGG ATT G	
>SP21-as	TTT GAA TCA TTT TTG GAA CTG	70 pb

Annexe 3 (3 sur 4) : Récapitulatif des amorces et la taille des fragments qu'elles permettent d'amplifier

>SP26-s	AAG AGA AGC AGC GGA AAA C	
>SP26-as	TTC ATT CAA GGA TAG ATC AC	89 pb
>SP28-s	GTC AAG TGG GGT TTT CTC C	
>SP28-as	ATT TGT GGG AAC TGT GAC C	120 pb
>SP29-s	ACA TCC ACT ACC ATC GCC G	
>SP29-as	ATT ACA GAT TAG CCC ACA C	147 pb
>SP39-s	ATG GAG AGG ATA GTG AAC C	
>SP39-as	AGG ACC AAA TGA GAT GAC C	99 pb
>SP44-s	GGC TGT CAT CAG TGA GGA GTG G	
>SP44-as	CAG GGT ACA AGT GGG TCA GAG G	98 pb
>SP56-s	AGT CCA GGT GTT TCC ATT G	
>SP56-as	CTC TCC TAT CCA TTT TCC AC	98 pb
>SP64-s	TGG TTG GAT GCT CCT GAG	
>SP64-as	CAA CAG TGT GTG AAG GTC	122 pb
>SP67-s	GCT CCT TCT TTT TCT AGT AC	
>SP67-as	GCC AGT TCT CTC GGA CTC	80 pb
>SP78-s	ACA TGA GCC CAG GCA GTG G	
>SP78-as	AAA ACC CAG AGG CTG GAA AGG	95 pb
>SP80-s	GCG GGT TCA GAG CAT AGC	
>SP80-as	CTT CCT TTT CCT TTA TCC AGC	120 pb
>SP87-s	TTT CCT CCT CAC TCA TCC C	
>SP87-as	TTC CAT GTC TGT ACT ATA AAC	72 pb
>SP92-s	AAT AAT GCC CCC TTA AAA TGC	
>SP92-as	TCT CAT GGT CCT TCA GCA C	96 pb
>WI1121-s	TGT AAC ATT AGA CCT CTC AC	
>WI1121-as	TAC AAA ATG AAG CCA GAG C	86 pb

Annexe 3 (4 sur 4) : Récapitulatif des amorces et la taille des fragments qu'elles permettent d'amplifier

>WI1126-s	TGA AAT TTG TTA AAA TGC AAA TCC A	
>WI1126-as	TGC AGC TCA TCC CTA CTC ATT TG	114 pb
>WI1305-s	CAC TGC TTT CAA TAA CTC ACT CC	
>WI1305-as	AAT GGT AGG CAG GGC TGT TAT G	115 pb
>WI1306-s	GAT GGC TTT TCA GTC TGT G	
>WI1306-as	AAG AGG CAG GAT ACC AAG	94 pb
>WI1327-s	GGA GCT GAG AGT CTG TGA AGT TGG	
>WI1327-as	GAT GGA ACC ATT GCC TCT TCA G	98 pb
>WI1341-s	ATC CAC TAT CTT CTC TCA AG	
>WI1341-as	AAA ATA AAC AGA TCC ATC AC	89 pb
>WI1349-s	GGA TAG TGC TGA ATT TCC ATC TCT G	
>WI1349-as	TGG TAG GTG GCA GGA TTC AGT G	150 pb
>WI1732-s	AAA GGG GAC ATA TTC AAG GAG TGT C	
>WI1732-as	CTG ATG CAC CCA TCA CTC ACA G	90 pb
>WI1803-s	AAT AAT AGT GTT GGA GAT AAG	
>WI1803-as	AAG TGT GAT GCT CTA CCC	91 pb
>WI2032-s	GAC TTT CAC AGA TCA TTG G	
>WI2032-as	ATG TTG TCC CAT CAA ATG C	85 pb
>WI2054-s	GGA CAA GAC AGG TGA CC	
>WI2054-as	TAA GAG ATA GGG CTG GTG	130 pb
>WI563-s	TAG AGG GTT TAA CAA TGG C	
>WI563-as	AGT AGG CAG TTA GGT AGA C	100 pb
>WI597-s	TGT GAA AGG TGG AAG GTA AGG ATG	
>WI597-as	TGA TGA GGT ACC AGA GTC ATG GAG	128 pb
>WI681-s	TGC CTG GTA TGT CAT AAG CAA TCC	
>WI681-as	TGC TGC ATA TCT TAA AGG ACA TTC G	116 pb

Annexe 4 (1 sur 4) : Récapitulatif des séquences (des amorces, du polymorphisme et de 3 nucléotides en 3' du SNP), masses molaires des amorces à l'état monoprotoné (Molecular Weight, MW⁺) et températures de fusion des amorces PEX (T_m et T_{mn}) utilisées pour l'étude des 50 SNP autosomaux sélectionnés en tant que marqueurs pour l'identification génétique et du marqueur permettant la détermination de sexe.

Les séquences sont décrites de leur extrémité 5' vers leur extrémité 3'. Le polymorphisme est indiqué par le signe α et les allèles sont représentés par les lettres du code IUB.

*Température approximative de fusion : $T_m = 2^\circ\text{C} \cdot (A+T) + 4^\circ\text{C} \cdot (G+C)$

**T_m normalisé selon la longueur : $T_{mn} = 22^\circ\text{C} + 1,46^\circ\text{C} \cdot ((A+T) + 2(G+C))$

Nom	Séquence (amorce PEX -SNP- séquence en 3')	MW ⁺ (Da)	T _m [*]	T _{mn} ^{**}
>AB1C-sr	GC ATC AAA ACA AAG GCT CAA -K- ATG - α - . . .	6113,09	56 °C	69 °C
>AB1D-sf	AT CTT AGC CTG TAC CTT TTA -M- ATC - α - . . .	6034,00	54 °C	70 °C
>AB1H-sr	TG AAC CAT TTC CTC CTG GAC -R- CAA - α - . . .	6028,99	60 °C	66 °C
>AB1M-sr	AAC TAA TGG GCA CCC TTC -Y- GAT - α - . . .	5444,62	54 °C	61 °C
>AB1T-sr	C GCA CCC CGC AGA CCT -R- TTG - α - . . .	4773,16	56 °C	51 °C
>AB2H-sf	T CTA TTC AGC AGG CAC TTA -Y- GAA - α - . . .	5763,83	54 °C	66 °C
>AB2I-sr	G AGC TGA CTT GCC GCC -R- GGA - α - . . .	4859,21	54 °C	53 °C
>AB2J-sf	GA AAT GAC AAC TGG TGA AAT -W- CCC - α - . . .	6199,14	54 °C	70 °C
>AB2P-sf	ATA AAG AAA GAA CCT GGA GAC -R- ACT - α - . . .	6506,35	58 °C	72 °C
>AB2Q-sr	G TCA GCG GAG CCT CTC -S- GCC - α - . . .	4859,21	54 °C	53 °C

Annexe 4 (2 sur 4) : Séquences des amorces, du polymorphisme et de 3' du SNP

Nom	Séquence (amorce PEX -SNP- Séquence en 3')	MW ⁺ (Da)	T _m	T _{mn}
>AB2T-sr	GA GAG AAC AAT TAA ACT CAC -K- GCA -α- ...	6128,10	54 °C	70 °C
>AmelSD-sr	C TGC AGG GTG TGA GTC -R- GGG -α- ...	4954,28	52 °C	60 °C
>Or05-sf	AG TTC AAA TAT ACT AAA CTA TTC -Y- CTG -α- ...	6990,67	56 °C	82 °C
>Or09-sf	CA GGC AGG TGG TCA GGC -Y- ATG -α- ...	5277,49	58 °C	54 °C
>Or11-sf	T AGA ACC AAT CAA GGC TCC -Y- ACT -α- ...	5766,84	56 °C	64 °C
>Or29-sr	A CAC TGG ATA ACC AAG GAC -R- TAA -α- ...	5815,88	56 °C	64 °C
>Or33-sf	GC AGC ATA CTC ATT CAC AGA -Y- GGT -α- ...	6071,04	58 °C	67 °C
>Or36-sr	G TAG ATA GAA TTG AAT AGA AGT -R- AAG -α- ...	6871,58	56 °C	63 °C
>Or54-sr	C TGT TTT CTG TTA TAT TCT GCT -R- TTC -α- ...	6664,39	58 °C	64 °C
>Or62-sr	GTG TAG TGG GAG CAC TGG -R- TAG -α- ...	5636,72	58 °C	59 °C
>Or68-sf	CCG CTG TGA GAC AAA GGG -Y- GAG -α- ...	5574,70	58 °C	59 °C
>SP02-sr	GT TGT TAA TGA GGA TGA TGC -Y- GTC -α- ...	6228,13	56 °C	69 °C
>SP12-sf	AG GCT ACT CTT CTT CTT AGG -Y- ATT -α- ...	6075,01	58 °C	67 °C
>SP19-sr	AAG AGA CCT GTA GGA GCG -R- TTA -α- ...	5598,72	56 °C	60 °C

Annexe 4 (3 sur 4) : Séquences des amorces, du polymorphisme et de 3 nucléotides en 3' du SNP

Nom	Séquence (amorce PEX -SNP- Séquence en 3')	MW ⁺ (Da)	Tm	Tmn
>SP21-sf	ATT GTC AGT CTC AAA AGC ACA -Y- TTG -α- ...	6399,26	58 °C	72 °C
>SP26-sr	GCA GGG GGA GTC CTT AAC -R- TTT -α- ...	5565,68	58 °C	59 °C
>SP28-sf	CAA ACC AGG GAA GGG AGG -R- GAG -α- ...	5632,75	58 °C	59 °C
>SP29-sf	CA ACA ACA GTG AAA AGG ATC -R- GTC -α- ...	6153,12	56 °C	69 °C
>SP39-sf	GTG CCA GAC CGC TTC TTG -M- TGC -α- ...	5467,61	58 °C	59 °C
>SP44-sr	CTC CGT GTG TCT GTG TGC -S- CAG -α- ...	5489,61	58 °C	59 °C
>SP56-sr	CT CTC CTA TCC ATT TTC CAC -W- CTT -α- ...	5914,91	58 °C	67 °C
>SP64-sf	A ATT GAT GTT GGG CTC TGG -R- CAA -α- ...	5890,89	56 °C	64 °C
>SP67-sr	CT TCT GTC CAC CTT TCC -W- TGT -α- ...	5024,32	52 °C	59 °C
>SP78-sr	T GAG AAT GTG GAC CTG AGG -Y- TGG -α- ...	5933,93	58 °C	63 °C
>SP80-sr	GC TGT TCA AAT TCC TCT CTC -Y- ATC -α- ...	5994,96	58 °C	67 °C
>SP87-sf	AGT AGG CTT TTC ATC AAT TAG -K- TTT -α- ...	6436,27	56 °C	73 °C
>SP92-sr	C AGG TAG GGT CAC AGT TTC -Y- GGG -α- ...	5844,87	58 °C	63 °C
>WI1121-sr	A GAG GAG AGT AAA AAC AGC -R- TAG -α- ...	5928,97	54 °C	66 °C

Annexe 4 (4 sur 4) : Séquences des amorces, du polymorphisme et de 3 nucléotides en 3' du SNP

Nom	Séquence (amorce PEX -SNP- Séquence en 3')	MW ⁺ (Da)	Tm	Tmn
>WI1126-sr	CAA GAA ATA AAA GAC AAG TCC -R- AAA -α- ...	6450,33	56 °C	73 °C
>WI1305-sr	GTA TAC TTA TTA CTT CTC CCT -R- CAT -α- ...	6298,17	56 °C	73 °C
>WI1306-sr	AG GTA GCC CAC AGA AGC -Y- AGG -α- ...	5214,47	54 °C	57 °C
>WI1327-sr	C CCA AGA ACT ACA TTT GGG GAG -S- CCT -α- ...	6769,49	66 °C	70 °C
>WI1341-sr	A ACA GAT CCA TCA CAA TCC -Y- GCT -α- ...	5710,82	54 °C	66 °C
>WI1349-sf	C AGA AAA TAT GAT TAG AAA AAG -M- GTG -α- ...	6833,59	54 °C	79 °C
>WI1732-sr	ATG TGG CAG AGA CTG AAT -R- AAG -α- ...	5588,72	52 °C	63 °C
>WI1803-sr	GTG ATG CTC TAC CCT ACT -Y- ATC -α- ...	5426,59	54 °C	61 °C
>WI2032-sr	C AGA AGC AAT TAC AGG AAG -S- TGA -α- ...	5879,93	54 °C	66 °C
>WI2054-sr	G GCT GGT GGA GGC AGA -R- GAA -α- ...	5028,33	54 °C	53 °C
>WI563-sr	T AGG TAG ACA TGA TCA GGG -Y- AGG -α- ...	5917,93	56 °C	64 °C
>WI597-sr	G AGT CAT GGA GGG TCA TTA -Y- CAT -α- ...	5908,92	56 °C	64 °C
>WI681-sr	CA AAT GGT GCC ATA GTA TAA -Y- AAT -α- ...	6150,10	54 °C	70 °C

VIII. Publications

SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology.

Petkovski E, Keyser-Tracqui C, Hienne R, Ludes B.

J. Forensic Sci. 2005 May;50(3):535-41.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology.

Petkovski E, Keyser-Tracqui C, Hienne R, Ludes B.

J. Forensic Sci. 2005 May, 50(3):535-41

Pages 535-541 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Il est possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise.

Dixon LA, Dobbins AE, Pulker H, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubweiser P, Mogensen HS, Morling N, Nielsen K, Sanchez JJ, Petkovski E, Carracedo A, Sanchez-Diz P, Brion M, Irwin JA, Just RS, Loreille O, Parsons TJ, Syndercombe-Court D, Schmitter H, Gill P.

Forensic Sci. Int; sous presse, article disponible sur Internet.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise.

Dixon LA, Dobbins AE, Pulker H, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubweiser P, Mogensen HS, Morling N, Nielsen K, Sanchez JJ, **Petkovski E**, Carracedo A, Sanchez-Diz P, Brion M, Irwin JA, Just RS, Loreille O, Parsons TJ, Syndercombe-Court D, Schmitter H, Gill P.

Forensic Science International; sous presse

Pages XXX à XXX.

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <http://dx.doi.org/10.1016/j.forsciint.2005.11.011>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

A set of 51 biallelic markers for human identification and filiation in French people.

Petkovski E, Keyser-Tracqui C, Strub JM, Hienne R, Van Dorsselaer A, Ludes B.

J Forensic Sci., soumis.

A set of fifty-one biallelic markers for human identification and filiation in French people¹

Elizabet Petkovski, ^{1,2} Ing.; **Christine Keyser-Tracqui**, ¹ Ph.D.; **Jean-Marc Strub**, ³ Ph.D.;
Rémi Hienne, ² Ph.D.; **Alain Van Dorsselaer**, ³ Ph.D.; **Bertrand Ludes**, ¹ M.D., Ph.D.

¹Institute of Legal Medicine, University Louis Pasteur, 11 rue Humann, 67085 Strasbourg Cedex, France.

²CODGENE Laboratory, 11 rue Humann, 67085 Strasbourg Cedex, France.

³ Laboratoire de Spectrométrie de Masse BioOrganique, UMR 7512, ECPM, 25 rue Becquerel, 67087 Strasbourg Cedex, France

Correspondence to Elizabet Petkovski, Ing.

Tel: + 33 3 90 24 33 49, Fax: + 33 3 90 24 33 62, E-mail: petkovski@ulp.u-strasbg.fr

Running header: **A 51 SNP set for genetic identification**

¹ This work has been financially supported by a French national grant through a CIFRE contract.

ABSTRACT: DNA markers used for individual identification in forensic sciences are based on repeat sequences in autosomal DNA. Filiations can also be achieved by the investigation of haplotypic markers, the mitochondrial and Y-chromosomal DNA. When due to degradation, the DNA lacks fragments of sufficient length to be investigated by the means of microsatellites, single nucleotide polymorphisms (SNPs) may represent an alternative tool for genetic identification.

In order to study the potential of SNP analysis in these fields, 50 autosomal SNPs were selected on the basis conservation, lack of phenotypic expression and frequency of occurrence in populations. Furthermore, a single base difference in the first intron of the amelogenin genes carried by the sex chromosomes was selected for biological sex determination of the sample. 75 unrelated French individuals and 24 paternity cases were investigated by genotyping this panel of markers.

Our approach to SNP typing is a multiplex PCR based amplification followed by simultaneous allelic discrimination by primer extension (PEX) whose products are analyzed by Matrix Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS). PEX and MALDI-TOF MS, with their high sensitivity, precision and speed, gave a powerful method for forensic and anthropological exploitation of biallelic markers.

KEYWORDS: forensic sciences, DNA typing, anthropology, single nucleotide polymorphisms, paternity testing, individual identification, MALDI-TOF MS

Single nucleotide polymorphisms (SNPs) represent the most common type of variation in the human genome and are usually biallelic (1). In the forensic field SNP typing is of great interest particularly when degraded samples are all that is available for analysis. Indeed, these markers require only short amplicons which increases the probability of obtaining a result when the analyses of classical STRs fail. To be on par with the STR multiplexes routinely used for genetic identifications, batteries of at least 50 SNPs must be typed (2). The plethora of these markers through out the human genome (one per kilobase (3)), their low mutation rate (in the range of 10^{-9} to 10^{-8} per generation (4, 5)) and their compatibility with high-throughput automated analysis from short DNA fragments, make them attractive as a complement to the existing forensic investigation methodologies.

Among the wide range of SNP typing technologies we chose the detection of Primer EXtension (PEX) products by Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). In this assay, allelic discrimination is achieved by hybridization of PEX primers to the target PCR amplicon immediately upstream the SNP position and its elongation by a single dideoxynucleotide or by both a deoxynucleotide and a dideoxynucleotide according to the present allele (6, 7). In this way, the mass difference between the extended primers, representing the different alleles, corresponds to at least one nucleotide (about 300Da) which can easily and unambiguously be distinguished in a MALDI-TOF MS linear mode and in a 4.5 to 8.5kDa mass range. MALDI-TOF MS SNP typing is very rapid, sensitive and precise and allows simultaneous detection of a large number of products (7).

In this context, we have selected a panel of 50 autosomal SNPs along with a single base difference between the amelogenin sequences on the gonosomes as markers for individual identification. Twenty eight of these markers have been evaluated in a previously published study (6) and the goal of the present study was to select a panel whose discriminatory power

is on the par with the one of the currently used STR multiplexes. The informative content of the panel regarding individual identification was investigated on a French Caucasian representative population consisting of 75 unrelated individuals and 24 disputed paternity trios among which 4 exclusions. The 51 amplicons (less than 150bp long) were obtained in 5 multiplexed PCR reactions. Each PCR was optimized to yield 9 to 11 equilibrated fragments. The products were split for allelic discrimination of 2 to 4 SNPs per PEX reaction allowing the application of the same reaction conditions by the use of primers showing similar melting temperatures and therefore similar sizes and masses. This combination provides a flexible detection platform for measuring the molecular mass per charge of the analyte, an intrinsic property obtained without mass tagging or product cleavage.

Materials and methods

The analysis of selected SNPs included the following major steps: PCR amplification of the target SNP containing DNA fragment, which was to be shorter than 150bp; Double stranded amplicon purification by Genopure™ ds magnetic bead assisted technology (Bruker Daltonics); Allele specific primer extension (PEX); Single stranded product purification by Genopure™ oligo technology (Bruker Daltonics) and MALDI-TOF MS analysis.

DNA samples

Two types of matrices were used in this study. Either DNA was purified from blood samples stored on FTA®-treated matrix (Whatman BioScience Ltd) or it was extracted from buccal swabs using the MagneSil® KF Genomic System (Promega) on a KingFisher® mL instrument (Thermo LabSystems). PCR was performed either on 1.2mm FTA® cardpunches after 3 washings with the manufacturer recommended FTA purification reagent, and 2 washings with TE light buffer (10mM Tris-HCl pH 8 buffer and 0.1mM EDTA) or on 300pg DNA eluted in TE light buffer.

SNP selection

For this study, the SNPs were chosen from freely available databases such as the SNP consortium. They are non-coding, at least 4.5Mbp distant and have allelic frequencies ranging between 0.2 and 0.8 in three distinct populations. Indeed, SNPs being biallelic, when single contributor samples are investigated the maximum power of discrimination is acquired by the combination of SNPs with an even allele distribution (9). However, in case of mixed DNA stains the likelihood ratios maximize when frequencies are unbalanced (2).

Sequencing

A control group was built from 15 individuals to ensure the analysis of a heterozygous individual at each SNP locus. Sequencing of each SNP of interest containing fragment was carried out using the BigDye™ terminator cycle sequencing kit (Applied Biosystems) according to the manufacturer's protocol. Each locus was sequenced twice with primers complementary to both the forward and the reverse strands, once using the short amplicons designed for the multiplexed reactions and once using amplicons yielded by primers binding outside these final fragments.

PCR amplification

Primers (18 to 22-base oligonucleotides) were designed according to the DNA sequence flanking the target SNP sites. Primers were designed so that the amplicons ranged in size from 50 to 150bp with the OLIGO-4.0-s software (<http://www.oligo.net/>). Each target DNA fragment was tested in singleplex PCR and Big Dye Terminator™ reactions (Applied Biosystems) followed by capillary electrophoresis on an ABI 3100™ analyzer (Applied Biosystems) with primers outside those for final PCR and PEX reactions (200 to 300bp fragments). This allowed us to verify the polymorphism, its flanking sequences and the primer hybridization region.

Primers were designed to have melting temperatures ranging between 58 and 62°C. Nine to eleven target fragments were amplified in a 30 µL reaction volume containing 1X Magnesium free reaction buffer; 3mM MgCl₂; 200µM each dNTP; 0.2–0.4µM each primer (adjusted to obtain equilibrated amounts of each PCR product); 1 unit DNA Polymerase (MBI Fermentas). All DNA amplifications were performed in a GeneAmp 9600 thermal cycler (Perkin-Elmer) using the following program: denaturation for 15min at 95°C followed by 30 cycles (95°C-30sec; 51°C-45sec; 72°C-30sec) and 5min at 72°C final elongation step. PCR products were purified with magnetic bead DNA purification system Genopure™ ds (Bruker Daltonics) according to the manufacturer's recommendations and eluted in 10µL elution buffer.

Primer Extension

Primers for SNP detection were designed with the 3' base corresponding to the one upstream the target SNP. The sequences of the primers were checked for the possibility of primer-dimer and hairpin formation and investigated in self-extension reactions (negative controls). A total of 2µL of purified PCR product were used for the PEX reaction. Ten µL PEX medium was optimized for the extension of all target detection primers with 1X Magnesium free reaction buffer C (Solis BioDyne®); 10mM MgCl₂(Solis BioDyne®); 200µM appropriate dNTP and ddNTPs; 0.5 - 1µM each detection primer; 1 unit ThermiPol DNA Polymerase (Solis BioDyne®) in 35 cycles (95°C–10sec; 51°C –30sec; 72°C –30sec) and 2min at 72°C final elongation step. Each PEX was designed to detect simultaneously 2 - 4 SNPs with no interference from dNTPs and ddNTPs (Table 1). Each PEX medium contains 4 nucleotide types and is named according to the dNTPs (for example the medium A contains dATP, ddTTP, ddCTP and ddGTP). PEX products were desalted and purified with magnetic bead DNA purification system Genopure™ oligo (Bruker Daltonics) according to the manufacturers' recommendations and eluted in 6µL elution buffer.

MALDI-TOF MS

One μL eluted product was deposited on an Anchor-ChipTM (384-400 μm) (Bruker Daltonics) with 0.5 μL 3-HPA matrix (3-hydroxy-picolinic acid 20g/L, di-ammonium-hydrogen-citrate 2g/L) according to the dried droplet method. The samples were analyzed on an UltraflexTM (Bruker Daltonics) instrument in linear positive mode. Calibration was performed with primers present in negative control samples. Mass to charge ratios were labelled using the centroid function.

Statistical analysis

Divergence from Hardy-Weinberg expectations and linkage disequilibria across loci located on the same chromosome were tested using the Genepop software (10). The Markov chain parameters were set to 10000 dememorizations, 1000 batches and 5000 iterations per batch for all tests, in order for the standard error (S.E.) of the estimates to be below 0.001.

Results

Validation of the method

The results obtained by sequencing were fully concordant with those obtained by mass spectrometry. No null alleles were observed and no additional SNPs in the PEX primer binding regions were revealed. Reproducibility was tested and consistent results were obtained by typing each SNP three times by MALDI-TOF MS for every DNA sample. Two opposite homozygous and a heterozygous sample, determined by sequencing, was included as a positive control for each experiment. Furthermore, no discrepancies were observed relative to the sex determination by the use of the selected marker.

Though in five separate reactions (table 1), the 51 fragments containing the selected SNPs were amplified in identical medium (except for primers) and thermal conditions. The same goal was achieved for the primer extension reactions by the use of primers with similar length.

Samples were amplified using varying amounts of starting DNA (cascade dilution series) and complete and unambiguous profiles were obtained with a minimum of 180pg per amplification reaction. Beneath this target DNA amount, the heterozygous balance observed as peak height equilibrium decreases due to stochastic variation. A 33% minimal threshold value for heterozygous balance with optimal amounts of template DNA was experimentally determined.

The simultaneous PEX reactions were designed to yield products differing by at least 60Da in protonated mass allowing a broad window for product labelling. In this regard, we investigated the precision of mass determination by MALDI-TOF from 96 spectral representations of each of the 153 possible products (non extended primers and alleles) by the use of the X-MassTM software (Bruker Daltonics) and the centroid labelling function. The determined masses for each analyte in 96 spectra showed a normal distribution with standard deviations below 4Da (2.03Da - 3.85Da). The numeric deviation between observed and expected mass ranged between 1.6 and 33%. Therefore a window of 12Da (corresponding to 3 standard deviations) would allow 99% correct peak labelling of each analyte. The window applied to sample analysis was 60Da allowing the false allele call risk to be ignored.

Statistical analysis

Allele frequencies were derived from the data (Table 2) and the panel of markers was characterised by a number of tests including Hardy-Weinberg equilibrium tests and exact tests for linkage disequilibrium. All the autosomal markers presented dimorphism with heterozygosity per locus values ranging between 0.354 and 0.5 with 86% of the SNPs showing a value above 0.45 and an overall mean heterozygosity of 0.474.

The studied population sample revealed no significant deviation from Hardy-Weinberg equilibrium as the calculated p values were significantly above 0.05 (S.E.<0.001), indicating that the SNP loci were behaving as expected within a randomly mating population.

Furthermore, no linkage disequilibria among loci carried by the same chromosome were suggested by genotyping these samples ($p > 0.05$, S.E. < 0.001).

The probability of identity obtained by the analysis of this panel of SNPs is $3.75 \cdot 10^{-21}$ and the discriminatory power is on the par with the routinely used STR multiplexes. The highest observed 50-SNP-profile frequency was $2.90 \cdot 10^{-21}$ when the restriction due to the sex determination was not taken into account.

No mutations through generations were observed and the parentage relationships previously investigated by means of STRs were confirmed. Exclusions of biologically unrelated alleged fathers were based on at least 7 genotypes inconsistent with paternity. One exclusionary case was based on only 4 SNPs and therefore we decided to test alternative relationship hypotheses. We obtained a likelihood ratio exceeding $5 \cdot 10^{12}$ in favour of an uncle-nephew relationship. This relation was confirmed by posterior information. When investigating inclusionary paternity the likelihood ratios obtained and combined with an a priori probability of 50% led to probabilities ranging between 99.983% and 99.999997%.

Discussion

In this study we have evaluated the typing and the information content of a panel of SNPs by primer extension and MALDI-TOF MS for forensic studies and potentially for anthropological investigations on a French representative population.

Our results corroborate the polymorphism of the 50 selected autosomal markers, their independent inheritance and their ability to provide individual specific genetic profiles. However, increasing DNA sample number would be necessary in order to assess the specificity of the profiles that we have obtained.

The selected autosomal markers gave clear results in paternity and identification investigation with genotype derived numerical evaluations comparable to those obtained by existing STR

multiplexes (2). In this paper we excluded the presence of null alleles, as all analysis reactions as well as sequencing fragments containing the amplification primer binding sites and the PEX analysis, repeated in triplicate, brought the same and reproducible results. Furthermore, the SNP mutation rates can be ignored in practice. However, assuming that, as SNP and STR analysis systems are both PCR based, the frequency of appearance of null alleles can be considered as common (11) and ranging between 10^{-3} and $5 \cdot 10^{-3}$ (12, 13, 14). Taking into consideration the allelic frequencies of the 50 SNPs in the panel studied here, the probability to observe such an incompatibility at a locus when studying a trio where the alleged father is the biological father reaches the value of 0.111. This value is almost twofold the one observed when analysing a 15-locus-system such as the common STR multiplexes (0.063) and therefore "covers" the mutation and the null allele appearance rates of the latter. Furthermore, these calculations may also underestimate frequency of null allele appearance in the case of SNPs due to the existence of an additional step in their analysis. Indeed, whereas STRs are detected directly as PCR products SNPs often require an allelic discrimination step such as the PEX reaction applied here. This reaction is built on an additional primer hybridization where not only *in vivo* occurred mutations but also *in vitro*, in early cycle step Taq induced nucleotide misincorporations, may cause null alleles. No matter how low the Taq error rate is, it should be taken under consideration. Therefore, though SNP mutation rates are very low, we deemed that an incompatible result at one locus might occur as the result of a mutation, or due to the presence of a null allele, whereas 3 exclusionary loci (PI=0) are sufficient to prove a non-paternity (15).

Ternary encoding of genotype data, whose effectiveness for the straightforward identification of individuals has been previously demonstrated (16, 6), is compatible with any SNP analysis platform. The combination of single-point genomic markers detectable from degraded DNA, with specific multiplexed PEX detection and very sensitive, precise and rapid MALDI-TOF

MS analysis brings a powerful method of DNA investigation for forensic and anthropological applications. The primer extension major advantage is that the distinction between the variants relies on the high accuracy of the incorporated nucleotides rather than on differences in thermal stability between matched and mismatched probes. PEX assays give rise to allele-specific products which can be separated according to their intrinsic differences in mass without any labeling by MALDI-TOF MS. This allows excellent discrimination between the homozygous and heterozygous genotypes and the assays are robust and tolerant to small variations in the reaction conditions. Therefore, it is the most direct method of detection and provides more robust signal-to-noise ratios when compared to the other assay formats that infer the identity of the products by monitoring the fluorescence emitted by labelled molecules (17). The majority of PEX products being monoprotonated (18), the observed value is virtually m (m/z , with $z = +1$). Minisequencing has become the most widely used MALDI-TOF MS based method for SNP analysis. Despite the fact that the smallest mass difference of 9Da induced by incorporation of ddNTPs only can be detected (19), it is difficult to distinguish between A/T, A/A and T/T genotypes (20). Therefore we chose to incorporate nonlabelled ddNTPs complementary to the nucleotide corresponding to one allele of the SNP and to the first 3' position following the SNP, and nonlabelled dNTP complementary to the nucleotide corresponding to the other SNP allele. This led to a time of flight separation not only according to the composition of PEX products but also according to their lengths (21). This technique avoids labeling and enzymatic shortening of the products as it allows working in a high mass range and the sensitivity necessary to distinguish between alleles is lowered. The mass determination accuracy (i.e. the distributions of the determined mass in relation to the expected mass) of the used method allowed us to work in non-overlapping allele windows and therefore face quasi inexistence false allele call risk. The achievement of the same accuracy when products are detected by capillary electrophoresis requires a several

nucleotides length difference between two consecutive products. This variant of the classical PEX reaction has also an advantage in specificity over other modified-PEX based methodologies. Indeed, it relies on the detection of both alleles, unlike allele dependent extension (22) where only one allele of a locus is physically detected per reaction, which implies genotype deduction or requires double reactions per locus corresponding to both alleles. However, the method has two major drawbacks; the specificity of the medium limits the multiplexing degree and disables the detection of a newly arisen allele.

The high throughput that is possible for SNP genotyping in an automated MALDI-TOF MS system has been demonstrated, and the major challenge in application of this method to forensic casework lies in multiplexing amplification reactions. Although this method allows analyzing such a set of SNPs in very short time, forensic samples generally do not enclose enough template-DNA. The increase of the degree of PCR multiplexing would lessen the required starting DNA amount. MALDI-TOF MS can theoretically allow simultaneous detection of more than 80 SNPs in the 5500-10500Da range (23). Therefore further multiplexing of the PEX assays yielding SNP specific products with non-overlapping mass distributions would also significantly reduce manipulation steps (splitting PCR products in several reactions), as well as time and consumable expenses.

Conclusions

The multiplex PCR SNP typing format presented here could be useful for forensic and anthropological investigations because degraded DNA templates can be typed. However, it is necessary to either increase the number of SNP containing fragments amplified simultaneously in the PCR step, or to apply an allelic discrimination reaction which does not require pre-amplification in order to reduce the required template DNA amount. At this stage SNPs are a solution for facing template DNA degradation by the analysis of short fragments,

but the probability of accessing intact marker containing fragments in a sample depends on the global DNA quantity it contains. Reaching sufficient discriminatory power by means of SNP based investigations requires a high number of markers to be analyzed which increases the necessary template amount. However, SNP markers should be considered as a complement to STR analysis rather than as an alternative. Typing a set of SNPs on a sample whose degradation state allowed a partial STR profile only, would allow the increase of the discriminatory power to a significant value. In this context the number of SNPs to be typed would be less important. The chosen set of SNPs assents to Hardy–Weinberg expectations and shows no linkage disequilibrium and is adaptable to routine and high throughput forensic and anthropological work. The SNP data obtained in this study offer an accurate and French population specific database that will be an essential reference for SNP-based forensic application in the future.

Acknowledgements

We gratefully thank the people who made it possible for us to run our investigations on their DNA by giving their informed consent prior to their inclusion in the study.

References

1. Brookes AJ. The essence of SNPs. *Gene* 1999;234:177-86.
2. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 2001;114:204-10.
3. Li J, Butler JM, Tan Y, Lin H, Royer S, Ohler L, et al. Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 1999;20:1258-65.
4. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000;156:297-304.
5. Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 2003;21:12-27.
6. Petkovski E, Keyser-Tracqui C, Hienne R, Ludes B. SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. *J Forensic Sci* 2005;50:535-41.
7. Sun X, Ding H, Hung K, Guo B. A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs. *Nucleic Acids Res* 2000;28:E68.
8. Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. Typing of multiple single-nucleotide polymorphisms using ribonuclease cleavage of DNA/RNA chimeric single-base extension primers and detection by MALDI-TOF mass spectrometry. *Anal Chem* 2005;77:5229-35.
9. Krawczak M. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 1999;20:1676-81.
10. <http://wbiomed.curtin.edu.au/genepop/>
11. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 2005;150:17-21.

12. Clayton TM, Hill SM, Denton LA, Watson SK, Urquhart AJ. Primer binding site mutations affecting the typing of STR loci contained within the AMPFISTR SGM Plus kit. *Forensic Sci Int* 2004;139:255-9.
13. Ballard DJ, Phillips C, Wright G, Thacker CR, Robson C, Revoir AP, Court DS. A study of mutation rates and the characterisation of intermediate, null and duplicated alleles for 13 Y chromosome STRs. *Forensic Sci Int* 2005;155:65-70.
14. Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G et al. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. *Forensic Sci Int* 2003;133:220-7.
15. Brenner C. <http://dna-view.com/>
16. Griffin TJ, Smith LM. Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms ternary encoding of genotypes. *Anal Chem* 2000;72:3298-302.
17. Sobrino B, Brion M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 2005;154:181-94.
18. Karas M, Gluckmann M, Schafer J. Ionization in matrix-assisted laser desorption / ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom* 2000;35:1-12.
19. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* 1997;7:378-88.
20. Fei Z, Ono T, Smith LM. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res* 1998;26:2827-8.
21. Pusch W, Kraeuter KO, Froehlich T, Stalgies Y, Kostrzewa M. Genotools SNP manager: a new software for automated high-throughput MALDI-TOF mass spectrometry SNP genotyping. *Biotechniques* 2001;30:210-5.

22. Blondal T, Waage BG, Smarason SV, Jonsson F, Fjalldal SB, Stefansson K et al. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 2003;31:e155.
23. Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N. MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 Y-chromosome SNPs. *Anal Chem* 2004;76:6039-45.

Table 1: Selected SNPs sorted according to the PCR and PEX groups they were analysed in.

The primer orientations are indicated by an f for forward and an r for reverse.

PCR group	PEX group	dbSNPrs	PEX primer 5'-3' sequence	PCR group	PEX group	dbSNPrs	PEX primer 5'-3' sequence
1	A-1	rs910170	r TGAACCATTTCTCCTGGAC	4	C-1	rs874746	f GTGCCAGACCGCTTCTTG
		rs997556	r ACACTGGATAACCAAGGAC			rs240	r CAGAAGCAATTACAGGAAG
		rs1674139	r GAGCTGACTTGCCGCC		G-1	rs763443	r AAGAGACCTGTAGGAGCG
		rs1003204	r GTGTAGTGGGAGCACTGG			rs820129	f ATAAAGAAAGAACCTGGAGAC
	A-2	rs1371370	r GCAGGGGGAGTCCTTAAC		rs225	r CCAAGAACTACATTTGGGGAG	
		rs235	r GTATACTTATTACTTCTCCCT			G-2	rs936250
	rs243	r GGCTGGTGGAGGCAGA	rs757655		f AATTGATGTTGGGCTCTGG		
	A-3	rs222	r AGAGGAGAGTAAAAACAGC		rs228047	r CGCACCCCGCAGACCT	
		rs713053	f CAACAACAGTGAAGGATC		rs2260	r CAAGAAATAAAAGACAAGTCC	
		rs767453	r CTCTGTCCACCTTTCC			G-3	rs234
		rs727206	r CTGTTTTCTGTTATATTCTGCT		rs997750		r GTAGATAGAATTGAATAGAAGT
	2	T-1	rs882937		r AACTAATGGGCACCCTTC	rs833069	f CAAACCAGGGAAGGGAGG
rs738518			f CCGCTGTGAGACAAAGGG	A/T	rs1924609	f TCTATTCAGCAGGCACTTA	
rs892503		f GCAGCATACTCATTACAGA	rs737612		r CAGGTAGGGTCACAGTTTC		
T-2		rs237	r AGGTAGCCACAGAAGC	rs279536	f ATTGTCAGTCTCAAAGCACA		
		rs2123	r TAGGTAGACATGATCAGGG		rs241	f CAGAAAATATGATTAGAAAAAG	
T-3		rs991191	r GTTGTTAATGAGGATGATGC	rs544021	r GTCAGCGGAGCCTCTC		
		rs1020636	f CAGGCAGGTGGTCAGGC		rs768058	r GCTGTTCAAATTCCTCTCTC	
		rs230	r GTGATGCTCTACCCTACT	rs724784	f ATCTTAGCCTGTACCTTTA		
		rs276922	r GAGAGAACAATTAACCTCAC		rs948438	r TGAGAATGTGGACCTGAGG	
3		T-4	rs2013526	f AGTTCAAATATACTAACTATTC		rs964224	f AGTAGGCTTTTCATCAATTAG
	rs1551995		f AGGCTACTCTTCTTCTTAGG	Amel	r CTGCAGGGTGTGAGTC		
	rs1793286	f GAAATGACAACCTGGTGAAAT	rs227	r AACAGATCCATCACAATCC			
	rs889012	r GCATCAAAAACAAAGGCTCAA		T-5	rs811	r GAGTCATGGAGGGTCATTA	
	rs1019264	f TAGAACCAATCAAGGCTCC	rs725		r CAAATGGTGCCATAGTATAA		
	rs1013582	r CTCTCCTATCCATTTCCAC					

Table 2

Features of selected autosomal SNPs sorted according to their genomic position. / stands when a single SNP is selected on a chromosome, Δ for the distance to the closest SNP and f for allelic frequencies.

dbSNPrs	G- band	Alleles	f	Δ (Mbp)	dbSNPrs	G- band	Alleles	f	Δ (Mbp)
rs991191	1p31	A/G	0.573/0.427	101.4	rs237	9q33	A/G	0.56/0.44	18.7
rs2013526	1q24	C/T	0.534/0.466	101.4	rs936250	9q34	G/C	0.23/0.77	18.7
rs1019264	2p16	C/T	0.637/0.363	38.7	rs997750	10p12	C/T	0.486/0.514	50.2
rs1020636	2p22	T/C	0.527/0.473	26.8	rs241	10q22	A/C	0.5/0.5	20.1
rs222	2q11	T/C	0.747/0.253	26.8	rs243	10q24	T/C	0.507/0.493	20.1
rs1551995	3p26	C/T	0.527/0.473	6.5	rs544021	11q13	C/G	0.567/0.433	27.5
rs227	3p26	G/A	0.673/0.327	188.2	rs882937	11q21	A/G	0.447/0.553	21.2
rs225	3q29	C/G	0.587/0.413	188.2	rs1013582	11q23	A/T	0.653/0.347	17.3
rs763443	4q22	C/T	0.513/0.487	25.1	rs1793286	11q25	A/T	0.527/0.473	17.3
rs279536	4q26	C/T	0.623/0.377	57	rs2123	12q14	G/A	0.713/0.287	27.9
rs230	4q34	G/A	0.56/0.44	57	rs811	12q21	A/G	0.653/0.347	27.9
rs724784	5q21	A/C	0.553/0.447	15.6	rs1924609	13q14	C/T	0.367/0.633	/
rs1371370	5q23	T/C	0.466/0.534	10.9	rs757655	14q24	A/G	0.61/0.39	/
rs889012	5q31	C/A	0.587/0.413	10.9	rs767453	15q23	A/T	0.601/0.399	/
rs910170	6p21	C/T	0.647/0.353	4.7	rs727206	17q12	C/T	0.582/0.418	42.2
rs833069	6p21	A/G	0.6/0.4	118.7	rs820129	17q25	A/G	0.64/0.36	5.3
rs713053	6q26	G/A	0.413/0.587	118.7	rs276922	18q12	A/C	0.52/0.48	6.2
rs997556	7p12	T/C	0.616/0.384	44.3	rs948438	18q12	A/G	0.493/0.507	26
rs2260	7q21	T/C	0.6/0.4	9.1	rs768058	18q21	G/A	0.74/0.26	26
rs234	7q22	C/T	0.533/0.467	35.2	rs1674139	19q13	C/T	0.513/0.487	/
rs725	7q34	A/G	0.673/0.327	35.2	rs1003204	20p11	C/T	0.445/0.555	34.6
rs235	8q21	C/T	0.593/0.407	24.9	rs964224	21q21	T/G	0.513/0.487	15.3
rs892503	8q22	C/T	0.603/0.397	37.1	rs228047	21q22	C/T	0.553/0.447	15.3
rs874746	8q24	A/C	0.658/0.342	37.1	rs738518	22q12	T/C	0.712/0.288	6.5
rs240	9q31	C/G	0.607/0.393	9.9	rs737612	22q13	A/G	0.439/0.561	6.5

Validation of SNPs as markers for individual identification.

Petkovski E, Keyser C, Ludes B, Hienne R.

Progress in Forensic Genetics 9; ICS 1239 (2003) 33– 36

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Validation of SNPs as markers for individual identification.

Petkovski E, Keyser C, Ludes B, Hienne R.

International Congress Series, 2003, Vol. 1239, Pages 33– 36

Pages 33 à 36 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :
[http://dx.doi.org/10.1016/S0531-5131\(02\)00517-4](http://dx.doi.org/10.1016/S0531-5131(02)00517-4)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

SNPs: tools for individual identification.

Petkovski E, Keyser-Tracqui C, Niemeyer D, Hienne R, Ludes B.

Progress in Forensic Genetics 10; ICS 1261 (2004) 21– 23

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

SNPs: tools for individual identification.

Petkovski E, Keyser-Tracqui C, Niemeyer D, Hienne R, Ludes B.

International Congress Series, 2004, Vol. 1261, Pages 21– 23

Pages 21 à 23 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :
[http://dx.doi.org/10.1016/S0531-5131\(03\)01721-7](http://dx.doi.org/10.1016/S0531-5131(03)01721-7)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

Detection of genetic variation by MALDI-TOF mass spectrometry: rapid SNP genotyping using the GENOLINK system.

Niemeyer D, Elssner T, Fahr K, Peters D, Wenzel T, Petkovski E, Kostrzewa M.

Progress in Forensic Genetics 10; ICS 1261 (2004) 9– 11

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Detection of genetic variation by MALDI-TOF mass spectrometry: rapid SNP genotyping using the GENOLINK system

D. Niemeyer, T. Elssner, K. Fahr, D. Peters, T. Wenzel, **E. Petkovski** and M. Kostrzewa

International Congress Series, 2004, Vol. 1261, Pages 9– 11

Pages 9 à 11 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :
[http://dx.doi.org/10.1016/S0531-5131\(03\)01836-3](http://dx.doi.org/10.1016/S0531-5131(03)01836-3)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

Does the Tat polymorphism originate in northern Mongolia?

Keyser-Tracqui C, Blandin P, Ricaut FX, Petkovski E, Crubézy E, Ludes B.

Progress in Forensic Genetics 10; ICS 1261 (2004) 325– 327

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Does the Tat polymorphism originate in northern Mongolia?

C. Keyser-Tracqui, Blandin, F. X. Ricaut, **E. Petkovski**, E. Crubézy and B. Ludes

International Congress Series, 2004, Vol. 1261, Pages 325– 327

Pages 325 à 327 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :
[http://dx.doi.org/10.1016/S0531-5131\(03\)01701-1](http://dx.doi.org/10.1016/S0531-5131(03)01701-1)

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

MALDI-TOF MS analysis of Y-SNPs in ancient samples.

Petkovski E, Keyser-Tracqui C, Crubézy E, Hienne R, Ludes B.

Progress in Forensic Genetics 11; ICS (2005) ; sous presse.

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

MALDI-TOF MS analysis of Y-SNPs in ancient samples.

Petkovski E, Keyser-Tracqui C, Crubézy E, Hienne R, Ludes B.

International Congress Series, 2006, Volume 1288, Pages 25-27

Pages 25 à 27 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :
<http://dx.doi.org/10.1016/j.ics.2005.12.008>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

Polymorphismes ponctuels de séquence (SNP) et identification génétique. Etude par spectrométrie de masse MALDI-TOF.

L'investigation des polymorphismes du génome humain permet d'identifier les individus avec certitude et d'établir des relations de parenté. L'étude des polymorphismes ponctuels de séquence (SNP) présente un réel intérêt dans les domaines de la médecine légale et de l'anthropologie moléculaire. En effet, les SNP autosomaux réunissent les avantages des méthodes appliquées aujourd'hui et permettent d'atteindre un haut pouvoir discriminant à partir d'échantillons dégradés par l'analyse de fragments d'ADN courts.

Nous avons sélectionné 50 SNP autosomaux et une différence de séquence entre les copies du gène de l'amélogénine porté par les gonosomes. La définition de cet ensemble de marqueurs est un travail original tant d'un point de vue quantitatif, menant à un pouvoir de discrimination élevé, que d'un point de vue qualitatif, ne considérant que des marqueurs non codants en accord avec la législation française en vigueur. Les marqueurs ont été étudiés par détection des produits d'extension allèles spécifique d'amorce par spectrométrie de masse MALDI-TOF. L'étude d'un échantillon de la population française a permis de fournir des informations sur la distribution allélique de ces marqueurs, de valider l'ensemble des SNP sélectionnés comme outil pour la filiation et l'identification génétique et de développer une méthode d'analyse directe, sensible et rapide permettant l'analyse simultanée d'un grand nombre de molécules de manière reproductible.

L'analyse des marqueurs bialléliques validés représente un outil complémentaire à celle des microsatellites dans le cadre d'échantillons fortement dégradés tels qu'ils sont fréquemment rencontrés sur les scènes de crimes ou lors de catastrophes de masse. La définition d'un protocole de routine permettra d'introduire cette technique basée sur les marqueurs SNP au niveau des laboratoires d'identification génétique.

Single nucleotide polymorphisms (SNPs) and genetic identification. Study by MALDI-TOF mass spectrometry.

Human genome polymorphism investigation allows accurate individual identification and genetic relationship establishment. The study of single nucleotide polymorphisms (SNPs) requires short DNA fragments and therefore has a particular advantage over classical markers in the analysis of degraded samples. This and the capacity of yielding high discriminatory powers confer a great value to autosomal SNP markers in the fields of forensics or molecular anthropology.

In the present study 50 autosomal SNPs and a sex determining sequence difference between the amelogenin gene gonosomal copies were selected. The characterization of this set of markers represents an innovative work as it allows generating strong discriminatory information and is restricted to non-coding DNA regions, in harmony with the in force French legislation. Our approach to SNP typing is a multiplex PCR based amplification followed by simultaneous detection of primer extension products by MALDI-TOF mass spectrometry.

The study of these markers in a French representative population allowed their allelic distribution investigation, their validation as tools for genetic identification and filiation and the development of a direct, sensitive, rapid and multiplexed analysis method yielding reproducible results.

The analysis of the selected binary markers represents a complementary means of great help in cases where microsatellite investigation fails due to extensive DNA degradation rather than lack of DNA template. Their specific advantage relies in the identification of discrete samples, such as highly degraded tissues commonly encountered on crime scenes or in mass-disasters. The establishment of a routine protocol will lead to the implementation of the method based on SNP typing in genetic identification laboratories.