

Université
de Toulouse

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Anthropologie moléculaire

Présentée et soutenue par :

Fanny MENDISCO

le : 13 décembre 2011

Titre :

Apports de la paléogénétique à l'histoire du peuplement précolombien
des Andes méridionales (Vème - XVème siècles).

Ecole doctorale :

Biologie, Santé, Biotechnologies (BSB)

Unité de recherche :

UMR 5288, Anthropologie Moléculaire et Imagerie de Synthèse (A.M.I.S)

Directeur(s) de Thèse :

Christine KEYSER, Professeur à l'Université de Strasbourg

Bertrand LUDES, Professeur à l'Université de Strasbourg

Rapporteurs :

Eva-Maria GEIGL, Directeur de recherche, Institut Jacques Monod, Paris

Peter FORSTER, Directeur de recherche, Université de Cambridge

Membre(s) du jury :

Michel BRUNET (président du jury), Professeur au Collège de France

Eric CRUBEZY (examineur), Professeur à l'Université Paul Sabatier, Toulouse

Georges LARROUY (examineur), Professeur émérite à l'Université Paul Sabatier, Toulouse

Remerciements

En tout premier lieu, je voudrais exprimer toute ma reconnaissance à mes directeurs de thèse, les professeurs Christine Keyser et Bertrand Ludes, sans qui ce travail n'aurait pu être mené à bien. J'adresse toute ma gratitude au Professeur Bertrand Ludes, pour m'avoir permis de réaliser cette thèse au sein de son laboratoire dans d'excellentes conditions, ainsi que pour son soutien et ses encouragements dans les moments les plus importants. Mes sincères remerciements s'adressent au Professeur Christine Keyser qui m'a encadrée et accompagnée tout au long de ces trois années de thèse, qui se sont révélées extrêmement enrichissantes. Votre disponibilité, votre patience ainsi que votre gentillesse et votre bonne humeur quotidienne m'ont permis d'avancer et d'acquérir la confiance nécessaire à l'aboutissement de ce travail de thèse.

Je souhaite également remercier tout particulièrement le Professeur Eric Crubézy, pour m'avoir accueillie au sein de son équipe, et pour m'avoir fait confiance et permis de continuer ce travail sur l'Amérique du Sud dont il est l'un des initiateurs.

Je remercie chaleureusement l'ensemble des membres du jury pour avoir accepté de juger ce travail, et particulièrement Eva-Maria Geigl, et Peter Forster pour avoir accepté d'en faire un rapport et avoir apporté leur vision sur ce sujet. Je remercie sincèrement les professeurs Michel Brunet, Eric Crubézy et Georges Larrouy d'avoir accepté de participer à ce jury de thèse.

Je voudrais profiter de l'occasion qui m'est donnée au travers de la présentation de ce travail pour adresser ma reconnaissance au Professeur Michel Brunet m'a fait partager sa passion et ses connaissances au cours d'une mission au Chili, ce qui fût extrêmement instructif et enrichissant.

Ce travail de thèse n'aurait pas été possible sans échantillons, je remercie donc sincèrement toutes les personnes qui ont œuvré pour obtenir les échantillons les plus intéressants possibles :

- Les Professeurs Carnese et Nielsen de l'Université de Buenos Aires, qui nous ont permis d'accéder à des collections si précieuses.

- Veronica Seldes, anthropologue argentine, qui a réalisé une partie des prélèvements et qui a été une interlocutrice patiente et intéressée tout au long de ce travail.
- Le docteur Marc Hubbe, qui nous a permis d'étudier les collections du musée Gustavo Le Paige de San Pedro de Atacama et a réalisé les prélèvements sur ces collections.
- François-Xavier Ricaut pour son implication dans le projet et pour m'avoir permis de participer à ses projets de recherche et Jean-Michel Dugoujon pour nous avoir fourni des échantillons contemporains qui nous ont permis de réaliser la mise au point de certaines manipulations.

Le travail en laboratoire est tout aussi important que l'accès aux échantillons, ainsi, je remercie Prisca et Aurélie pour m'avoir formée et accompagnée tout au long de ce travail. Angela je te remercie pour ton aide concernant toute la partie sur la spectrométrie de masse. Ta bonne humeur et ton optimisme ont permis de surmonter tous les obstacles que nous avons pu rencontrer. Enfin, je remercie grandement Daniel Montagnon pour son aide précieuse et sa grande patience en ce qui concerne les analyses statistiques.

Caroline, Marie, Clémence, vous avez contribué largement à la réussite de ce travail. Merci pour m'avoir transmis vos connaissances, m'avoir écoutée et épaulée à chaque doute et pour avoir contribué très largement à la bonne ambiance de travail au sein de notre équipe, indispensable pour mener à bien un tel travail.

Je souhaite remercier l'ensemble de l'équipe AMIS ainsi que tout le personnel de l'institut de médecine légale de Strasbourg qui ont participé, tous à leur manière, au bon déroulement de ces trois années.

Mes dernières pensées vont à ma famille et plus particulièrement à mes parents, mes sœurs, mais aussi à tous les membres du G8 (Anaiz, Anna, Maiana, Mairer, Magalie, Oiana et Pia ; par ordre alphabétique !) et bien sûr à Brice : je voudrais simplement vous dire à tous merci d'être là.

Résumé

L'aire andine abrite le berceau de civilisations parmi les plus extraordinaires et complexes d'Amérique, suscitant encore aujourd'hui un très grand engouement pour l'histoire de ces civilisations. Les populations andines précolombiennes, restées isolées du reste du monde pendant près de 15 000 ans, ont accumulé des connaissances sur leur environnement et inventé toute une série de techniques d'exploitation du milieu naturel, leur permettant de diversifier leur organisation sociale jusqu'au développement d'états et d'empires, dont le plus marquant est probablement l'empire Inca. L'origine de ces populations, leur évolution ou encore leur organisation sur le territoire restent des problématiques largement débattues, d'autant plus que toutes les régions de l'aire andine n'ont pas atteint le même degré de développement. Ainsi, l'histoire des populations de la région andine méridionale (nord du Chili, nord-ouest de l'Argentine et extrême sud de la Bolivie) est méconnue.

Le principal objectif de ce travail de thèse était de reconstruire la dynamique de peuplement des Andes méridionales grâce à une étude diachronique (entre le V^e et le XV^e siècle de notre ère) de groupes anciens issus de différents étages écologiques de la région (hauts-plateaux, vallées tempérées et piedmont andin). Pour appréhender cette problématique, nous avons choisi d'étudier directement des populations anciennes des Andes méridionales, grâce aux outils de la paléogénétique. L'étude de l'ADN conservé au sein de restes humains anciens permet d'accéder à des informations inestimables concernant la composition biologique de populations du passé, et ainsi, d'obtenir des informations aussi bien sur la structure sociale (recrutement funéraire, liens de parenté), que sur la structure génétique de populations anciennes. Par conséquent, ce travail a été axé, dans un premier temps, sur l'étude de l'origine des populations des Andes méridionales et sur la compréhension de l'organisation des populations anciennes sur ce territoire. Puis, dans un deuxième temps, grâce à une comparaison avec des données génétiques d'autres populations amérindiennes, nous avons pu appréhender des problématiques plus générales telles que la compréhension de l'histoire évolutive de l'aire andine et de la colonisation de l'Amérique du Sud.

Des analyses génétiques ont été menées sur des échantillons (fragments d'os longs ou dents) correspondant à 103 individus anciens. Pour cela, une approche originale, combinant l'étude de différents marqueurs génétiques, situés aussi bien sur le génome mitochondrial que nucléaire, à l'aide de méthodes classiques (séquençage, analyse de fragments) et innovante (spectrométrie de masse MALDI-TOF), a été mise en place.

Nous avons pu déterminer les lignées maternelles, les lignées paternelles ainsi que les profils génétiques (plus ou moins partiels) pour la plupart des individus anciens analysés (entre 50 % et 70 %, suivant le marqueur et/ou la technique analytique employé). A partir de ces résultats, des analyses ont été effectuées afin d'évaluer la diversité intra-populationnelle des populations anciennes de la région mais aussi d'estimer leur distance génétique avec environ

70 populations amérindiennes contemporaines et anciennes.

Nous avons pu démontrer que si d'un point de vue biologique une certaine unité se dégage de l'ensemble de la région, laissant supposer une origine commune de l'ensemble des groupes de la région, d'un point de vue culturel, des disparités sont évidentes, notamment au niveau des pratiques funéraires. En effet, alors que pour les populations des vallées tempérées à l'époque tardive les *ayllus* (clans) revêtaient une importance capitale aussi bien dans la vie sociale que pour le traitement des défunts, les groupes du piedmont andin, à la période céramique initiale ne semblaient pas prendre en compte les liens de parenté lors de l'inhumation de leurs morts. La comparaison de ces résultats avec les données génétiques disponibles concernant l'aire andine démontre que les habitants des Andes méridionales semblent former une entité homogène légèrement différenciée du reste du complexe andin. Ainsi, même si nous ne pouvons pas rejeter que les populations des aires andines centrales et méridionales possèdent une origine commune, il semble que ces deux régions aient évolué de manière indépendante au cours du temps, sans migrations massives interrégionales. Il est donc possible d'imaginer, comme cela a déjà été proposé sur la base de données archéologiques, que l'influence des grandes civilisations (notamment Tiwanaku) n'ait été qu'indirecte dans cette région méridionale. Seul l'empire Inca a eu un impact significatif sur le pool génétique des populations autochtones. En effet, la structure génétique actuelle des populations de cette région diffère sensiblement de la diversité précolombienne, ce qui peut s'expliquer par l'expansion des Incas, et les relocalisations de populations qu'ils ont initié. De manière générale, la structure génétique actuelle de l'ensemble des populations de l'aire andine ne semble pas être le résultat de la colonisation initiale, mais plutôt le résultat d'un renouvellement, débuté avec l'expansion Inca et accentué par la colonisation européenne.

Ce travail de thèse démontre que la paléogénétique permet de retracer des événements de migrations et/ou l'histoire évolutive de populations anciennes de manière plus nuancée qu'à partir des données génétiques contemporaines. Ainsi, l'analyse de l'ADN conservé au sein d'individus précolombiens nous a permis de présenter des données originales permettant la reconstruction au moins partielle de l'histoire des populations des Andes méridionales. De plus, ce travail a permis de révéler au sein de notre échantillon de nombreux haplotypes encore non décrits à ce jour. Il est probable que seule une petite partie de la diversité génétique de populations précolombiennes soit connue à ce jour. Il apparaît ainsi évident que de nombreuses études paléogénétiques sont encore nécessaires pour retracer finement le processus de colonisation de l'Amérique du Sud.

Abstract

The Andean area is the home of civilizations among the most extraordinary and the most complex of America, generating still a huge craze. The pre-Columbian Andean populations, which remained isolated from the outside world during almost 15,000 years, have accumulated knowledge and invented numerous techniques, allowing them to diversify their social organization to the development of states and empires. The origin of these populations, their development or their organization in the territory are issues widely debated, especially since all parts of the Andean area have not reached the same level of development. Thus, the population history of the southern Andean region (northern Chile, northwestern Argentina and southernmost Bolivia) is unknown.

The main goal of this thesis was to reconstruct the peopling of the southern Andes, thanks to a diachronic study (between the 5 and 15 centuries AD) of ancient groups from different ecological zones of the region (highlands, temperate valleys and Andean foothills). To address this issue, we chose to directly study ancient populations of the southern Andes, using palaeogenetic tools. The study of DNA preserved in ancient human remains makes it possible to gain access to valuable information about the biological composition of past populations, and thus to obtain information both on the social structure (e.g. funeral recruitment, family relationships), and on the genetic structure of ancient populations. Therefore, at first, this work focused on the study of the origin of the populations of the Southern Andes and on the understanding of the organization of the ancient populations on this territory. Then, thanks to a comparison with genetic data obtained on other Amerindian populations', we were able to study more general issues such as the understanding of the evolutionary history of even the Andean area or the colonization of South America.

Genetic analyses were carried out on biological samples consisting in fragments of long bones or teeth from 103 ancient individuals. An original approach, combining the study of various genetic markers (located on the mitochondrial and the nuclear genomes), using conventional (DNA sequencing, fragment analysis) and innovative (MALDI-TOF mass spectrometry) methods, was established.

We were able to determine the maternal lineages, the paternal lineages as well as the genetic profiles (more or less partial) for the majority of the ancient samples analyzed (between 50 % and 70 %, depending on the marker and/or the analytical method used). From these results, analyses were performed to assess the genetic diversity of the ancient populations of the region studied but also to estimate their genetic distance with about 70 ancient and contemporary Amerindian populations.

On the one hand, the analyses revealed that all the studied populations are genetically homogeneous, suggesting a common origin of all the groups of the region. On the other hand,

from a cultural point of view, disparities are evident, in particular in terms of burial practices. Indeed, for the populations of the temperate valleys, during the late period, *ayllus* (clans) were of crucial importance in the social life and also for the funeral practices. On the contrary, for the groups of the Andean foothills, during the initial ceramic period, the dead people were not grouped according to their familial relationships.

The comparison of these results with the available genetic data concerning the Andean area demonstrates that the inhabitants of the southern Andes seem to form a homogeneous entity slightly differentiated from the rest of the Andean complex. So, even if we cannot reject that the populations of the central and southern Andean areas have a common origin, it seems that these two regions evolved in an independent way over time, without massive interregional migrations. It is thus possible to imagine, as it was already proposed on the basis of archaeological data, that the influence of the Andean civilizations (in particular Tiwanaku) was only indirect in this southern region. Only the Inca Empire had a significant impact on the gene pool of the autochthonous populations. Indeed, the current genetic structure of the populations of this region differs significantly from the pre-Colombian variability, which can be explained by the expansion of the Incas, and the relocations of populations that they initiated. In a general way, the genetic structure of all the current populations of the Andean area does not seem to be the result of the initial colonization. It rather seems to be the result of a replacement of population, which began with the Inca expansion and which was increased by the European colonization.

This thesis demonstrates that palaeogenetic analyses allow to retrace events of migrations and/or the evolutionary history of ancient populations in a more differentiating way than from contemporary genetic data. The analysis of the DNA preserved within pre-Colombian individuals allowed us to present original data leading to the reconstruction, at least partial, of the history of the populations of the southern Andes. Furthermore, this work revealed in our sample many haplotypes still not yet described. It is likely that only a small part of the genetic diversity of pre-Colombian populations is known today. Numerous palaeogenetic studies are still necessary to better understand the process of colonization of South America.

Table des matières

Remerciements	3
Résumé	5
Abstract	7
Table des matières	9
Liste des tableaux	15
Liste des figures	17
Liste des annexes.....	19
Liste des abréviations	20
Introduction générale	23
Première partie. Contexte général de l'étude.	27
I. Peuplement de l'Amérique du Sud.	29
1. Etat de la connaissance actuelle.	29
1.1. Modèle des trois vagues de migration.....	30
1.2. Données génétiques.....	31
1.2.1. Les premières études : de nombreuses hypothèses mais peu de certitudes..	32
1.2.2. Avancées technologiques et nouvelles perspectives : vers une meilleure caractérisation de la diversité génétique.....	34
1.3. Peuplement de l'Amérique du Sud : un gradient est/ouest.	40
2. L'ADN ancien un outil essentiel pour la compréhension des populations du passé....	42
2.1. Caractéristiques de l'ADN ancien.....	43
2.2. ADN ancien et peuplement de l'Amérique du Sud.....	45

II. Peuplement et histoire culturelle des Andes méridionales.	49
1. Contexte géographique.....	49
1.1. Etages écologiques et ressources associées.....	49
1.2. Mobilité des populations.	52
2. Contexte historique.	53
2.1. Contexte chronologique.	53
2.2. Du peuplement initial à la colonisation.....	55
2.2.1. Période archaïque : 10 000 – 1 000 av. J.-C.....	55
2.2.2. Période céramique initiale : 1 000 av. J.-C. – 650 ap. J.-C.	56
2.2.3. Période céramique intermédiaire : 650 – 900 ap. J.-C.	57
2.2.4. Période céramique tardive : 900 – 1480 ap. J.-C.	59
2.2.5. Phase Inca : 1480 – 1530 ap. J.-C.	61
3. Diversité biologique contemporaine dans les Andes méridionales.....	62
III. Hypothèses et objectifs.	64
Deuxième partie. Matériel et méthodes.	67
I. Sites et échantillons analysés.	69
1. Stratégie d'échantillonnage.	69
2. Situation et description des échantillons.	71
2.1. Quebrada de Humahuaca.	72
2.1.1. Zones résidentielles simples.....	75
2.1.2. Terrasses domestiques.....	75
2.1.3. Agglomérations.	76
2.2. Vallée Calchaqui.	79
2.3. Puna : Doncellas.....	82
2.4. Sierras : Pampa Grande.	82
2.5. Désert d'Atacama : Solcor 3.	83

II. Marqueurs moléculaires : à chacun son histoire.....	85
1. Les marqueurs à transmission biparentale.	86
1.1. Détermination du sexe des individus : le gène de l'amélogénine.	86
1.2. Détermination des profils génétiques : analyse de marqueurs autosomaux.....	87
2. Les marqueurs à transmission uni-parentale.	90
III. Préparation et extraction des échantillons d'ADN.....	92
1. Pré-traitement des échantillons.	92
2. Extraction de l'ADN.	93
2.1. ADN ancien.....	93
2.2. ADN moderne.	94
3. Quantification de l'ADN extrait.....	94
IV. Analyses génétiques menées.	96
1. Première région hypervariable de l'ADNmt (HVR-1).....	96
1.1. Amplification par PCR.....	96
1.2. Purification des produits PCR.....	96
1.3. Séquençage et analyse.....	97
2. Analyse des STR du chromosome Y.....	98
3. Analyse des marqueurs autosomaux.	99
3.1. STR.....	99
3.2. Petites insertions délétions (Indels).....	100
4. Analyse de SNP par spectrométrie de masse.	101
4.1. Principe général et protocole.....	101
4.2. Choix des SNP d'intérêt.....	103
V. Authentification des résultats.....	105
1. Préventions contre les contaminations.	105
1.1. Prélèvement des échantillons.	106
1.2. Précautions lors de l'analyse.....	107
1.2.1. Bonnes pratiques de laboratoire.	107

1.2.2.	Contrôles d'extraction et de réaction PCR.....	107
2.	Stratégies pour garantir l'authenticité.	108
2.1.	Réplication des analyses.....	108
2.2.	Stratégie multi-marqueurs.	108
2.3.	Cohérence phylogénétique.	109
2.4.	Comparaison des profils génétiques.....	109
VI.	Analyses des données génétiques.	110
1.	Populations de comparaison : établissement de bases de données de référence.	110
2.	Identification individuelle et détermination des liens de parenté.....	116
3.	Analyses statistiques.	116
	 Troisième partie. Résultats et analyses génétiques.	 119
I.	Caractérisation des échantillons anciens étudiés.	121
1.	Taux de génotypage et conservation de l'ADN.	121
2.	STR autosomaux et amélogénine.....	123
3.	Détermination des haplotypes.	128
3.1.	Variations de séquences de la région HVR-1 et distribution des haplotypes anciens.....	128
3.2.	Y-STR et haplotypes du chromosome Y.....	131
4.	Détermination des haplogroupes par spectrométrie de masse MALDI-TOF.	133
5.	Authenticité des résultats.	135
6.	Recrutement funéraire et détermination des liens de parenté.	136
6.1.	Recrutement funéraire du site de Pampa Grande.	136
6.2.	Cortaderas.....	138
6.3.	Los Amarillos.....	140
6.3.1.	Unité 320.....	140
6.3.2.	Unité 400.....	142
7.	Constitution des populations archéologiques.....	143
7.1.	Population des vallées tempérées.....	143

7.2.	Population des sierras.....	144
7.3.	Individus de la puna.	144
II.	Diversité génétique des populations des Andes méridionales.	145
1.	Diversité des lignées maternelles et paternelles.....	145
2.	Diversité des haplotypes mitochondriaux et du chromosome Y.....	146
III.	Comparaisons inter-populationnelles	148
1.	Distribution des haplogroupes mitochondriaux et Y en Amérique du Sud.....	148
1.1.	Comparaisons de fréquences.....	148
1.2.	Distances génétiques et analyses multidimensionnelles.	150
2.	Distribution et comparaison des haplotypes mitochondriaux et du chromosome Y en Amérique du Sud.....	153
2.1.	Comparaison de diversités génétiques.	153
2.2.	Haplotypes partagés.	155
2.2.1.	Analyses phylogénétiques.....	158
2.3.	Calculs de distances et analyses multidimensionnelles.....	161
	 Quatrième partie. Discussion.	 165
I.	Stratégie d'analyse de l'ADNa.	168
1.	Etat de conservation de l'ADN.	168
2.	Perspectives méthodologiques.	169
2.1.	Bénéfices du génotypage par spectrométrie de masse MALDI-TOF pour l'analyse de l'ADNa.....	170
II.	Populations et peuplement des Andes méridionales.	173
1.	Pratiques funéraires : le culte des ancêtres.....	173
2.	Populations de la période céramique initiale.....	175
3.	Populations de la période céramique tardive.....	178

4. Les Andes méridionales : entre disparités culturelles et homogénéité génétique.....	183
III. Apports pour la compréhension de la dynamique de peuplement de l'Amérique du Sud.....	185
1. Les Andes méridionales au sein du complexe andin.....	185
1.1. Interactions entre populations andines anciennes : l'impact des grandes civilisations sur les Andes méridionales.	187
1.2. Histoire évolutive et mise en place du complexe andin.	189
2. Populations andines vs amazoniennes : une frontière de moins en moins évidente...	191
Conclusion générale et perspectives.	195
Bibliographie.....	203
Annexes	231
Publications.....	243

Liste des tableaux

Tableau 1 : Dates de divergence (en milliers d'années) des différents haplogroupes et sous-haplogroupes fondateurs amérindiens présentés dans ce travail.....	34
Tableau 2 : Lignées maternelles (2.A) et paternelles (2.B) fondatrices des populations amérindiennes et polymorphismes associés.....	36
Tableau 3 : Récapitulatif et données générales concernant l'ensemble des études d'ADN ancien réalisées dans le sous-continent sud-américain.	48
Tableau 4 : Description des sites et des échantillons de la Quebrada de Humahuaca analysés.	73
Tableau 5 : Informations principales concernant les échantillons du site de Pampa Grande étudiés.....	83
Tableau 6 : Tableau récapitulatif des populations anciennes et contemporaines sud-américaines dont les données ont été utilisées pour réaliser les analyses statistiques concernant le génome mitochondrial.	112
Tableau 7 : Récapitulatif des populations amérindiennes contemporaines dont les données sur le chromosome Y ont été utilisées pour les analyses statistiques.	115
Tableau 8 : Taux de génotypage obtenu pour chacun des sites anciens étudiés et chacun des marqueurs génétiques analysés.	122
Tableau 9 : Profils génétiques consensus obtenus, pour les individus anciens de l'ensemble des sites archéologiques analysés, grâce à l'utilisation de deux kits complémentaires : Minifiler et Identifier + (Applied Biosystems).	126
Tableau 10 : 26 haplotypes obtenus pour les 64 individus anciens dont la région HVR-1 a été séquencée de manière reproductible. Les mutations sont notées en fonction de la séquence de référence de Cambridge (CRS) (Anderson <i>et al.</i> , 1981).....	130
Tableau 11 : Haplotypes Y obtenus pour les individus anciens masculins.	131
Tableau 12 : Résultat de la déduction des haplogroupes et de la comparaison avec la base de données en ligne Y-HRD.	132
Tableau 13 : Haplogroupes déterminés par l'analyse de SNP des régions codantes de l'ADN mitochondrial et de la portion non recombinante du chromosome Y, par spectrométrie de masse MALDI-TOF.	134

Tableau 14 : Paires d'échantillons présentant un haplotype mitochondrial et/ou Y identique pour le site de Pampa Grande.....	138
Tableau 15 : Résultats des différents marqueurs analysés pour les trois individus de Cortaderas, et valeurs des LR et pourcentages des différentes hypothèses de parenté testées.	139
Tableau 16 : LR (likelihood ratio) et probabilités calculées pour les liens de parentés supposés entre les individus inhumés au sein de l'unité domestique 320 du site de Los Amarillos.....	141
Tableau 17 : LR (likelihood ratio) et probabilités calculées pour les liens de parentés supposés entre les individus inhumés au sein de l'unité domestiques 400 du site de Los Amarillos.....	143
Tableau 18 : Pourcentage des lignées maternelles et paternelles détectées grâce à l'analyse des SNP mitochondriaux et du chromosome Y dans les deux populations analysées...	145
Tableau 19 : Paramètres de diversité calculés à l'aide du logiciel arlequin pour les populations anciennes des vallées et de Pampa Grande.....	147
Tableau 20 : Fréquences des lignées amérindiennes dans les populations amérindiennes anciennes et dans les différentes régions géographiques d'Amérique du Sud.....	149
Tableau 21 : Indices de diversité génétique et indice de neutralité, calculés à l'aide du logiciel Arlequin pour 4 populations anciennes et 39 populations contemporaines sud-américaines, à partir des variations de séquence HVR-1.	154
Tableau 22 : Répartition et nombre des haplotypes mitochondriaux partagés (HVR-1 : 16024-16380), entre les populations anciennes de Pampa Grande, des Vallées, de Doncellas, et les populations anciennes et contemporaines sud-américaines.....	156
Tableau 23 : Haplotypes Y partagés avec les populations sud-américaines. Les comparaisons ont été effectuées à partir de sept loci (haplotype minimal moins le loci DYS385).....	157

Liste des figures

Figure 1 : Schéma global de la dynamique de peuplement de l'Amérique.....	39
Figure 2 : Représentation des dégradations post-mortem subies par la molécule d'ADN.....	43
Figure 3 : Localisation des populations anciennes amérindiennes pour lesquelles des données d'ADNa ont été obtenues.	48
Figure 4 : Présentation des différents étages écologiques des Andes méridionales.....	50
Figure 5 : Périodes et traditions culturelles des différentes aires écologiques des Andes méridionales, mentionnées au cours de ce travail.	54
Figure 6 : Distribution des principales ethnies des Andes méridionales.....	61
Figure 7 : Carte représentant la situation des sites anciens étudiés.....	70
Figure 8 : Nombre et répartition des échantillons anciens analysés, en fonction de la période et de la région.	71
Figure 9 : Localisation des différents sites archéologiques de la Quebrada de Humahuaca étudiés.....	74
Figure 10 : Plan schématique du site archéologique de Los Amarillos et de l'unité domestique 400.	77
Figure 11 : Photographie du site de Tero (S Sal Cac 14) montrant l'étendue des fouilles réalisées, ainsi que les habitations contemporaines (à gauche de la photo) construites sur un des secteurs du site.	80
Figure 12 : Présentation des polymorphismes de l'ADN analysés dans le cadre de ce travail.	85
Figure 13 : Illustration de la transmission des marqueurs uni- et bi-parentaux.	88
Figure 14 : Présentation des différentes étapes des deux protocoles d'extraction testés au cours de ce travail.....	95
Figure 15 : Illustration représentant la taille des amplicons des différents marqueurs amplifiés grâce aux différents kits utilisés.	99
Figure 16 : Principe général de la technologie de génotypage iPLEX Gold® de Sequenom.	102
Figure 17 : Arbres phylogénétiques illustrant les haplogroupes mitochondriaux et Y déterminés, les marqueurs typés ainsi que les polymorphismes associés.....	104
Figure 18 : Représentation des principales aires écologiques d'Amérique du Sud.	111

Figure 19 : Localisation des populations de comparaison utilisées dans ce travail : (A) populations dont les données ADNmt ont été utilisées, (B) populations dont les données Y ont été utilisées	114
Figure 20 : Echantillons pour lesquels les prélèvements réalisés ont révélé un profil génétique identique.....	124
Figure 21 : Analyse phylogénétique des 64 séquences HVR-1 obtenues à partir des échantillons anciens analysés.....	129
Figure 22 : Représentation schématique des liens de parenté supposés au sein de l'unité domestique 320 du site archéologique de Los Amarillos.....	142
Figure 23 : Arbre phylogénétique (NJ) généré à partir d'une matrice de distances F_{ST} obtenues en comparant les fréquences de lignées maternelles de 15 populations amérindiennes anciennes.....	151
Figure 24 : Analyse en composante principales (ACP) générée à partir des fréquences de lignées maternelles de 69 populations anciennes et contemporaines amérindiennes. ...	152
Figure 25 : Réseaux phylogénétiques illustrant la distribution des haplotypes au sein des quatre lignées maternelles amérindiennes fondatrices.....	159
Figure 26 : Réseau phylogénétique présentant les haplotypes Y (haplotype minimal moins DYS385) amérindiens associés à la lignée paternelle fondatrice Q.....	160
Figure 27 : MDS généré à partir des paires de distances F_{ST} de 44 populations andines contemporaines et anciennes, à partir des variations de séquences HVR-1 (16024-16380).....	162
Figure 28 : Arbre phylogénétique (NJ) présentant les distances génétiques entre populations, calculées à partir des haplotypes Y.....	163

Liste des annexes

- Annexe 1 :** Tableau récapitulatif présentant la séquence des amorces PCR et SBE utilisées pour le génotypage par spectrométrie de masse MALDI-TOF, et la taille des fragments amplifiés. 232
- Annexe 2 :** Résultats de la quantification obtenue pour au moins un extrait de chaque échantillon ancien (la valeur la plus haute est présentée). 233
- Annexe 3 :** (A.) Profils génétiques des fouilleurs (F) et personnels ayant eu accès au laboratoire d'ADNa (M), obtenus avec le kit Identifiler + ; (B.) Séquences HVR-1 obtenues pour le personnel ayant eu accès au laboratoire d'ADNa (M) et les fouilleurs ayant eu accès au matériel (F). 234
- Annexe 4 :** Ensemble des génotypes consensus obtenus par spectrométrie de masse MALDI-TOF pour les 70 échantillons anciens amérindiens analysés. Seuls les génotypes des 15 SNP conservés dans la dernière réaction sont présentés. 235
- Annexe 5 :** Présentation des profils génétiques consensus obtenus à partir de l'utilisation du kit Mentype DIPplex (Qiagen), pour 14 échantillons anciens pour lesquels des liens de parenté étaient supposés. 237
- Annexe 6 :** Fréquences des haplogroupes mitochondriaux de 69 populations d'Amérique du Sud compilées à partir des données disponibles dans la littérature. 238
- Annexe 7 :** Matrices de distances F_{ST} calculés à partir des fréquences d'haplogroupes mitochondriaux de 15 populations anciennes amérindiennes. 239
- Annexe 8 :** Matrice de distances F_{ST} calculés à partir des haplotypes mitochondriaux de 44 populations sud-américaines, avec la méthode Kimura 2P, corrigé avec Slatkin. 240

Liste des abréviations

ADN : acide désoxyribonucléique

A : adénine

ADNa : ADN ancien

ADNmt : ADN mitochondrial

ADNnu : ADN nucléaire

BET : bromure d'éthidium

BSA : albumine sérique bovine (Bovine Serum Albumin)

C : cytosine

ddNTP : didésoxyNucléotide TriPhosphate

DMG : Dernier Maximum Glaciaire

dNTP : désoxyNucléotide TriPhosphate

DTT : Dithiothréitol

EDTA : acide éthylène diamine tétra acétique

G : guanine

J.-C. : Jésus Christ

Kb : kilo base

KCL : chlorure de potassium

HV-1 / HVR-1 : première région hyper variable de la région de contrôle de l'ADN mitochondrial

LR : « Likelihood Ratio » (rapport de vraisemblance)

M : molaire (mole/litre)

MALDI-TOF : Matrix Assisted Laser Desorption/Ionization-Time Of Flight

MDS : « MultiDimensional Scaling »

MgCL² : chlorure de magnésium

min : minute

ml : millilitre

mM : millimolaire

n.d. : non déterminé

ng : nanogramme

NJ : « Neighbor-Joining »

NRY : partie non recombinante du chromosome Y

nt : nucléotide

pb : paire de base

pg : picogramme

PCR : « Polymerase Chain Reaction »

pH : potentiel hydrogène

RFLP : « Restriction Fragment Length Polymorphism » (polymorphisme de longueur des fragments de restriction)

rpm : rotation par minute

SBE : « Single Base Extension »

SNP : « Single Nucleotide Polymorphism » (polymorphisme ponctuel de séquence)

sec : seconde

STR : « Short Tandem Repeat » (microsatellite)

STRa : microsatellite présent sur les autosomes

STR-Y : microsatellite présent sur le chromosome Y

T : thymine

TBE : tris-borate-EDTA

UV : Ultra Violets

μL : microlitre

°C : degré Celsius

Introduction générale

Depuis plusieurs siècles, l'étude de la lignée humaine passionne. En particulier, l'origine de l'**Homme moderne**, mais aussi sa **dispersion** et son **adaptation** à des environnements extrêmement variés dans le monde entier intriguent et font l'objet de vifs débats. Si de nombreuses théories ont dans un premier temps été émises sur le fondement de données paléontologiques et archéologiques, à partir du milieu des années 1980, les données génétiques ont été intégrées dans le débat. Depuis cette époque, les techniques ont été considérablement améliorées et l'ADN est devenu un outil essentiel pour la compréhension de l'**histoire récente de la lignée humaine**.

Grâce aux données génétiques de populations contemporaines des avancées remarquables ont été faites dans la compréhension du peuplement des différentes grandes régions du monde (Goebel *et al.*, 2008; Li and Durbin, 2011; Macaulay *et al.*, 2005; Mellars, 2006; Torroni *et al.*, 2006). Néanmoins, si des schémas généraux de peuplement peuvent être proposés, la dynamique précise de ces processus est difficilement abordable à partir des données contemporaines seules, étant donné que des événements démographiques récents (mouvements de populations, métissages ...) ont pu modifier les pools géniques ancestraux. Etudier les populations anciennes directement apparaît donc comme indispensable pour retracer de manière fiable la dispersion de l'Homme moderne et éviter les biais qui peuvent être induits par l'analyse de populations contemporaines.

Dans ce cadre, la **paléogénétique** se présente comme un outil de choix permettant de s'intéresser directement aux génomes anciens, entre autre d'individus humains anciens. Dès l'émergence de cette discipline vers la fin des années 80, les études d'ADN ancien ont été utilisées pour essayer de répondre aux questions concernant la dispersion de l'Homme moderne dans le monde, et plus particulièrement en **Amérique**, qui constitue une des dernières grandes régions qu'il ait colonisé (Gibbons, 1996; Horai *et al.*, 1993). Aussi bien l'**origine** des premiers Hommes en Amérique que le moment de cette **colonisation** ou les modalités de leur diffusion à tout le continent restent sujets à controverses. La colonisation de l'Amérique du Sud est particulièrement mal comprise actuellement, notamment à cause d'un schéma de diversité génétique des populations actuelles extrêmement complexe. La paléogénétique semble donc un outil indispensable pour cette problématique.

Les particularités de l'ADN ancien induisent de nombreuses difficultés qui rendent les problématiques aussi larges que le peuplement d'un continent difficiles à aborder directement.

Ainsi, les études paléogénétiques se concentrent le plus souvent sur des populations ou des régions restreintes, permettant d'aborder des problématiques plus précises, et de mettre en évidence des **processus micro-évolutifs** sur de plus petites régions. Ces études permettent de mieux caractériser la variabilité génétique des populations du passé, et associées les unes aux autres permettent de tester des hypothèses plus globales.

L'objectif très général de ce travail de thèse a donc été d'essayer de contribuer au débat concernant le peuplement de l'Amérique du Sud, au travers de l'étude de la **diversité génétique de populations anciennes des Andes méridionales**. Cette région, comprenant l'extrême sud de la Bolivie, le nord du Chili et le nord-ouest de l'Argentine, a été occupée depuis au moins 11 000 ans par des groupes de chasseurs-cueilleurs, puis par des sociétés sédentaires complexes. La diversité biologique de ces groupes anciens est quasiment inconnue, ainsi, leur origine, les relations qu'elles ont pu entretenir entre elles ou avec d'autres régions, leur évolution jusqu'à aujourd'hui sont très peu comprises et discutées. Cette étude portant sur des populations de l'ère céramique (entre le V^{ème} et le XV^{ème} siècle de notre ère) est l'occasion d'appréhender ces diverses questions et ainsi d'essayer de révéler la dynamique de peuplement de cette région. Une meilleure connaissance de l'histoire évolutive de ces groupes anciens permettra potentiellement de discuter des hypothèses actuellement les plus plausibles concernant le peuplement de l'Amérique du Sud.

La première partie de ce mémoire sera consacrée à la mise en place du contexte général de l'étude réalisée. Pour cela, les contextes géographique et historique de la région des Andes méridionales seront présentés afin de connaître les hypothèses principales concernant le peuplement de la région. Avant cela, la connaissance actuelle concernant le peuplement de l'Amérique et plus particulièrement de l'Amérique du Sud sera résumée, ainsi que les caractéristiques principales de l'ADN ancien, essentielles pour comprendre la nature des études réalisées. Dans la deuxième partie, le matériel analysé, ainsi que les contextes archéologiques dont ils sont issus seront décrits, tout comme l'ensemble des protocoles d'analyse mis en place pour ce travail de thèse. Enfin, après la présentation des principaux résultats obtenus dans la troisième partie, ces derniers seront remis dans leur contexte et discutés dans la quatrième partie de ce travail afin de mettre en relief les éléments essentiels concernant les populations anciennes analysées et de dégager des perspectives à ce travail.

Première partie.

Contexte général de l'étude.

I. Peuplement de l'Amérique du Sud.

Depuis la découverte de l'Amérique par Christophe Colomb en 1492 et la conquête de l'ensemble du continent pendant le XVI^{ème} siècle, l'intérêt pour le « nouveau monde » n'a jamais faiblit. Cet engouement est d'autant plus important pour l'Amérique du Sud qui a été tardivement colonisée par l'homme moderne mais qui a fait partie des premières régions du monde à avoir vu émerger de grandes civilisations stables et complexes. Comprendre comment et quand l'homme moderne est arrivé dans cette partie du monde, comment il a réussi à s'adapter à des environnements extrêmement variés et parfois hostiles, et comment les civilisations précolombiennes ont émergé et conquis une grande partie de ce nouveau territoire, font partie des questions récurrentes débattues par les historiens, les archéologues, les linguistes ou encore les anthropologues.

Il n'est pas possible de s'intéresser à la dynamique de peuplement de l'Amérique du Sud sans évoquer auparavant le peuplement de l'Amérique dans sa globalité, les deux étant étroitement liés.

1. Etat de la connaissance actuelle.

Comme pour l'ensemble des problématiques de peuplement, les questions principales concernent l'origine des individus, le moment et les modalités (nombre de vagues de migration et voies empruntées) de la colonisation. Durant ces deux dernières décennies, l'apport des données génétiques a permis des avancées importantes concernant ces questions.

Au cours du XX^{ème} siècle de nombreuses populations natives américaines, depuis les régions arctiques jusqu'à l'extrême sud de la Terre de Feu, ont été génétiquement caractérisées pour différents marqueurs classiques, comme les groupes sanguins, les protéines sériques, et les polymorphismes enzymatiques, ce qui a permis notamment de **confirmer l'origine Asiatique des populations natives américaines** (Neel *et al.*, 1980; Schanfield *et al.*, 1990; Szathmary, 1984), évoquée dès 1589 par le frère José de Acosta. De plus, ces premières études ont démontré que les fréquences des marqueurs classiques présentaient des distributions

géographiques particulières tout au long de l'Amérique, ce qui a conduit à la proposition de différentes hypothèses concernant le nombre de vagues de migrations, les voies empruntées et l'ancienneté de la première entrée en Amérique. Une théorie en particulier, fondée sur des données de génétique classique, mais aussi sur des données linguistiques et des données de morphologie dentaire, a marqué un tournant dans ce domaine de recherche. En effet, au milieu des années 1980, Greenberg et ses collaborateurs (1986) proposent le premier modèle interdisciplinaire sur le peuplement de l'Amérique (appelé modèle tripartite) auquel toutes les nouvelles données obtenues ultérieurement seront comparées.

1.1. Modèle des trois vagues de migration.

Ce modèle tripartite ou modèle des trois vagues de migration est essentiellement fondé sur la classification de l'ensemble des populations natives américaines en trois groupes linguistiques distincts : les Amérindiens, les Na-Dene et les Esquimaux-Aléoutes (Greenberg *et al.*, 1986). Pour expliquer cette répartition, les auteurs ont proposé trois vagues de migrations indépendantes et successives : (1) la première ayant conduit à la dispersion des groupes de la famille linguistique amérindienne à quasiment toute l'Amérique, entre 10 000 et 8 000 ans avant notre ère ; (2) la seconde plus tardive, entre 7 000 et 5 000 ans avant notre ère, aurait conduit à la diffusion des Na-Dene vers l'Alaska et vers les régions allant de la côte Pacifique Nord aux états du sud des Etats-Unis ; (3) enfin, les Esquimaux-Aléoutes auraient peuplé les territoires arctiques de l'Amérique au cours de la troisième et dernière vague de migration ayant eu lieu il y a environ 4 000 ans.

Outre le fait que ce modèle a été le premier à combiner des données de différentes disciplines, son impact important est également dû à la concordance de cette théorie avec le modèle Clovis qui constituait à cette époque le modèle de peuplement référence pour l'Amérique.

Ce modèle, fondé sur des données strictement archéologiques, proposait que la tradition lithique Clovis, apparemment la plus ancienne et la plus répandue en Amérique du Nord entre 12 000 et 11 000 ans avant aujourd'hui, soit la culture véhiculée par les premiers migrants ayant colonisé l'Amérique (Haynes, 1964; Martin, 1973). Il a donc été proposé, à partir de ces deux modèles, qu'une expansion de chasseurs venus d'Asie et s'étant dispersés à toute

l'Amérique du Nord en passant par la Béringie¹ puis par le Corridor d'Alberta² en diffusant la culture Clovis, serait à l'origine de tous les groupes natifs américains de langue amérindienne. Selon ce modèle, la préhistoire de l'Amérique est donc relativement jeune, puisque la présence humaine sur ce continent ne serait pas antérieure à 12 000 ans.

Même si encore aujourd'hui certains chercheurs restent partisans de ce modèle et d'une préhistoire courte en Amérique, de nombreuses données l'ont rapidement remis en question. Tout d'abord, la classification linguistique réalisée par Greenberg a été critiquée, mais surtout, de nombreux sites archéologiques pré-Clovis (datés de plus de 12 000 ans) ont été découverts dans toutes les régions de l'Amérique (Dalton, 2003; Dillehay, 2003; Dillehay *et al.*, 2008; Dixon, 2001; Neves, 2000; Neves *et al.*, 2005; Waters and Stafford, 2007) (Figure 1 ; p.39), jetant un doute sérieux sur l'ancienneté de la première migration en Amérique. De plus, ces sites (Cactus Hill, Monte Verde, Pedra Furada...) ayant été occupés alors que les glaciers recouvraient une grande partie de l'Amérique du Nord, impliquaient une colonisation par une voie alternative à celle du corridor d'Alberta.

Par la suite, plusieurs modèles alternatifs ont donc été proposés, notamment fondés sur des données de génétique moléculaire, qui était une discipline en plein essor à cette époque.

1.2. Données génétiques.

L'amélioration des techniques de biologie moléculaire a permis dans les années 80 d'analyser les variations de séquences (polymorphismes) de différentes régions du génome humain. En particulier, l'analyse de génomes non-recombinants et à transmission uni-parentale, l'ADN mitochondrial (ADNmt) ainsi que la portion non recombinante du chromosome Y (NRY), a été très souvent utilisée pour étudier la préhistoire de différentes populations du monde. La combinaison de certaines variations de séquences de ces génomes permet de caractériser ou d'identifier des **lignées génétiques, maternelles** dans le cas de l'ADNmt et **paternelles** dans le cas du chromosome Y. L'identification de ces lignées permettra de déterminer la variabilité génétique d'une population, de connaître leur origine bio-géographique, mais aussi de retracer des événements de migration ou d'évolution.

¹ La Béringie comprend les terres situées entre la Sibérie et l'Alaska, qui ont été émergées pendant le dernier maximum glaciaire (entre 25 000 et 12 000 ans environ).

² Le corridor d'Alberta est un corridor libre de glace, situé entre les deux glaciers continentaux qui recouvraient le Nord de l'Amérique pendant le dernier maximum glaciaire (Figure 1 ; p.39).

Ainsi, dès leur apparition, ces outils moléculaires ont été utilisés pour essayer de répondre aux questions très controversées de la dynamique de peuplement de l'Amérique.

1.2.1. *Les premières études : de nombreuses hypothèses mais peu de certitudes.*

Les toutes premières études moléculaires sur des populations natives américaines ont eu pour objectif de caractériser la diversité génétique présente au sein de ces populations. Bien que dans un premier temps il ait semblé exister une diversité génétique différente dans les populations des trois groupes proposés par Greenberg, avec par exemple des groupes Na-Dene et Esquimaux-Aléoutes qui semblaient présenter une diversité beaucoup plus réduite que celle des amérindiens (Merriwether *et al.*, 1995; Wallace, 1995), il a très vite été démontré que les mêmes lignées maternelles et paternelles étaient distribuées dans toute l'Amérique, même dans les populations de la zone arctique (Lorenz and Smith, 1996; Merriwether *et al.*, 1995). Il semblait donc peu probable que des migrations successives et indépendantes depuis l'Asie aient conduit à la diffusion des mêmes lignées à chaque fois (Kolman *et al.*, 1996). Ainsi, les trois groupes à la base du modèle Tripartite dériveraient du même pool génique¹.

Ces premières études ont révélé que la diversité génétique sur le continent américain était très faible, avec seulement **cinq lignées maternelles** majeures et **une lignée paternelle** prépondérante. Les lignées mitochondriales, appelées **A, B, C, D et X** (Torroni *et al.*, 1992), ont été caractérisées dans un premier temps par des polymorphismes de restriction (RFLP) spécifiques à chaque lignée (Schurr *et al.*, 1990; Wallace *et al.*, 1985), puis par des variations de séquences de la région contrôle de l'ADNmt (HVR-1) (Forster *et al.*, 1996; Horai *et al.*, 1993; Torroni *et al.*, 1993a; Torroni *et al.*, 1993b; Ward *et al.*, 1991). Les études pionnières concernant les lignées paternelles natives américaines ont permis de détecter une seule lignée majeure en Amérique (Bianchi *et al.*, 1997; Karafet *et al.*, 1997; Lell *et al.*, 1997; Pena *et al.*, 1995; Santos *et al.*, 1996; Torroni *et al.*, 1994; Underhill *et al.*, 1996), aujourd'hui appelée lignée **Q** et identifiée par le polymorphisme de séquence **M242** (Seielstad *et al.*, 2003). Mis à part la lignée maternelle X, absente de toutes populations sud-américaines (Dornelles *et al.*, 2005), toutes les autres lignées, maternelles et paternelle ont été décrites dans des populations

¹ Dans la suite de ce travail le terme Amérindien sera donc employé pour se référer à l'ensemble des populations natives américaines.

amérindiennes de toutes les régions de l'Amérique mais avec des gradients de fréquence particuliers. A partir de la diversité établie de chaque lignée ainsi que des gradients de distribution observés, de nombreuses hypothèses concernant le peuplement de l'Amérique ont été proposées, souvent très divergentes.

Les premiers modèles formulés à partir de données génétiques, allant de deux vagues de migrations (Karafet *et al.*, 1999; Lell *et al.*, 2002), à quatre évènements migratoires indépendants (Torrioni *et al.*, 1992; Torrioni *et al.*, 1993b) ont assez vite été remplacé par un modèle de peuplement repris par de nombreux auteurs. En effet, des études prenant en compte un plus grand nombre de données ont été réalisées, conduisant toutes à l'hypothèse d'une **origine unique et simultanée** de l'ensemble des natifs américains (Bonatto and Salzano, 1997a; Bonatto and Salzano, 1997b; Forster *et al.*, 1996; Kolman *et al.*, 1996; Lorenz and Smith, 1997; Merriwether *et al.*, 1995; Stone and Stoneking, 1998). L'utilisation des variations de séquences observées sur le génome mitochondrial et/ou sur le chromosome Y comme une horloge moléculaire (Zuckermandl, 1987; Zuckermandl and Pauling, 1965), a permis d'estimer la date de la colonisation initiale (Bonatto and Salzano, 1997a; Schurr and Sherry, 2004; Silva *et al.*, 2002; Torrioni *et al.*, 1994). Cependant étant donné les disparités, tant au niveau des échantillons étudiés, que des méthodes d'analyses utilisées, ces dates varient fortement, entre approximativement 10 000 et 40 000 ans (Tableau 1), et restent peu fiables. Toutefois, des temps de divergence similaires ont été estimés pour l'ensemble des lignées renforçant l'hypothèse d'une colonisation initiale à partir d'une source unique, et ayant donné lieu à la diffusion simultanée de l'ensemble des lignées maternelles et paternelles.

Si avec ce modèle, une origine asiatique unique semblait faire consensus, la ou les voies empruntées, le moment ainsi que l'origine exacte et l'effectif des migrants fondateurs restaient des questions irrésolues.

Avec la découverte de polymorphismes spécifiques des populations amérindiennes¹ (Bonatto and Salzano, 1997b), certains auteurs ont commencé à imaginer que la Béringie ait pu jouer un rôle important dans le processus de colonisation, alors qu'elle n'était vue jusque là que comme un lieu où les populations seraient passées rapidement avant d'entrer en Amérique. Etant absents de toute population asiatique, cela suggère que ces polymorphismes sont

¹ La mutation de la région contrôle de l'ADNmt notée 16111T est caractéristique de la lignée maternelle A2 qui est spécifique des populations natives américaines.

apparus après la divergence entre les ancêtres des natifs américains et les asiatiques, mais avant l'entrée de ces ancêtres en Amérique, probablement donc en Béringie (Bonatto and Salzano, 1997b).

Pour résumer, ces premières études ont permis de s'accorder sur les paramètres généraux de la première colonisation, avec un sous-ensemble des lignées asiatiques, probablement du sud de la Sibérie (Derenko *et al.*, 2001), qui aurait été diffusé en Amérique par le biais d'une population ancestrale, isolée un certain temps en Béringie avant sa première entrée en Amérique pendant le dernier maximum glaciaire (DMG).

Tableau 1 : Dates de divergence (en milliers d'années) des différents haplogroupes et sous-haplogroupes fondateurs amérindiens présentés dans ce travail.

	Torrioni <i>et al.</i> , 1994	Bonatto <i>et al.</i> , 1997	Silva <i>et al.</i> , 2002	Tamm <i>et al.</i> , 2007	Fagundes <i>et al.</i> , 2008	Perego <i>et al.</i> , 2009
A	30,0 ± 4,0	28,1 ± 4,2	20,5 ± 4,1			
A2				13,9 ± 2,0	20,5 ± 5,7	15,8 ± 2,5
B	13,0 ± 2,5	26,7 ± 3,9	18,1 ± 3,6			
B2				16,5 ± 2,7	20,3 ± 5,0	18,7 ± 2,5
C	38,3 ± 5,3	27,9 ± 4,5	21,6 ± 4,3		17,2 ± 5,7	
C1						19,6 ± 2,5
C1b				14,7 ± 4,7		16,5 ± 2,5
C1c				15,8 ± 4,7		17,2 ± 2,5
C1d				9,5 ± 3,4		18,7 ± 2,5
D	21,1 ± 3,0	32,1 ± 5,3	23,8 ± 4,5			
D1				18,0 ± 2,0	21,5 ± 8,0	16,9 ± 2,5
X2a					17,9 ± 11,0	14,2 ± 2,5

1.2.2. Avancées technologiques et nouvelles perspectives : vers une meilleure caractérisation de la diversité génétique.

Si les premières études réalisées ont rendu possible la caractérisation des lignées fondatrices majeures des populations natives américaines, ainsi qu'une meilleure compréhension de la dynamique de peuplement de l'Amérique, le modèle proposé semblait trop simpliste. Dans ce cadre, les avancées technologiques et méthodologiques importantes de ces dernières années, comme par exemple la possibilité d'analyser de manière simultanée un grand nombre de positions de la séquence codante de l'ADNmt ou encore des génomes mitochondriaux complets, ont permis une étude à plus haute résolution des génomes amérindiens. Cette méthodologie a rapidement été utilisée pour une étude plus fine des événements migratoires

humains (Ingman *et al.*, 2000; Macaulay *et al.*, 2005), mais en ce qui concerne les populations natives américaines, les études ont été tardives.

Les génomes mitochondriaux amérindiens complets disponibles (Achilli *et al.*, 2008; Fagundes *et al.*, 2008; Hartmann *et al.*, 2009; Kitchen *et al.*, 2008; Perego *et al.*, 2009; Tamm *et al.*, 2007; van Oven and Kayser, 2009) ont permis de mieux caractériser la diversité génétique des amérindiens. Ainsi, plusieurs lignées maternelles mineures, distribuées de manière inégale sur le continent ont été découvertes : **C1b**, **C1c**, **C1d**, **C4c**, **D4h3**, **X2a** (Malhi *et al.*, 2010; Tamm *et al.*, 2007). De plus, avec l'augmentation du nombre de données concernant le chromosome Y (Bailliet *et al.*, 2009; Bisso-Machado *et al.*, 2011; Bolnick *et al.*, 2006; Bortolini *et al.*, 2003; Karafet *et al.*, 2008; Karafet *et al.*, 2002; Seielstad *et al.*, 2003; Zegura *et al.*, 2004) ainsi que l'harmonisation des nomenclatures (Karafet *et al.*, 2008; YCC, 2002), plusieurs sous-haplogroupes de la lignée paternelle principale Q ont pu être caractérisés. Le paragroupe **Q1a3a*** (M3), apparu semble-t-il en Béringie (Bianchi *et al.*, 1998; Lell *et al.*, 1997; Ruiz-Linares *et al.*, 1999), est ainsi le plus fréquent dans quasiment toutes les populations amérindiennes, suivi par le paragroupe **Q1a3*** (M346) (Bisso-Machado *et al.*, 2011; Blanco-Verea *et al.*, 2010). Certaines lignées mineures sont retrouvées très localement, comme la lignée **Q1a3a1** (M19) dans les populations Ticuna et Wayuu de Colombie (Bortolini *et al.*, 2003) ou la lignée récemment décrite, **Q1a3a4** (SA01), dans les populations andines (Jota *et al.*, 2011). Mis à part cette lignée Q, les populations natives américaines présentent de faibles fréquences de l'haplogroupe **C3b** (P39), détecté seulement en Amérique du Nord (Bisso-Machado *et al.*, 2011; Karafet *et al.*, 2008; Zegura *et al.*, 2004). Le paragroupe plus ancien **C3*** a par ailleurs été retrouvé dans quelques populations natives du nord-ouest de l'Amérique du Sud (Geppert *et al.*, 2011; Zegura *et al.*, 2004).

L'ensemble des lignées fondatrices natives américaines et les polymorphismes qui leurs sont associés sont présentés dans le tableau 2.

Il est important de noter que les phylogénies sont en constante évolution. La lignée Q1a3a4 typique des populations andines a été découverte très récemment (Jota *et al.*, 2011). De plus, une version révisée de la lignée Q a été publiée le 29 juillet 2011 (http://www.isogg.org/tree/ISOGG_HapgrpQ.html), avec de nouvelles lignées ainsi que de nouveaux marqueurs, mais aussi une réattribution des lignées précédentes. Ainsi, le marqueur M3 n'est plus caractéristique de la lignée Q1a3a* mais de la lignée Q1a3a1, le marqueur M19 devient caractéristique de la lignée Q1a3a1a, et ainsi de suite. Cependant, pour une meilleure compréhension, et pour plus de facilité, l'ancienne nomenclature sera conservée dans le texte, l'important étant la fréquence des marqueurs et non pas le nom des lignées.

Tableau 2 : Lignées maternelles (2.A) et paternelles (2.B) fondatrices des populations amérindiennes et polymorphismes associés.

2.A

Haplogroupes mt	SNP HVR-1	SNP HVR-2	SNP de la région codante
A	16290, 16319	235, 663	1736, 4248, 4824, 8794
A2	16111, 16362	146, 153	8027, 12007
A2a	16192		3330
A2b	16265		11365
B	16189		8281-8289 d
B4	16217		
B4b		499, 827	4820, 13590, 15535
B2			3547, 4977, 6473, 9950, 11177
X	16189, 16278	153	6221, 6371, 13966, 14470
X2		195	1719
X2a	16213	200	8913, 12397, 14502
C	16327	249d 3552A, 9545, 11914, 13263, 14318	
C1	16325	290-291 d	
C1b		493	
C1c			1888, 15930
C1d	16051		7697
C4			2232A, 6026, 11969, 15204
C4c	16245		11440, 13368, 14433, 15148
D	16362		4883, 5178A
D4			3010, 8414, 14668
D4h3	16301, 16342	152	3396, 3644, 5048
D1*	16325	2092	
D2*	16129, 16271	3316, 7493, 8703, 9536, 11215	
D2a			11959
D4b1*	16319	8020, 10181, 15440, 15951	

2.B

Haplogroupes Y	SNP
C*	RPS4Y711, M216, P184, P255, P260
C3	M217, PK2, P44
C3b	P39
Q*	M242
Q1	P36.2
Q1a	MEH2
Q1a3	M346
Q1a3a	M3
Q1a3a1	M19
Q1a3a2	M194
Q1a3a3	M199, P106, P292
Q1a3a4	SA01

L'étude de la répartition et de la diversité de toutes ces lignées fondatrices nouvellement décrites rend compte de la complexité de la dynamique de peuplement en Amérique. L'ensemble des travaux réalisés permettent de définir plusieurs étapes clé du processus de colonisation :

1) Divergence des ancêtres des amérindiens du pool génétique asiatique.

Cette divergence a probablement commencé avant le DMG (25 000 - 12 000 ans), autour de 40 000 ans (Kitchen *et al.*, 2008; Mulligan *et al.*, 2008; Perego *et al.*, 2010), avec un mouvement de populations sud-sibériennes vers le nord-est de la Sibérie et une partie de la Béringie. L'ensemble des lignées maternelles majeures (Derenko *et al.*, 2001; Kolman *et al.*, 1996), ainsi que les deux lignées paternelles Q et C (Zegura *et al.*, 2004) ont été retrouvées dans certaines régions de l'Altaï et de l'Amur, faisant de ces populations sud-sibériennes les groupes source les plus probables (Starikovskaya *et al.*, 2005). Les données archéologiques disponibles semblent concorder puisque des traces d'occupation humaine ont été découvertes dans le nord ouest de la Béringie, datant d'au moins 30 000 ans (Goebel, 2007; Pitulko *et al.*, 2004).

2) Isolement et diversification des ancêtres des amérindiens.

La présence de nombreux polymorphismes spécifiques des populations amérindiennes, comme la mutation caractéristique de la lignée A2 (Achilli *et al.*, 2008; Bonatto and Salzano, 1997b), les sous-haplogroupes mitochondriaux C1b, C1c et C1d (Tamm *et al.*, 2007), la lignée paternelle caractérisée par le marqueur M3 (Bortolini *et al.*, 2003; Karafet *et al.*, 2008; Rasmussen *et al.*, 2010) ou encore les 9 répétitions du marqueur autosomal D9S1120 (Phillips *et al.*, 2008; Schroeder *et al.*, 2007), signifie que la population ancestrale qui a colonisé l'Amérique devait déjà avoir acquis un certain degré de différenciation par rapport aux populations asiatiques. Un isolement relativement long (qui aurait duré entre 7 000 et 15 000 ans au minimum) permet d'expliquer l'apparition de ces nombreuses variations spécifiques (Mulligan *et al.*, 2008). Les données paléo-écologiques récoltées ces dernières années confirment que les ressources disponibles en Béringie devaient être suffisantes pour subvenir aux besoins d'une petite population (Zazula *et al.*, 2003). De plus, les traces d'occupation humaine les plus anciennes, dans la région arctique de la Béringie concordent avec cette période et pré-datent les plus anciennes traces d'occupation en Amérique (Hamilton and Buchanan, 2010; Waters and Stafford, 2007).

3) Expansion et colonisation de l'Amérique.

Les estimations des temps de coalescence des différentes lignées, même mineures, varient entre 18 000 et 15 000 ans (Tableau 1 ; p.34) (Fagundes *et al.*, 2008; Perego *et al.*, 2009; Perego *et al.*, 2010; Tamm *et al.*, 2007). De plus, les signes d'une expansion de population, importante et rapide, ont également été détectés entre 16 000 et 12 000 ans environ (Fagundes *et al.*, 2008; Kitchen *et al.*, 2008; Mulligan *et al.*, 2008), ce qui concorde avec une période de réchauffement climatique (entre 18 000 et 12 000 ans) qui a conduit à l'élévation du niveau de la mer mais aussi à la diminution des terres émergées de Béringie. Même si les glaciers étaient encore largement présents, des routes, notamment côtières et entre les glaciers, ont pu devenir praticables (Dixon, 2001). Un évènement migratoire majeur par la voie littorale Pacifique a été confirmé par la distribution géographique de certaines lignées mineures, notamment la lignée maternelle D4h3 (Perego *et al.*, 2009), mais aussi par la réalisation de simulations (Fix, 2005; Surovell, 2000). Si cette voie littorale semble avoir été la voie principalement empruntée (Dixon, 2001; Goebel *et al.*, 2008), un passage par le corridor d'Alberta ne peut être exclu, notamment sur la base de la distribution de certaines lignées maternelles (X2a, D2, D1) et paternelles (C3), retrouvées seulement dans les populations du nord de l'Amérique du Nord (Perego *et al.*, 2009; Tamm *et al.*, 2007). La diffusion de ces lignées s'est faite soit simultanément à la colonisation majeure par le littoral Pacifique (Perego *et al.*, 2009), soit ultérieurement, ce qui conforterait les toutes dernières études démontrant que suite au premier évènement majeur qui a conduit à la diffusion des lignées fondatrices à toute l'Amérique, des flux de populations et de gènes ont pu avoir lieu régulièrement entre la Béringie et le nord de l'Amérique pendant tout le processus de colonisation (de Azevedo *et al.*, 2011; Ray *et al.*, 2010).

Ces différentes étapes semblent aujourd'hui faire consensus et permettent d'expliquer la dynamique globale du peuplement de l'Amérique (Figure 1), même s'il reste encore plusieurs points à éclaircir, dont la diffusion des populations amérindiennes en Amérique du Sud, après le passage de l'isthme du Panama.

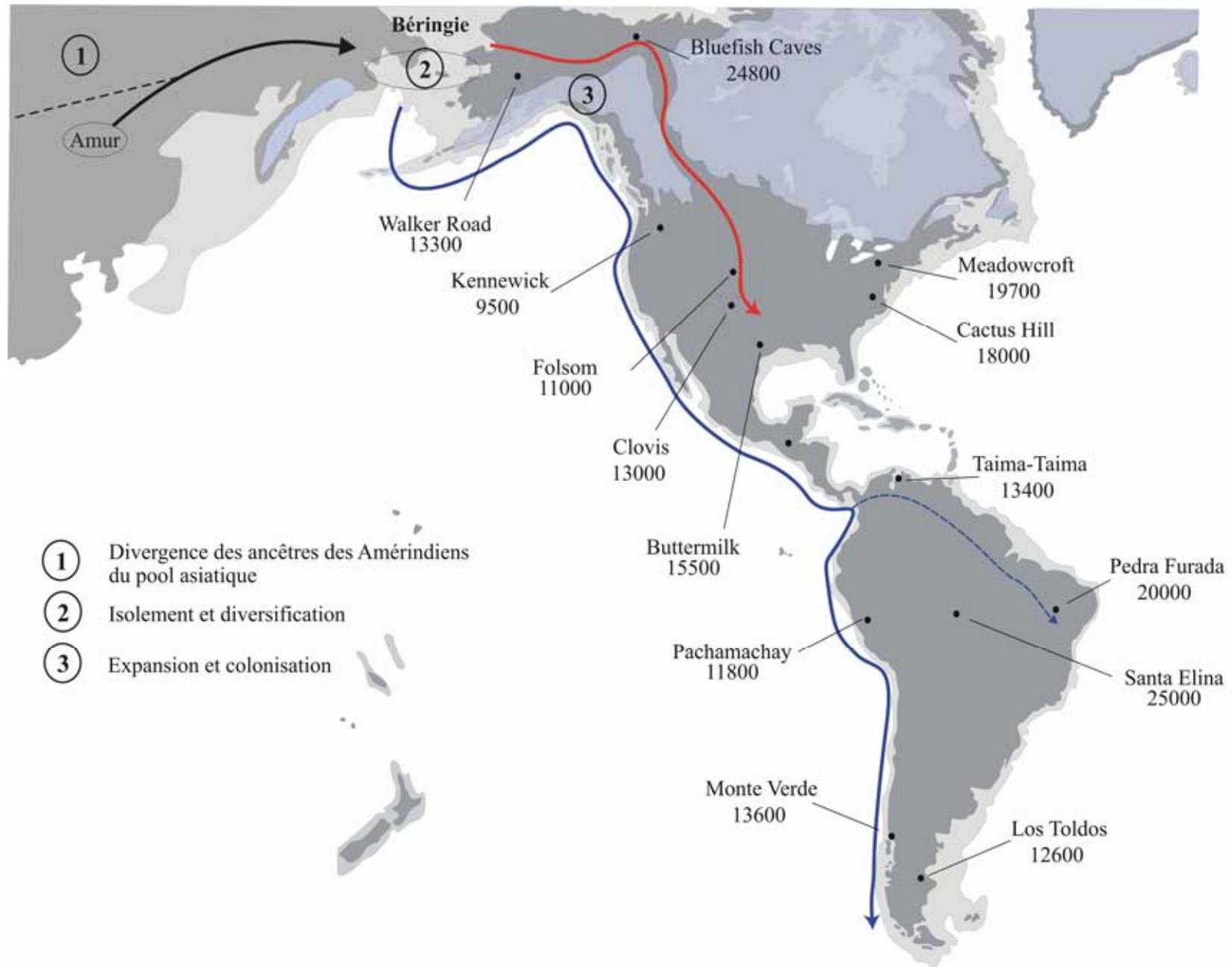


Figure 1 : Schéma global de la dynamique de peuplement de l'Amérique. *Adapté de(Nemecek, 2011).*

1.3. Peuplement de l'Amérique du Sud : un gradient est/ouest.

Après le passage par la Mésoamérique, la diffusion des amérindiens dans le sous-continent sud reste très énigmatique. Deux éléments principaux induisent des difficultés dans la compréhension du peuplement initial. Premièrement, contrairement à ce qui est observé en Amérique du Nord, aucune domination d'une culture particulière n'a été remarquée en Amérique du Sud. Cette diversité culturelle précoce peut être expliquée en partie par la présence sur le territoire de nombreuses **barrières géographiques** (Andes, fleuve Amazone) qui ont pu freiner les mouvements de populations, mais aussi par des changements climatiques importants (Rothhammer and Dillehay, 2009). Deuxièmement, l'Amérique du Sud est caractérisée par des **différences marquée entre est et ouest**, avec entre autre des données archéologiques, des chronologies ainsi que des histoires évolutives différentes. En effet, plusieurs études, fondées sur différents systèmes génétiques tels que les autosomes (Luiselli *et al.*, 2000; Wang *et al.*, 2007), le chromosome Y (Tarazona-Santos *et al.*, 2001) ou l'ADNmt (Fuselli *et al.*, 2003; Lewis *et al.*, 2007b) ont fait état d'une différence significative entre les populations homogènes de l'ouest et les populations hétérogènes de l'est. Les populations andines présentent une diversité génétique (intra-populationnelle) plus élevée et une divergence (inter-populationnelle) plus faible que les populations amazoniennes. Ces observations peuvent être la conséquence : (1) soit de facteurs historiques et/ou écologiques, ayant induit des intensités de dérive génétique et de flux génétiques différents dans les deux zone ; (2) soit d'une division du pool génétique ancestral ayant conduit à deux migrations indépendantes en Amérique du Sud. Encore aujourd'hui, aucune des deux hypothèses n'est privilégiée, les études menées favorisant alternativement l'une ou l'autre.

La distribution des lignées maternelles amérindiennes, avec un gradient nord sud décroissant pour la lignée A2 et inversement croissant pour les lignées C1 et D1, et une très forte fréquence de lignées B2 dans les populations andines, a été expliqué, par de nombreux auteurs, par une diffusion de ces lignées par plusieurs voies migratoires différentes : une vague vers l'est, depuis le nord de l'amazone jusque vers l'extrême sud du continent, et une deuxième vague le long de la côte Pacifique ayant conduit à la colonisation de la région andine (Keyeux *et al.*, 2002; Lalueza *et al.*, 1997; Rodriguez-Delfin *et al.*, 2001; Rothhammer and Silva, 1989; Tarazona-Santos *et al.*, 2001). D'autres auteurs proposent une migration initiale le long des hauts-plateaux andins, puis une migration consécutive de la partie

amazonienne à partir du pool génétique andin (Rothhammer and Dillehay, 2009). Il existe pratiquement autant d'hypothèses que d'auteurs.

La région andine est une des deux régions de l'Amérique, avec la Mésoamérique, qui a vu le développement de grandes civilisations socio-économiquement très stables, avec des effectifs de populations relativement élevés. La dérive génétique a donc pu avoir beaucoup moins d'impact sur ces populations que sur les petits groupes relativement isolés d'Amazonie. Ainsi, le patron de diversité observé a probablement au moins en partie été influencé par l'histoire démographique différente des populations andines et amazoniennes (Fuselli *et al.*, 2003; Gaya-Vidal *et al.*, 2011). De plus, les Andes ainsi que le fleuve Amazone ont pu constituer des barrières, conduisant à l'isolement géographique et génétique de certaines populations, en interrompant régulièrement les flux de gènes entre certains groupes pourtant proches géographiquement (Rothhammer and Dillehay, 2009).

Selon certains auteurs, les différences entre est et ouest ne sont pas si évidentes et au contraire l'homogénéité des populations amérindiennes serait un argument en faveur d'un évènement migratoire unique en Amérique du Sud (Barbieri *et al.*, 2011; Lewis *et al.*, 2007b; Lewis *et al.*, 2005; Moraga *et al.*, 2000). Selon ces auteurs, la disparité des données disponibles, avec un manque important de données des populations amérindiennes amazoniennes peut induire un biais dans les études, conduisant à de mauvaises interprétations de la diversité génétique observée.

Enfin, une question importante est de savoir si l'organisation actuelle des populations amérindienne reflète réellement l'organisation qui existait à leur toute première arrivée dans la région ? Il est possible, par exemple, que l'homogénéité andine et la fréquence élevée de l'haplogroupe B caractéristique de ces populations ne soit que le résultat d'évènements récents (Barbieri *et al.*, 2011; Rodriguez-Delfin *et al.*, 2001). Le développement de sociétés complexes et leur expansion (par exemple l'expansion Inca), a pu rendre certaines lignées prédominante par des effets de dérive génétique, puisque de nombreuses petite populations ont été soumises et/ou ont disparues à ces périodes (Rodriguez-Delfin *et al.*, 2001).

A la vue des données disponibles à l'heure actuelle nous pouvons remarquer qu'il existe de très nombreuses interrogations, et qu'il manque des études interdisciplinaires permettant de révéler de manière précise la dynamique de peuplement de l'Amérique du Sud.

Les populations contemporaines permettent-elles vraiment de comprendre la préhistoire, et les évènements complexes qui ont conduit à la diversité génétique actuellement observée ? Il semble que tous les aspects d'une telle dynamique ne soit pas compréhensible simplement en analysant la diversité génétique contemporaine, même avec les technologies les plus avancées. Il est donc indispensable d'étudier la diversité génétique des populations du passé directement, ce qui est possible aujourd'hui grâce à l'**ADN ancien** (ADNa).

2. L'ADN ancien un outil essentiel pour la compréhension des populations du passé.

Il apparaît à la vue des nombreuses études réalisées à partir de populations contemporaines que les évènements migratoires anciens peuvent être masqués par des évènements démographiques plus récents, qui ont pu modifier considérablement le pool génique des populations anciennes. Etudier directement les populations du passé semble donc le seul moyen pour éviter les biais induits par l'analyse de populations contemporaines. Les disciplines telles que l'archéologie, l'anthropologie, ou encore la paléo-écologie vont permettre de s'intéresser à de nombreux aspects des populations anciennes, sans toutefois pouvoir approcher vraiment la diversité biologique des individus anciens. De nouvelles perspectives sont apparues au milieu des années 1980, lorsque deux études indépendantes ont démontré que des molécules d'ADN ancien pouvaient « survivre » pendant une longue période de temps (l'étude de Paabo portait sur une momie de 2430 ans) et qu'en plus il était possible de séquencer ces molécules (Higuchi *et al.*, 1984; Paabo, 1985). Par la suite, des progrès importants ont été faits avec l'apparition de la PCR, qui permet une amplification exponentielle d'un fragment d'ADN cible présent en très faible quantité de départ, voire en une seule copie (Mullis and Faloona, 1987), conduisant à la multiplication des études à partir de nombreux substrats différents (Cano *et al.*, 1993; Golenberg *et al.*, 1990; Lindahl, 1993; Soltis *et al.*, 1992; Woodward *et al.*, 1994). Ces premières études ont très vite révélé l'omniprésence des dommages et des contaminations présents sur les séquences d'ADNa (Paabo, 1989) et donc le besoin de méthodologies analytiques adaptées à ce matériel biologique très particulier (Cooper and Poinar, 2000; Gilbert *et al.*, 2005; Hofreiter *et al.*, 2001; Paabo *et al.*, 2004), menant à l'apparition d'une discipline à part entière : la **paléogénétique**.

2.1. Caractéristiques de l'ADN ancien.

L'ADN ancien peut être défini comme tout ADN, extrait d'organismes anciens, qui n'a pas été conservé délibérément pour des analyses génétiques. Cet ADN présente des caractéristiques spécifiques, très différentes d'un ADN contemporain, étant donné que dès la mort d'un organisme, les mécanismes cellulaires de réparation de l'ADN cessent immédiatement, et les cellules subissent rapidement une autolyse, par l'action des endo- et exo-nucléases. En plus de ces dégradations enzymatiques rapides, des processus chimiques, plus lents, vont également contribuer à la dégradation des molécules anciennes. Ces phénomènes chimiques sont surtout des hydrolyses et/ou des oxydations qui pourront provoquer des coupures simple ou double brin d'ADN mais aussi la formation de bases modifiées ou de sites abasiques (Hofreiter *et al.*, 2001; Hoss *et al.*, 1996; Lindahl, 1993), comme illustré sur la figure 2. D'autres phénomènes chimiques conduiront plutôt à la formation de liaisons intra- et/ou inter-moléculaires, produisant par exemple des complexes ADN-protéine (Willerslev and Cooper, 2005).

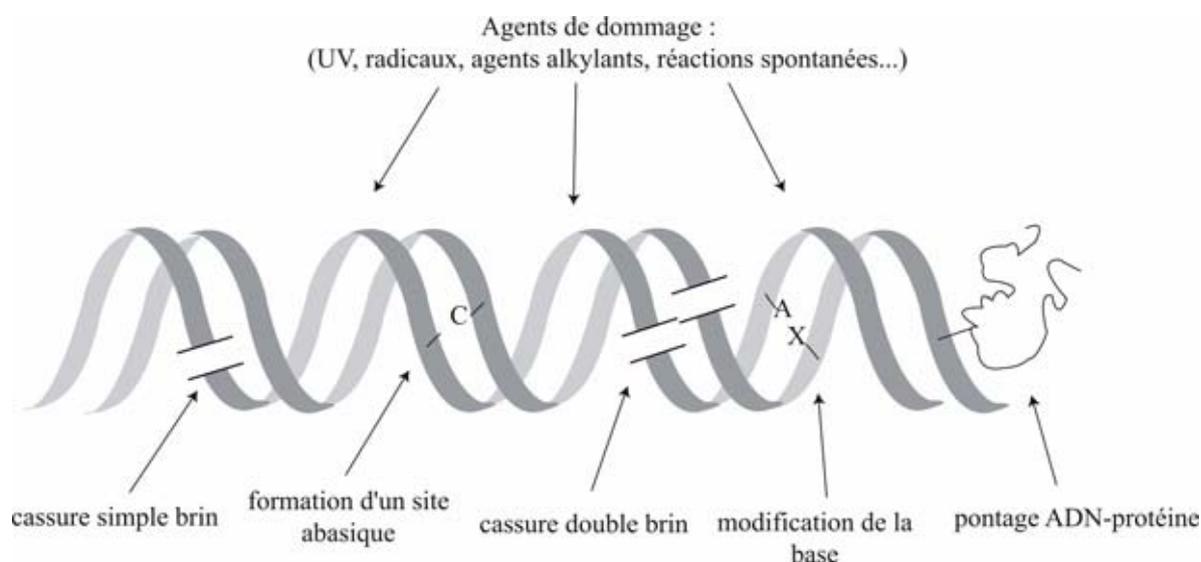


Figure 2 : Représentation des dégradations post-mortem subies par la molécule d'ADN.

D'après (Jobling et al., 2004).

Les mécanismes de dégradation des molécules sont identiques pour tous les individus mais leur taux peut varier fortement d'un individu à l'autre, en fonction de nombreux facteurs

physico-chimiques. Il apparaît donc que la préservation de l'ADN au sein d'organismes anciens est dépendante des conditions dans lesquelles ces organismes ont été maintenus depuis leur mort : le pH, la température, la pression, le taux d'oxygène, le taux d'humidité, et la concentration en sels pouvant affecter le nombre de modifications *post-mortem* que l'ADN subi (Lindahl, 1993). Certains environnements, comme les milieux froids et/ou secs, semblent favoriser la préservation des molécules anciennes, alors qu'au contraire des milieux humides et des températures élevées apparaissent comme particulièrement défavorables (Hofreiter *et al.*, 2001). Le type d'échantillon analysé peut aussi avoir un impact sur la réussite de l'amplification de l'ADNa. Si beaucoup de tissus de types différents peuvent être analysés (tissus mous, ongles, cheveux, os et dents), les dents et les os longs sont parmi ceux où la préservation de l'ADN semble la meilleure, et peuvent être facilement soumis à des décontaminations de surface. Ils sont donc les tissus privilégiés pour les études paléogénétiques. Cependant, tous les mécanismes de dégradation de l'ADN n'étant pas connus, il n'est pas possible aujourd'hui de prédire l'état de préservation de l'ADN au sein d'un organisme ou d'un échantillon ancien. L'ancienneté des échantillons anciens analysables est donc très variable, mais ne semble pas pouvoir dépasser les 100 000 ans à l'heure actuelle (pour l'ADN humain) (Lindahl, 1993; Lindahl, 1997; Willerslev *et al.*, 2003).

Quoi qu'il en soit, les dégradations *post-mortem* ont des implications très importantes pour les études d'ADNa : la quantité ainsi que la qualité de l'ADN disponible seront moindres. En effet, les clivages enzymatiques et chimiques fragmentent l'ADN en de très courtes régions isolées qui souvent ne sont pas plus grandes que 200 à 300 paires de bases (Briggs *et al.*, 2009). De plus, les modifications ou pertes de bases vont augmenter le risque d'erreurs dans les séquences obtenues (Briggs *et al.*, 2007; Brotherton *et al.*, 2007).

L'état **dégradé** et **fragmenté** de l'ADNa va induire trois difficultés majeures pour les analyses d'échantillons anciens et d'autant plus pour les échantillons humains anciens.

La première difficulté est essentiellement **technique**. En effet, certaines lésions de l'ADNa pourront conduire à l'inhibition complète de la PCR, ou à l'incorporation dans les séquences de bases erronées (Gilbert *et al.*, 2003; Paabo, 1989). Certains inhibiteurs également pourront empêcher une amplification efficace. Les protocoles d'analyse devront donc être adaptés aussi bien au niveau de la technique (protocole d'extraction particulier, paramètres de PCR spécifiques, dilutions ...), que des marqueurs analysés (courts amplicons), tout en conservant un pouvoir informatif suffisant.

La deuxième difficulté vient des risques très élevés de **contamination** par des molécules

contemporaines exogènes, qui n'étant ni fragmentées, ni dégradées, seront préférentiellement amplifiées. Ce risque de contamination est présent tout au long du processus d'analyse (par les personnes ayant accès aux échantillons anciens, le matériel utilisé ou encore les réactifs employés), ce qui implique des précautions rigoureuses à prendre tout au long du traitement des échantillons anciens. Il existe plusieurs stratégies permettant de minimiser le problème des contaminations, comme le choix des tissus étudiés (Gilbert *et al.*, 2007; Willerslev *et al.*, 2009), les méthodes de décontamination employées, ou encore le matériel utilisé.

Enfin, la troisième difficulté, qui découle des deux précédentes, concerne **l'authentification** des résultats obtenus. Parmi les premières études paléogénétiques publiées, plusieurs résultats se sont avérés être des contaminations, jetant le doute sur toute la discipline. Ainsi, plusieurs guides rigoureux et détaillés ont depuis été publiés, permettant l'évaluation de l'authenticité des séquences d'ADNa (Cooper and Poinar, 2000; Gilbert *et al.*, 2005; Hofreiter *et al.*, 2001; Paabo *et al.*, 2004). Cependant, tous les critères décrits dans ces guides ne sont pas toujours applicables à toutes les études d'ADNa. Il semble donc préférable d'adapter une stratégie d'authentification à chaque échantillon ancien ou à chaque étude menée.

Les critères employés au cours de ce travail, aussi bien pour éviter les contaminations que pour authentifier les résultats, seront détaillés dans le chapitre 2.V. (partie 2).

Malgré toutes ces difficultés, l'étude directe de génomes d'individus anciens donne accès à des informations inestimables, impossible à obtenir par d'autres méthodes. La paléogénétique a donc assez rapidement été déclinée pour des applications variées telles l'étude d'espèces éteintes, la reconstruction de phylogénies, la paléo-pathologie, mais aussi l'étude de l'histoire de l'Homme moderne. En effet, les données paléogénétiques ont été utilisées pour tester les nombreuses hypothèses émises sur la base de données archéologiques, linguistiques ou génétiques, concernant la diffusion de l'homme moderne dans différentes régions du monde, comme en Europe (Francalacci *et al.*, 1996; Lacan *et al.*, 2011), en Océanie (Adcock *et al.*, 2001; Hagelberg, 1997; Hagelberg and Clegg, 1993), ou en Amérique.

2.2. ADN ancien et peuplement de l'Amérique du Sud.

Il existe à ce jour assez peu de données d'ADNa sur les populations amérindiennes sud-américaines, et elles se révèlent très inégales quant à la qualité et aux régions étudiées. Les caractéristiques générales des études réalisées ainsi que la localisation des populations anciennes analysées sont présentées dans le tableau 3 et la figure 3 suivants.

De manière intéressante, toutes ces études sont le reflet de l'évolution de la paléogénétique, aussi bien du point de vue technique, que du point de vue de l'approche adoptée. Dans un premier temps (jusqu'au début des années 2000), les chercheurs ont utilisé l'ADNa essentiellement dans le but de connaître la diversité génétique des populations anciennes mais aussi, avec l'objectif ambitieux de retracer la dynamique de peuplement de l'Amérique du Sud.

Ainsi, des scénarios de peuplement à plusieurs vagues de migrations ont été proposés à partir des premières études menées (Lalueza *et al.*, 1997; Monsalve *et al.*, 1994). En effet, étant donnée l'absence apparente des lignées maternelles A et B dans les populations de la Terre de Feu, Lalueza a proposé que la première vague de migration entrée en Amérique du Sud ait permis la diffusion des lignées C et D et que donc les deux autres lignées fondatrices auraient été introduites ultérieurement. (Demarchi *et al.*, 2001b; Lalueza *et al.*, 1997). Ces premières études ont également montré qu'un certain nombre d'échantillons anciens (39 % pour les individus amazoniens) n'appartenaient à aucune lignée maternelle amérindienne. Une hypothèse de baisse de la variabilité génétique ancienne causée par l'arrivée des colons européens a donc été proposée sur la base de ces résultats (Monsalve, 1997; Monsalve *et al.*, 1996; Monsalve *et al.*, 1994; Ribeiro-dos-Santos *et al.*, 1996). Cependant, quelques doutes subsistent quant à l'authenticité d'une partie de ces premiers résultats. Il apparaît très probable, à la lecture des protocoles, que ces résultats soient la conséquence de contaminations. Pour exemple, Ribeiro-dos-Santos et son équipe ont amplifié des fragments de 424 et 598 paires de bases, ce qui paraît étonnant pour de l'ADN ancien, et obtenu un certain nombre de séquences HVR-1 présentant le motif 16223T – 16278T qui semble être caractéristique de la lignée africaine L2.

Ces premières études, même si elles ont permis de montrer que les populations amérindiennes anciennes présentaient l'ensemble des lignées maternelles détectées dans les populations contemporaines, soulignent les difficultés inhérentes à l'analyse d'échantillons anciens (contaminations, effectif réduit).

Si en théorie la paléogénétique permet de retracer des événements anciens tels que des migrations ou des effets fondateurs, il apparaît en réalité que de telles études nécessitent un effectif important d'échantillons, les plus anciens possibles et non apparentés. Toutes ces conditions sont difficiles à réunir, ainsi, à partir des années 2000, les chercheurs se sont intéressés à des phénomènes moins généraux, et ont étudié des problématiques locales plus ciblées, notamment concernant certaines périodes ou régions particulières des Andes

centrales, où les conditions de conservation de l'ADN sont favorables.

Ces études, en plus de s'intéresser à des problématiques particulières telles que l'évolution des populations dans la vallée d'Azapa au Chili (Moraga *et al.*, 2005), l'origine des populations de la période Inca au Machu Picchu (Shinoda *et al.*, 2006), les pratiques funéraires de la culture Sican au Pérou (Shimada *et al.*, 2004), la continuité génétique après le déclin de l'empire Wari au Pérou (Kemp *et al.*, 2009) ou encore la dynamique de peuplement de la région de Palpa sur la côte péruvienne (Fehren-Schmitz *et al.*, 2010a), ont permis de démontrer une continuité entre la diversité génétique des populations amérindiennes anciennes et contemporaines. En effet, même si certains gradients de fréquences, comme pour les populations actuelles, peuvent être perçus, notamment avec l'absence des lignées A et B en Terre de Feu (Garcia-Bour *et al.*, 2004) et une fréquence plus élevée des lignées B dans les populations andines (Fehren-Schmitz *et al.*, 2010b; Lewis Jr *et al.*, 2007; Rothhammer *et al.*, 2003), aucune lignée supplémentaire n'a été détectée. A la vue de ces résultats, l'hypothèse d'une arrivée séparée des différentes lignées apparaît donc peu probable.

Les études d'ADNa sont encore trop peu nombreuses pour pouvoir confronter ces données anciennes aux hypothèses de peuplement émises sur la base des données contemporaines. Notamment, comme cela est illustré sur la figure 3, on remarque qu'il existe encore de nombreuses aires géographiques vierges de toute étude en Amérique du Sud, et en particulier, l'ensemble de l'est du continent, qui est pourtant une zone clé pour comprendre la diffusion des premiers amérindiens. Le climat de cette région amazonienne n'apparaît pas comme étant le plus favorable à la bonne conservation des restes humains anciens, et d'autant plus pour l'ADN, ce qui peut expliquer en partie l'absence de données provenant de ces populations.

Un corpus plus important de donnée sera nécessaire, aussi bien au niveau des régions étudiées que des périodes, pour appréhender correctement la diffusion de l'homme moderne en Amérique du Sud.

Nous avons choisi pour ce travail de thèse de nous intéresser à une région encore vierge en ce qui concerne les études paléogénétiques : les Andes méridionales. Cette région qui fait partie du monde andin, tout en étant en relation avec les grandes plaines menant à la région amazonienne pourrait permettre de mieux comprendre les relations entre ces deux environnements.

Tableau 3 : Récapitulatif et données générales concernant l'ensemble des études d'ADN ancien réalisées dans le sous-continent sud-américain.

N°	Populations	Datations	N	Type de données	Référence
1	Momies Colombie	1300-500 B.P.	8	RFLP mt	Monsalve <i>et al.</i> , 1994, 1996
2	Momies Amazonie	500-4000 B.P.	26	HVR-1 (16047-16380)	Ribeiro-dos-Santos <i>et al.</i> , 1996
3	Terre de Feu	100-200 B.P.	60	RFLP mt	Lalueza <i>et al.</i> , 1997
4	Arica	2900-600 B.P.	42	RFLP mt	Moraga <i>et al.</i> , 2001
5	Tiwanaku	n.d.	18	RFLP mt	Rothhammer <i>et al.</i> , 2003
6	Terre de Feu	100-200 B.P.	60	HVR-1 (16154-16400) / 10 STR Y	Garcia-Bour <i>et al.</i> , 2004
7	Sican	1000 B.P.	34	HVR-1 (16209-16402), RFLP mt	Shimada <i>et al.</i> , 2004
8	Arica	3900-500 B.P.	83	HVR-1 (16081-16390)	Moraga <i>et al.</i> , 2005
9	Ollantaytambo	600-400 B.P.	57	HVR-1,2 (128-267; 16209-16402)	Shinoda <i>et al.</i> 2006
10	Chen Chen	1200-1000 B.P.	94	n.d.	Lewis <i>et al.</i> , 2007a
11	Ayacucho	1400-600 B.P.	69	HVR-1 (16011-16382)	Kemp <i>et al.</i> , 2009
12	Palpa, Paracas	2200-1200 B.P.	130	HVR-1 (16021-16408), SNP mt	Fehren-Schmitz <i>et al.</i> , 2010a
13	Laramate	1200-800 B.P.	90	HVR-1 (16021-16408), SNP mt / 6 SNP Y	Fehren-Schmitz <i>et al.</i> , 2010b



Figure 3 : Localisation des populations anciennes amérindiennes pour lesquelles des données d'ADNa ont été obtenues.

Les numéros correspondent aux numéros des populations présentés dans le tableau 3.

II. Peuplement et histoire culturelle des Andes méridionales.

L'objectif de ce chapitre est de présenter les caractéristiques principales de l'histoire de cette région particulière, afin de bien comprendre dans quel contexte se situent les populations anciennes que nous allons étudier. Ainsi, les aspects principaux de leur histoire, notamment leur origine dans la région, leur mode de subsistance, les interactions qu'elles pouvaient avoir avec d'autres populations anciennes seront présentés brièvement.

1. Contexte géographique.

La géographie a influé sur les modes d'établissement de toutes les populations natives américaines et plus particulièrement sur les populations andines.

1.1. Etages écologiques et ressources associées.

Le vaste territoire des Andes méridionales, délimité par le sud de la Bolivie, le nord-ouest de l'Argentine et le nord du Chili (Figure 4) est composé d'une **mosaïque d'environnements écologiques extrêmement contrastés**, qui sont la conséquence de la position de la Cordillère des Andes et des conditions d'altitude et d'humidité qui en résultent. Les populations humaines adaptent leur mode d'établissement et de subsistance aux divers environnements qu'elles rencontrent et donc aux ressources qui sont disponibles (Acreche *et al.*, 2004). Afin de comprendre les raisons de l'établissement des populations dans certaines régions ainsi que leurs modes d'interactions, il est important de rappeler les caractéristiques propres ainsi que les capacités productives de chacun des étages écologiques de cette région qui sont, d'ouest en est : la frange côtière, les hauts-plateaux, les vallées tempérées, le piedmont andin et la plaine subtropicale, représentés sur la figure 4 suivante.



Figure 4 : Présentation des différents étages écologiques des Andes méridionales.

Frange côtière : ce secteur, enserré entre l'océan Pacifique et la cordillère occidentale, est un des environnements les plus arides du monde, notamment au niveau du Chili. Au niveau de l'océan, les ressources ichthyologiques sont très nombreuses, ce qui contraste avec l'absence de ressources alimentaires terrestres, et ce qui a depuis des milliers d'années attiré des groupes de pêcheurs¹ dans cette zone (Rivera, 2008). En allant vers l'est, au pied de la cordillère occidentale cette région est dominée par le désert d'Atacama qui est probablement le plus aride au monde. Seuls quelques cours d'eau arrivent à traverser cette zone depuis la cordillère jusqu'à l'océan, donnant naissance à des **oasis** qui ont été occupés par l'homme moderne tout au long de son histoire dans la région, comme par exemple l'oasis de San Pedro de Atacama (Stovel, 2008). Une des ressources principales de cette région était et est encore aujourd'hui, la présence de nombreux minerais (cuivre essentiellement mais aussi turquoise...) qui ont été utilisés pendant plus de 2 000 ans pour la fabrication de métal ou la réalisation d'objets de culte (Albeck *et al.*, 2003; Albeck and Ruiz, 2003).

Hauts-plateaux : cette zone s'étend depuis le Pérou et la Bolivie (où ils sont appelés

¹ La population Chinchorro a peuplé cette zone il y a environ 10 000 ans et a constitué une des premières populations au monde ayant pratiqué la momification intentionnelle de ces morts.

Altiplano) jusqu'au nord-ouest de l'Argentine (où ils sont appelés **Puna**), à une altitude qui varie entre 3 600 et 4 500 mètres et avec une largeur entre 150 et 220 km. La puna est soumise à des conditions climatiques très rigoureuses ce qui a limité l'agriculture à des espaces très restreints et à des variétés de végétaux résistants aux basses températures. Cette zone a été d'une grande importance pour le développement de l'élevage de camélidés, ce qui a notamment permis l'organisation de **circuits de caravanes de lamas** qui parcouraient la puna autant en direction nord-sud qu'est-ouest pour le transport de produits vers les zones écologiques adjacentes (Albeck, 2001). La richesse de cette région provenait de toute la matière première ainsi que des produits dérivés de l'élevage de camélidés (viande, textiles, laine...) qui étaient des ressources très recherchées dans les autres régions (Tarrago, 1999).

Vallées tempérées (Valliserrana) : cette zone de vallées fertiles qui entrecoupe les hauts-plateaux à des altitudes comprises entre 2 500 et 3 600 mètres est une aire essentiellement agricole où prospère une grande variété de végétaux. Dans cette zone se trouvent plusieurs vallées profondes, ou quebrada¹, qui ont été utilisées dès l'arrivée des hommes dans cette région comme **voies naturelles de passage**, notamment entre les hauts-plateaux et les plaines. Depuis près de 11 000 ans, la circulation des populations, des biens et des produits a été intense dans ces vallées. Parmi les voies de passage les plus empruntées se trouve la Quebrada de Humahuaca, tout au nord de l'Argentine, qui est considérée par certains auteurs comme une région à part, notamment car une autre langue que celle parlée dans le reste des Andes méridionales y était parlée lors de l'arrivée des Européens (Nielsen, 2001).

La richesse de cette aire environnementale était fondée sur la production de céréales et de légumes en excédent, et sur une spécialisation artisanale dans la production de produits manufacturés de grande valeur (Tarrago, 1999).

Piedmont andin : cette zone que l'on appelle grossièrement « jungle » est divisée en deux aires distinctes, les yungas au niveau de la Bolivie, et les sierras (ou selvas) au niveau de l'Argentine. Dans cette zone, une agriculture itinérante par défrichage et brûlis permettaient la culture de nombreuses espèces, parmi lesquelles le maïs occupait une place particulière. Cet étage écologique offrait de nombreuses ressources, recherchées par toutes les populations anciennes des Andes méridionales, comme le bois (pour les arcs notamment), les plumes ou encore les hallucinogènes (importants pour les rituels) (Tarrago, 1999).

¹ Quebrada est le nom espagnol donné aux vallées étroites, s'apparentant parfois à des canyons, desservant les hauts-plateaux andins.

Plaines subtropicale : ces plaines couvertes de forêts et parcourues de cours d'eau rejoignant l'Atlantique, ne font pas à proprement parlé partie de l'aire des Andes méridionales. Il s'agit plutôt d'une **aire transitoire** faisant le lien entre le monde andin et le monde amazonien. Si des liens entre les aires mentionnées précédemment sont avérés, les liens avec ces plaines subtropicales ne sont pas encore vraiment clairs.

1.2. Mobilité des populations.

Chaque étage écologique présenté précédemment possède des ressources essentielles. Les populations ont donc la nécessité d'établir des **communications** entre les différentes zones écologiques complémentaires dans le but d'assurer leur survie (Albeck *et al.*, 2003). Les implications d'une telle mobilité nécessaire à la subsistance des populations ne sont pas seulement économiques puisque, parallèlement à la **circulation des biens et des produits** transitent aussi des connaissances technologiques, des idéologies et des pratiques symboliques (Albeck *et al.*, 2003).

Les interactions humaines et l'échange de ressources entre les différentes zones écologiques sont attestés depuis des périodes très reculées. Les plus anciennes preuves archéologiques d'une telle mobilité ont été découvertes dans la Quebrada de Humahuaca. Certains vestiges mis au jour dans des grottes de la vallée provenaient aussi bien des plaines du Chaco, du piedmont andin que de la puna et dataient, pour certains, du dixième millénaire avant J.-C. (Albeck *et al.*, 2003). Pour ces populations de chasseurs-cueilleurs, certaines ressources provenant d'autres environnements étaient indispensables pour leur survie, comme les pierres servant à tailler des outils ou le bois permettant de confectionner des arcs et des flèches.

Avec l'apparition des sociétés agricoles, les liens entre les groupes qui peuplaient les différents étages écologiques se sont maintenus et même développés. L'offre inégale des ressources naturelles de chaque zone a conduit à un flux constant des biens et produits permettant d'assurer le **contrôle vertical** du milieu. Pour cela, des groupes spécialisés dans le transport ont émergé : les caravanes de lamas (*caravaneros*), qui faisaient le lien entre les plaines à l'est, et les oasis de la zone côtière (Tarrago, 1999).

La collaboration entre les différents groupes, la nécessité d'une organisation et d'une logistique importante mais aussi la centralisation permettant de redistribuer la production,

sont très probablement parmi les raisons principales qui ont conduit au développement de sociétés complexes dans la région (comme dans les Andes centrales), contrairement à d'autres régions où les populations, ayant toutes les ressources indispensables facilement à disposition, ont vécu un développement moins important (jusqu'à l'arrivée des européens les groupes de la Pampa et la Patagonie argentine sont restés nomades et essentiellement chasseurs-cueilleurs).

2. Contexte historique.

2.1. Contexte chronologique.

Les Andes méridionales présentent une multitude de traditions locales, le plus souvent définies à partir de styles céramiques, qui ont été utilisés pour décrire des périodes ou des phases chronologiques. Cependant, selon chaque auteur la chronologie ou l'enchaînement de ces phases peut varier fortement, ce qui complique l'étude des populations locales. Afin de pouvoir comparer correctement les groupes des différentes aires écologiques, nous avons choisi d'utiliser une des chronologies le plus souvent employée pour cette région, qui semble la plus simple et qui n'étant pas seulement fondée sur les styles céramiques paraît plus fiable (Albeck and Ruiz, 2003; Nunez Regueiro and Tartusi, 2003).

Comme cela est présenté sur la figure 5 suivante, la chronologie utilisée est divisée en deux périodes principales : la **période pré-céramique** (ou archaïque), pendant laquelle la majorité des populations étaient encore mobiles et dont le mode de subsistance était essentiellement basée sur la chasse, et la **période céramique**, elle-même subdivisée en trois périodes (initiale, intermédiaire et tardive), caractérisée par des populations sédentaires et agricoles. Les différentes cultures ou traditions des populations anciennes étudiées ainsi que des populations des différentes aires écologiques mentionnées au cours de ce travail, sont représentées sur la figure 5.

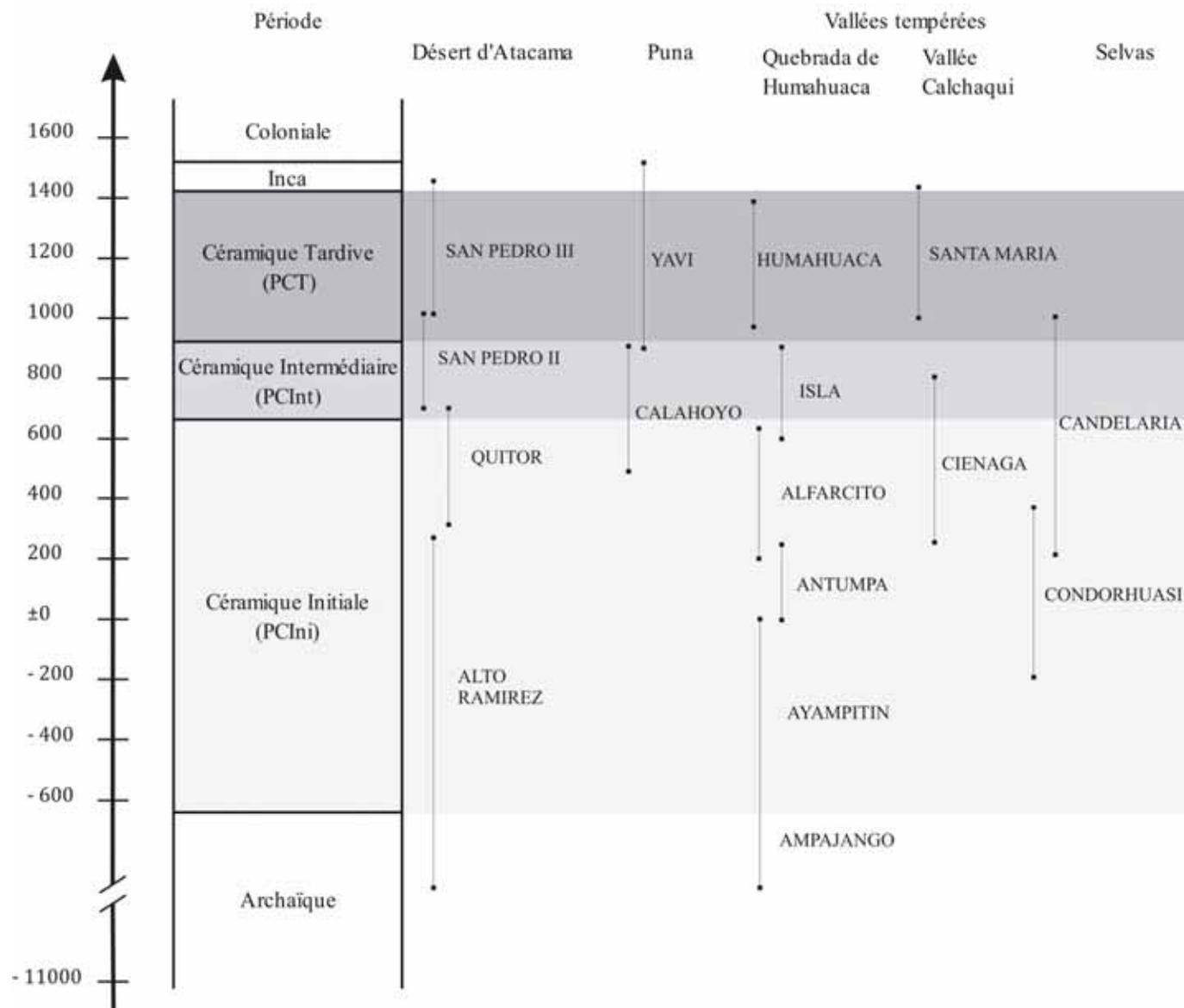


Figure 5 : Périodes et traditions culturelles des différentes aires écologiques des Andes méridionales, mentionnées au cours de ce travail.

2.2. Du peuplement initial à la colonisation.

2.2.1. *Période archaïque : 10 000 – 1 000 av. J.-C.*

Autour de 10 000 ans av. J.-C., à la fin du pléistocène, des groupes très mobiles de chasseurs-cueilleurs ont peuplé quelques zones de la puna, des vallées tempérées et du désert d'Atacama, en exploitant les aires proches des réseaux hydrographiques souvent riches en nourriture (Olivera, 2001; Tarrago, 1999). Les évidences archéologiques découvertes dans certains campements des vallées menant à la puna de Jujuy (la Quebrada de Humahuaca entre autre) suggèrent que la stratégie de subsistance de ces populations était fondée sur l'optimisation des ressources disponibles dans chaque étage écologique (Tarrago, 1999). Le schéma de mobilité devait inclure l'utilisation de plusieurs campements saisonniers, comme la grotte de Huachichocana, qui semble-t-il était occupé pendant le printemps par des groupes pratiquant la chasse de camélidés, qui occupaient par la suite des installations plus permanentes (comme Inca Cueva dans la Quebrada de Humahuaca) où la subsistance était fondée sur les rongeurs et les végétaux (Tarrago, 1999). Un mode de subsistance identique a été décrit pour les groupes de la région d'Atacama (Rivera, 2008).

La période archaïque finale voit apparaître la gestation des activités agricoles et pastorales, à partir de 2 000 ans av. J.-C. Cette période peut s'apparenter à un processus de néolithisation. L'apparition de la production agricole et de l'élevage a conduit à une **démographie croissante** ainsi qu'à la **sédentarisation des populations** (Tarrago, 1999). Au cours des derniers millénaires avant notre ère, des sociétés au mode de subsistance mixte (agriculture sans délaisser la chasse et la collecte) se sont développées dans la puna et les vallées (Tarrago, 1999). Ce nouveau type d'organisation a conduit à des innovations comme l'apparition de la céramique (environ 1 000 ans av. J.-C. dans la puna) (Olivera, 2001). L'origine de cette néolithisation reste encore très floue pour cette région comme pour le reste de l'Amérique du Sud en général. En effet, la transition vers une économie basée sur l'agriculture et la sédentarisation est mal comprise. Il n'est pas encore établi s'il s'agit de l'évolution sur place des populations, ou s'il s'agit de l'arrivée de populations, venant de régions plus au nord, (Andes centrales notamment) qui possédaient déjà ces techniques (Lavallée, 2006).

2.2.2. Période céramique initiale : 1 000 av. J.-C. – 650 ap. J.-C.

La période céramique initiale (P.C.Ini) voit le développement et la consolidation des **premières communautés villageoises sédentaires et égalitaires** (Leoni and Acuto, 2008; Tarrago, 1999; Tartusi and Nunez Regueiro, 2001). Ces petits villages ou hameaux ont été retrouvés dans l'ensemble des aires géographiques des Andes méridionales, depuis le désert d'Atacama jusqu'au piedmont andin.

Cultures et traditions.

Plusieurs traditions culturelles différentes ont été décrites pour cette période, comme la tradition Alto Ramirez dans l'oasis d'Atacama, ou les traditions Alfarcito, Cienaga et Condorhuasi dans les vallées et sierras (Figure 5). Cependant il apparaît que toutes ces traditions étaient en interaction constante et des céramiques de tous les styles ont été découvertes au travers de toutes les Andes méridionales (Olivera, 2001; Tartusi and Nunez Regueiro, 2003). Ainsi, même si de petites différences culturelles peuvent être observées, l'économie et l'organisation de toutes les populations de la région étaient relativement identiques.

Economie.

L'économie de ces populations était essentiellement basée sur l'agriculture et l'élevage qui étaient encore à cette époque complétés intensément par la chasse et la collecte. Les productions céramiques, textiles et métallurgiques développées durant cette période ont atteint un niveau très élevé, aussi bien en ce qui concerne la technique que l'esthétique (Leoni and Acuto, 2008). Même s'ils semblaient autonomes, ces villages n'étaient en aucun cas isolés. L'ensemble des populations des différents étages écologiques étaient connectées au travers d'échanges sur de courtes ou des plus longues distances, favorisant la circulation de matières premières, de produits manufacturés, d'idéologie ou encore de populations (Leoni and Acuto, 2008).

Organisation sociale.

Le patron d'établissement typique de cette période dans le désert d'Atacama, la puna et les vallées, est constitué d'habitations regroupées en hameaux, ou alors d'habitations disséminées entre les champs de culture (Tarrago, 1999). Ce schéma reflète une **organisation sociale simple**, essentiellement fondée sur la sphère domestique et les **liens de parenté** (Tartusi and

Nunez Regueiro, 2001). Cette organisation correspond peut être à la naissance des *ayllus* qui était, et sont encore aujourd'hui, des communautés composées de plusieurs familles dont les membres considèrent qu'ils ont une origine commune (réelle ou fictive). Tous les membres de ces *ayllus* travaillent de façon collective dans un territoire commun leur appartenant.

L'ensemble des habitations étaient en relation étroite avec les unités de production (champs, cours d'eau ou pâturages) (Leoni and Acuto, 2008). Ainsi, les tâches domestiques et agricoles étaient très souvent imbriquées. Il n'était pas rare que ces groupes complètent leur résidence principale par des sites complémentaires situés dans les étages écologiques voisins, permettant de récupérer d'autres types de matières premières (Leoni and Acuto, 2008). L'ensemble des activités (de production ou rituelles) étaient généralement effectuées au sein de la sphère domestique (Leoni and Acuto, 2008), et les preuves d'activités extra-domestiques rituelles ou d'espaces publics sont rares, laissant supposer que l'ensemble des individus de ces communautés possédaient le même statut social (Leoni and Acuto, 2008).

En ce qui concerne les sierras du piedmont, une organisation plus simple avec des huttes en matériaux périssables, s'est mit en place, comme par exemple dans certains villages de la culture Candelaria (Tarrago, 1999). Cependant, peu de données sont disponibles sur les groupes de cette zone.

Pratiques funéraires.

Les sépultures étaient le plus souvent des sépultures primaires, accompagnées de nombreuses offrandes funéraires, et situés le plus souvent dans les habitations ou les patios (Leoni and Acuto, 2008). Si les individus adultes étaient le plus souvent inhumés directement en pleine terre, les immatures pouvaient être inhumés dans des urnes. Les nécropoles séparées n'étaient pas nombreuses mais quelques unes ont été découvertes, notamment dans le nord de la vallée Calchaqui et dans la puna.

2.2.3. Période céramique intermédiaire : 650 – 900 ap. J.-C.

La période céramique intermédiaire (P.C.Int) est considérée comme une période d'**intégration régionale** qui voit le développement de **deux sphères d'interactions économiques et sociales indépendantes**. La première sphère comprenait les populations du nord des Andes méridionales : puna de Jujuy, les vallées tempérées de Salta et Jujuy (Quebrada de Humahuaca et vallée Calchaqui) ainsi que l'oasis d'Atacama, où plusieurs

traditions culturelles ont pu cohabiter. La deuxième sphère d'interaction plus au sud, était sous influence de la culture Aguada, dont le cœur se situait dans la vallée d'Ambato (province argentine de Catamarca) et qui s'est développée jusqu'au sud de la vallée Calchaqui (Nunez Regueiro and Tartusi, 2002; Tarrago, 1999).

Cultures et traditions.

Contrairement à la période précédente on observe pendant cette période intermédiaire le développement de quelques spécificités régionales notables, notamment des différences entre les deux sphères évoquées précédemment.

La première sphère d'interaction, la plus au nord, présente des interactions fortes avec les sociétés de *l'altiplano*, de la région du lac Titicaca en Bolivie. La tradition Quitar de la région de San Pedro de Atacama, par exemple, possède des liens forts avec la civilisation Tiwanaku (Rivera, 2008). Il n'est cependant pas encore établi si les similitudes observées sont dues à des mouvements de populations depuis la région des hauts-plateaux boliviens ou seulement à une forte influence idéologique (Rivera, 2008).

La deuxième sphère d'influence plus au sud, voit l'ensemble des populations regroupées ou au moins fortement influencées par la culture locale Aguada, sans relations avec d'autres régions. Dans les sierras adjacentes à la vallée Calchaqui on observe le développement de la culture candelaria, dont les premières phases sont apparues à la période précédente, et qui fait partie du réseau d'interactions de la Aguada (Baldini, 2003).

Economie.

L'agriculture atteint un degré supplémentaire pendant cette période avec l'apparition et le développement de **l'irrigation** qui permet de coloniser de nouveaux espaces agricoles. De plus, l'élevage de camélidés s'intensifie fortement (Tarrago, 1999). Le nombre important de céramiques manufacturées, d'instruments en bois ou en os mais aussi d'objets de cultes en or et en argent dans tous les sites de l'oasis d'Atacama, de la puna de Jujuy et des vallées, confirment les interactions importantes dans toute la zone ainsi qu'avec les hauts-plateaux boliviens (Tarrago, 1999).

Organisation sociale.

Les hameaux de la période précédente deviennent des villages, qui attestent d'une augmentation de la densité de la population durant cette période. A certains endroits (dans le site de Tilcara de la Quebrada de Humahuaca notamment) on commence même à remarquer

les prémices de réseaux urbains, et d'agglomérations (Tarrago, 1999). Un des changements significatif avec la période précédente réside dans l'augmentation du nombre d'espaces publics rencontrés (allant de petites places à des centres cérémoniels de grande envergure), révélant une évolution de l'identité des communautés au niveau local mais aussi régional (Leoni and Acuto, 2008).

Pratiques funéraires.

Les pratiques funéraires n'évoluent pas fondamentalement pendant cette période. Elles restent associées pour la plupart à la sphère domestique. La seule différence notable est l'apparition de crânes isolés que l'on appelle également « crânes trophées », souvent le signe de tensions entre différents groupes.

2.2.4. Période céramique tardive : 900 – 1480 ap. J.-C.

Cette période céramique tardive (P.C.T) est généralement caractérisée par des **développements régionaux** atteignant une complexité inégalée jusqu'alors, avec l'apparition de **conflits** et une **augmentation des inégalités sociale et de la complexité politique** (Leoni and Acuto, 2008; Rivolta, 2007; Seldes, 2006; Tarrago, 1999).

Cultures et traditions.

Les cultures de l'oasis d'Atacama (San Pedro III) ainsi que du nord de la puna (Yavi) présentent des interactions très fortes avec le sud de la Bolivie (Albeck, 2001; Avila, 2009; Stovel, 2008). Dans les vallées par contre, les cultures qui s'y développent possèdent des liens plus étroits avec les traditions observées dans le piedmont andin.

Economie.

L'agriculture et l'élevage intensif mis en place à la période précédente restent le fondement de l'économie de ces populations, en étant toujours associés à des circulations importantes de biens et de produits permettant le contrôle vertical de toutes les aires écologiques. Ce contrôle se faisait notamment avec l'implantation de « colonies » dans des régions voisines. Par exemple, les agglomérations de Humahuaca et Tilcara situés dans l'aire écologique des vallées, possédaient des installations dans la zone des sierras plus à l'est (Tarrago, 1999).

Les céramiques de cette période apparaissent plus simples, mais les modes de production sont

beaucoup plus complexes et sophistiquées, ce qui conduit à une production de biens artisanaux très importante, aussi bien en terme de quantité que de qualité (Tarrago, 1999).

Organisation sociale.

Les sites d'habitation de cette période sont de tailles différentes entre les étages écologiques, mais aussi au sein même de ces aires géographiques. En effet, dans la puna et dans le piedmont andin, les sites restent relativement petits, comme à la période précédente. Au contraire, dans les vallées tempérées, et plus particulièrement dans la Quebrada de Humahuaca ou la vallée Calchaqui, de véritables **agglomérations** se mettent en place, pendant la deuxième moitié de la période tardive (période de développements régionaux II, D.R.II) (Albeck *et al.*, 2003; Korstanje *et al.*, 2010). La croissance de ces villages en agglomérations est probablement liée à leur emplacement stratégiques sur les routes naturelles de communication entre la puna et les zones boisées orientales (Albeck *et al.*, 2003). Ces grands complexes résidentiels se présentent comme des agglomérations d'unités domestiques, parcourues de voies de circulations, et entrecoupées d'espaces ouverts, pouvant être des places publiques, des espaces réservés au stockage des déchets ou encore des parcs pour garder les animaux (Albeck and Ruiz, 2003). Les unités domestiques sont généralement composées, comme aux périodes précédentes, de plusieurs enceintes articulées autour d'un patio ouvert.

Les différences observées concernant la taille des sites ont été associées au développement d'un **pouvoir politique hiérarchisé** (Nielsen, 2001), ce qui est en accord avec les signes démontrant le développement de chefferies, comme ceux de Humahuaca, Tilcara (qui font partie de l'ethnie des omaguacas), Yavi (qui font partie de l'ethnie des atacamas ou atacamenos) ou encore Casabindo (Tarrago, 1999). La centralisation du pouvoir va induire, surtout pendant la deuxième phase de cette période (D.R.II) de plus en plus d'inégalités sociales, mais aussi des luttes, aussi bien intra qu'inter-régionales, attestées par la nette augmentation de **sites fortifiés** et de *pukaras*, qui sont des tours situées dans des endroits stratégiques permettant de surveiller les alentours (Tarrago, 1999). La localisation des principales ethnies précolombiennes présentent dans la région, ainsi que les interactions qu'elles entretenaient sont illustrés sur la figure 6.

Pratiques funéraires.

En ce qui concerne les vallées tempérées, la grande majorité des sépultures de cette période sont des **tombes collectives** (Acuto, 2007). Ces pratiques étaient le plus souvent associées à la

sphère domestique, et n'étaient pas des rites communautaires réalisés dans des espaces communs (Acuto, 2007). Même si pour certains sites, des espaces particuliers, pouvant être considérés comme des nécropoles, étaient dédiés aux inhumations, dans la grande majorité des cas, les inhumations étaient plutôt associées à des complexes résidentiels (Acuto, 2007). Lorsque des nécropoles sont identifiées, elles se situent toujours très près des sites d'habitations, associées aux populations. Enfin, les offrandes funéraires étaient pour la plupart des objets d'usage domestique et quotidien.

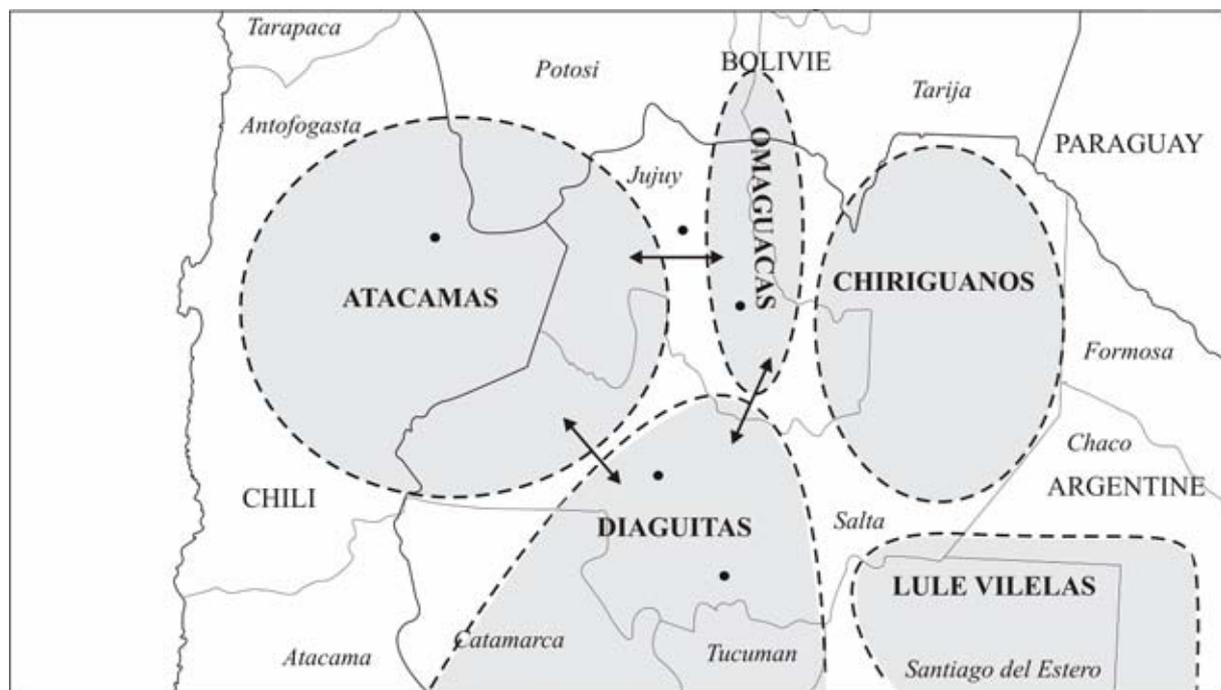


Figure 6 : Distribution des principales ethnies des Andes méridionales.

2.2.5. Phase Inca : 1480 – 1530 ap. J.-C.

Malgré la très courte durée de cette période, l'annexion de la région à l'empire Inca (province du *Tawantisuyu*) a provoqué des changements profonds dans le paysage, le mode de vie, mais aussi la répartition des populations autochtones des Andes méridionales. En effet, le schéma de complémentarité économique et les réseaux mis en place aux périodes précédentes ont été complètement bouleversés par les Incas, qui ont renforcé les circuits dans le sens nord-sud

tout en délaissant les circuits dans le sens est-ouest (Albeck and Ruiz, 2003; Rivera, 2008; Tarrago, 1999). De plus, alors qu'avant, tous les sites pratiquaient un grand nombre de type de productions différentes, les Incas vont spécialiser chaque site dans un type de production particulier, et envoyer la majorité des produits dans les régions de l'empire situées plus au nord. Enfin, la soumission des populations locales est perçue, à travers une baisse de la densité de population qui est à mettre en relation avec les grands mouvements de populations (*mitimaes*) initiés par les incas afin d'obtenir le contrôle aussi bien des populations réfractaires, que des ressources, mais aussi au travers des sanctuaires de hautes montagnes où étaient pratiqués des sacrifices (Acuto, 2001), témoignages de la religion officielle. Ainsi, les populations des Andes méridionales étant celles qui ont le plus longtemps résisté au pouvoir Incas, refusant pendant longtemps de réaliser des travaux au profit de l'empire, des tribus Chichas notamment de l'actuelle Bolivie y ont été déplacé en grand nombre par les Incas pour y réaliser les travaux agricoles ainsi que pour diffuser la langue officielle : le Quechua. Cette période est donc une période de luttes et de combats qui s'est poursuivie par la suite avec l'arrivée des colons européens vers 1536.

3. Diversité biologique contemporaine dans les Andes méridionales.

A l'arrivée des européens, la région du nord ouest argentin constituait, avec environ 200 000 individus, la région la plus peuplée d'Argentine (Acreche *et al.*, 2004; Sala *et al.*, 1999). Plusieurs langues étaient parlées dans cette région : la cacan dans les vallées, le kunsu dans la puna de Jujuy, l'aymara dans la zone frontalière avec la Bolivie ainsi que le quechua qui était la langue officielle de l'empire Inca (Tarrago, 1999). Les populations des Andes méridionales ont été celles de toute l'Amérique qui ont opposé la plus farouche opposition aux européens, conduisant à une guerre totale de cent ans jusqu'au XVII^{ème} siècle (Tarrago, 1999).

Aujourd'hui la population de ces régions est largement métissée, avec évidemment une part importante de la population qui est d'origine espagnole ou portugaise. Si ces populations métissées occupent actuellement les grandes villes et agglomérations, les régions les plus inhospitalières, notamment de la puna continuent à être peuplées par des populations amérindiennes, et entre autre par les Coyas (qui ont remplacé les omaguacas et les diaguitas de l'époque précolombienne). Il y aurait actuellement environ 600 000 individus se présentant comme d'origine strictement amérindienne en Argentine, confinés aux régions de l'extrême

nord ou du sud-ouest (Toscanini *et al.*, 2011). La population des Coyas que l'on retrouve actuellement dans les provinces de Salta et de Jujuy dans le nord ouest argentin possède une diversité génétique que l'on peut qualifier de typiquement andine avec de fortes fréquences de la lignée maternelle B2, et quasiment exclusivement la lignée paternelle Q1a3a* (Bobillo *et al.*, 2010).

Le métissage des populations semble s'être réalisé préférentiellement entre hommes européens et femmes amérindiennes (ce qui est en accord avec l'arrivée massive des colons européens qui étaient en majorité des hommes). Cependant, nous pouvons remarquer que dans certaines régions le métissage notamment des lignées paternelles a été moins marqué. Ainsi, alors que la population contemporaine de Humahuaca présente près de 84 % de lignées paternelles natives, les populations de la province de Tucuman ne possèdent que 12 % d'haplogroupes spécifiques de l'Amérique (Bailliet *et al.*, 2009). Le pool génique mitochondrial, mis à part dans les grandes villes, reste généralement dominé par des lignées amérindiennes (entre 20 % et 90 % en fonction des zones étudiées, et en moyenne 58 % de lignées natives sur l'ensemble de l'argentine) (Corach *et al.*, 2010; Corach *et al.*, 2006).

III. Hypothèses et objectifs.

Les objectifs de ce travail de thèse sont multiples. Grâce à l'analyse de la diversité génétique d'échantillons humains anciens provenant des différentes aires écologiques des Andes méridionales, plusieurs problématiques particulières pourront être abordées :

1. Le premier but principal sera de caractériser la diversité génétique des populations précolombiennes de l'aire andine méridionale, encore inconnue à ce jour. La diversité de marqueurs mitochondriaux aussi bien que nucléaires sera étudiée, afin d'obtenir le plus d'informations pertinentes possibles.

A partir de cette diversité, plusieurs points concernant l'histoire de cette région pourront être investigués :

2. Les relations génétiques entre populations anciennes des différentes aires écologiques pourront être analysées. Etant donné les interactions intenses entre les populations de la région, il serait intéressant de savoir si ces échanges ont été accompagnés de flux géniques, auquel cas on s'attendrait à observer peu de divergence génétique entre les populations anciennes étudiées.
3. Les relations et affinités avec les populations anciennes des Andes centrales seront également testées afin de savoir si l'influence des grandes civilisations andines (toutes localisées dans la région des Andes centrales) est perceptible dans le pool génique des populations de la partie méridionale des Andes.
4. En ce qui concerne la population particulière de l'oasis de San Pedro de Atacama, l'influence de la civilisation Tiwanaku pourra être révélée grâce à l'analyse de la diversité génétique de populations des périodes pré-Tiwanaku et Tiwanaku. Ainsi, une migration éventuelle de population depuis l'altiplano bolivien pourra être vérifiée par la discontinuité ou la continuité du pool génétique de ces populations.

5. Pour certains sites particuliers, les pratiques funéraires seront analysées, permettant d'évaluer l'apparente homogénéité culturelle et sociale de la région.
6. Enfin, l'évolution des populations amérindiennes dans les Andes méridionales sera évaluée grâce à la comparaison avec les populations amérindiennes contemporaines. Ainsi, l'impact de l'arrivée des Inca dans la région, ainsi que de la colonisation européenne, pourront être estimés.

L'ensemble de ces problématiques permettra d'appréhender de manière plus globale la **dynamique de peuplement de la région des Andes méridionales**, ainsi que l'**évolution des populations amérindiennes** depuis les périodes préhispaniques jusqu'à nos jours.

L'obtention de nouvelles données d'ADN ancien sur des populations amérindiennes permettra d'approcher quelques problématiques concernant la dynamique du peuplement de l'aire andine en général mais aussi la colonisation de l'Amérique du Sud.

7. L'homogénéité apparente des populations andine sera ainsi à nouveau testée grâce à ces nouvelles données, afin d'essayer de comprendre la mise en place d'un tel schéma de diversité.

Un second volet de cette étude concerne des aspects plus techniques de l'analyse de l'ADN ancien. Les caractéristiques des molécules d'ADN ancien impliquent une recherche constante pour l'amélioration des protocoles d'analyses. Ce travail a donc été l'occasion d'appliquer de nouvelles méthodologies à l'étude des molécules dégradées. Une technique de génotypage par spectrométrie de masse MALDI-TOF (Sequenom) a été mise en place pour l'analyse des échantillons anciens. Les bénéfices et implications de cette technique seront discutés au cours de ce travail.

Deuxième partie.

Matériel et méthodes.

I. Sites et échantillons analysés.

1. Stratégie d'échantillonnage.

Afin de comprendre au mieux la dynamique de peuplement et l'évolution des populations anciennes de l'aire andine méridionale, nous avons choisi de mettre en place une stratégie fondée sur l'analyse d'échantillons des différentes zones écologiques de la région, mais aussi de périodes différentes.

L'accès aux restes humains anciens est souvent un des facteurs les plus limitant pour les études paléogénétiques, en particulier en Amérique du Sud. Le faible nombre d'études en ADN ancien démontre la difficulté d'avoir accès à de tels échantillons.

Plusieurs raisons peuvent expliquer ce phénomène. Premièrement, l'accès aux échantillons dépend inévitablement des sites fouillés, ainsi que des méthodes de fouilles. En ce qui concerne les Andes méridionales, la majorité des collections sont issues de fouilles réalisées au début du XX^{ème} siècle, période pendant laquelle les méthodes ne préconisaient pas l'enregistrement systématique du contexte dans lequel les restes étaient découverts. Ainsi, les études peuvent être compliquées car les contextes de nombreuses collections ne sont pas connus. De plus, la majorité des sites des périodes archaïque, céramique initiale et intermédiaire se trouvent sous des constructions contemporaines et ne sont donc pas accessibles. Deuxièmement, l'aspect destructif des analyses et l'incertitude quant aux résultats peuvent poser problème aux dépositaires de ces échantillons, qui sont extrêmement précieux. Enfin, la situation en Amérique du Sud pour l'analyse d'échantillons humains anciens est particulière. De nombreux pays, ou régions possèdent des législations particulières pouvant restreindre l'accès à ce type d'échantillon. Ainsi, en Argentine la loi 25.517 promulguée le 14 décembre 2001 établit que :

«... deberán ser puestos a disposición de los pueblos indígenas y/o comunidades de pertenencia que lo reclamen, los restos mortales de aborígenes, que formen parte de museos y/o colecciones públicas o privadas.»

[... les restes humains anciens de populations autochtones faisant partie des musées et/ou de collections publiques ou privées devront être mis à disposition des communautés locales qui les réclament.] (Palma *et al.*, 2006).

Avec l'aide de l'équipe du Pr. Nielsen de l'université de Buenos Aires (CONICET¹), nous avons pu accéder à des échantillons provenant de plusieurs collections archéologiques du nord-ouest argentin. Parmi ces collections, nous avons eu accès à des échantillons provenant de nombreux sites archéologiques de la vallée de la Quebrada de Humahuaca, située dans la région de Jujuy et nous avons également eu accès à des échantillons anciens provenant de différents sites de la vallée Calchaqui. Les sites de la puna étant moins nombreux, nous n'avons eu accès qu'à quelques échantillons du site de Doncellas situé dans cette région. Avec l'aide de l'équipe du Pr. Carnese, de l'Institut des Sciences Anthropologiques de l'université de Buenos Aires, nous avons pu étudier le site de Pampa Grande situé dans le piedmont de la région de Salta. Nous avons également réalisé une mission au Chili, plus particulièrement au musée « Gustavo Le Paige » de San Pedro de Atacama, géré par l'université Catholique du Nord qui nous a permis d'étudier certains échantillons provenant de leurs nombreuses collections. Ainsi, nous avons eu accès à des échantillons provenant du site archéologique de Solcor 3, situé dans l'oasis de San Pedro de Atacama, au Nord du Chili.

L'ensemble des échantillons a été rapatrié en France après avoir obtenu toutes les autorisations nécessaires, et pour certains échantillons, après que des répliques exactes aient été réalisées.

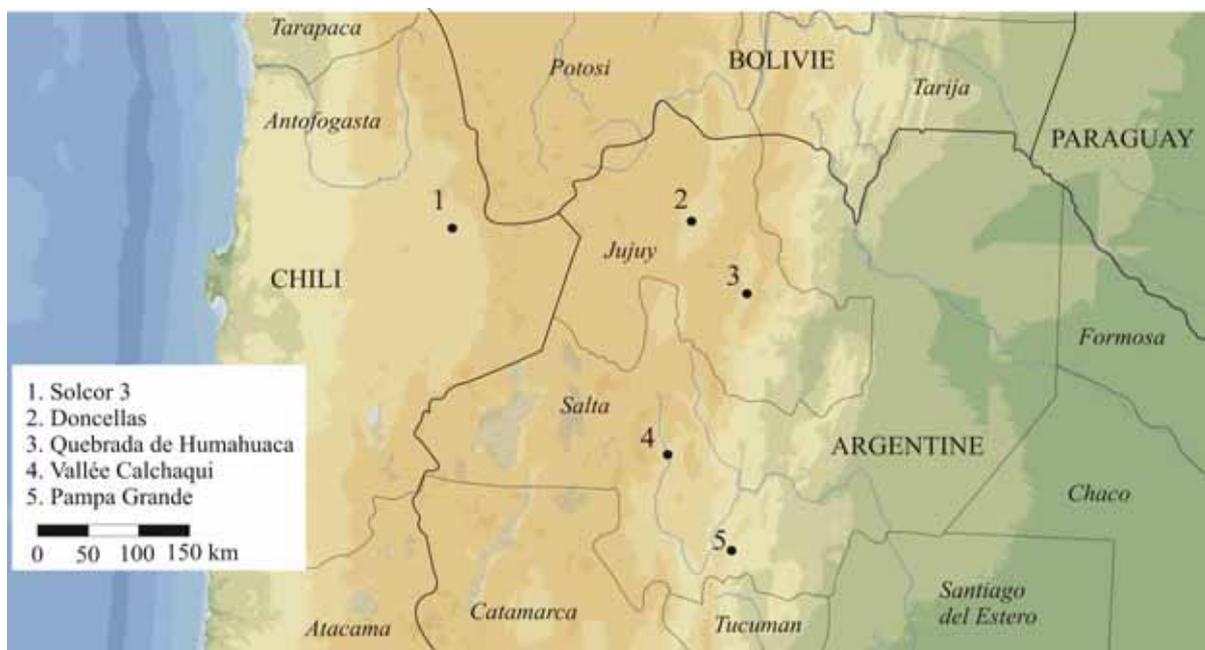


Figure 7 : Carte représentant la situation des sites anciens étudiés.

¹ Le CONICET argentin est l'équivalent du CNRS en France.

Ainsi, nous avons récupéré un ensemble de 145 échantillons, provenant du désert d'Atacama, des vallées tempérées (Quebrada de Humahuaca et Vallée Calchaqui), de la puna et du piedmont andin, datés des trois périodes céramiques principales. La répartition des échantillons en fonction des aires géographiques et des périodes, ainsi que leur localisation sont illustrées sur les figures 7 et 8.

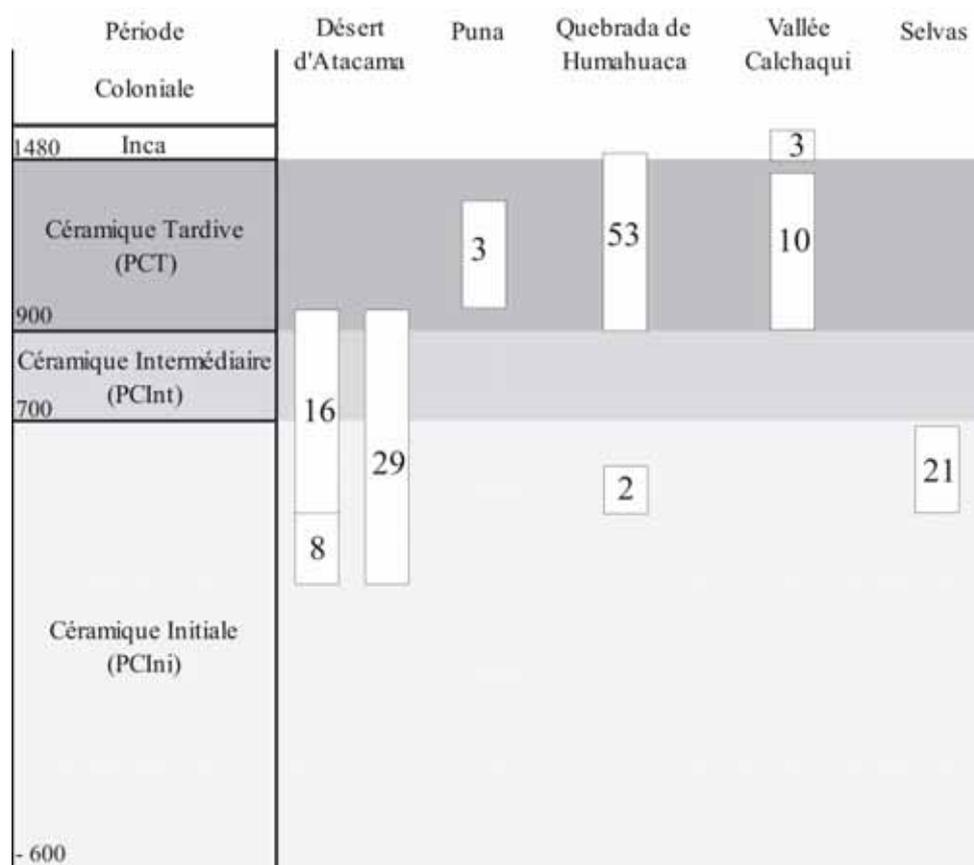


Figure 8 : Nombre et répartition des échantillons anciens analysés, en fonction de la période et de la région.

2. Situation et description des échantillons.

Même si l'objectif principal de l'étude est d'essayer d'appréhender la dynamique de peuplement de la région et l'évolution des populations amérindiennes locales, chaque site archéologique d'où proviennent les échantillons étudiés possède des caractéristiques propres

que nous détaillerons brièvement par la suite pour comprendre le contexte de chacune des études.

Il est cependant important de noter que les contextes dans lesquels ont été découverts les différents échantillons peuvent être difficiles à retracer, l'accès aux informations étant parfois restreint. Comme cela a déjà été mentionné, beaucoup des fouilles, pendant lesquelles les échantillons ont été prélevés, ont été réalisées durant la première moitié du XX^{ème} siècle. Les techniques de fouilles de cette époque ne préconisaient pas l'enregistrement systématique de la localisation du matériel prélevé (Zaburlin, 2008). De plus, de nombreux résultats de ces fouilles n'ont pas été publiés. Enfin, certains sites ont fait l'objet de très nombreuses études alors que d'autres restent quasiment inconnus. Il en résulte une inégalité quant à la quantité et la qualité des données disponibles, ainsi qu'un manque d'information quant à certains contextes archéologiques.

2.1. Quebrada de Humahuaca.

Enclavée au centre de la province argentine de Jujuy et parcourue par le Río Grande, la Quebrada de Humahuaca a constitué et constitue encore un passage naturel reliant différents étages écologiques. Plus encore, cette vallée est **le lien physique le plus important** entre les hautes terres du plateau andin et les grandes plaines tempérées du sud-est de l'Amérique du Sud, mais comprend aussi des chemins transversaux est-ouest qui relie la puna aux sierras et à la plaine du Chaco. La profusion de sites et de gisements archéologiques, de toutes les périodes, témoigne de **l'utilisation de cet itinéraire de manière constante tout au long des dix mille ans de l'histoire de l'homme dans la région**, sans avoir perdu de vigueur entre les moments les plus anciens jusqu'aux périodes contemporaines. La grande quantité de sites archéologiques (plus de cent sites sont répertoriés dans cette vallée qui fait environ 150 km de long) permet de suivre l'évolution des populations qui ont peuplé ou qui ont transité par cette vallée au cours des derniers 10 000 ans. Toutes ces caractéristiques ont conduit à l'intégration de la Quebrada de Humahuaca dans la liste du patrimoine mondial de l'humanité en 2002, en tant que patrimoine culturel et naturel.

Nous avons eu accès à des échantillons humains anciens provenant de 9 sites de cette vallée, essentiellement datés de la période céramique tardive (seulement 2 échantillons sont datés de la période initiale). Les caractéristiques principales ainsi que la localisation de chacun des

sites sont présentées dans le tableau 4 et la figure 9 suivants. Une attention plus particulière sera portée au site de Los Amarillos pour lequel un nombre significatif d'échantillons a été analysé et pour lequel une problématique particulière concernant le recrutement funéraire a été posée.

Tableau 4 : Description des sites et des échantillons de la Quebrada de Humahuaca analysés.

Sites	Période / Phase	Type de site	Code échantillons analysés
Banda de Perchel (BP)	P.C.T / D.R.II	Terrasse domestique	3 (BP-1; BP-2; BP-3)
Huacalera (Hu)	P.C.T / D.R.I	Résidentiel	1 (Hu-1)
Juella (Ju)	P.C.T	Agglomération	5 (Ju-1; Ju-2; Ju-3; Ju-4; Ju-5)
Los Amarillos (LA)	P.C.T	Agglomération	24
Complexe B		place	1 (LA-19)
Complexe E	D.R.II	habitation	3 (LA-1; LA-17; LA-18)
Enceinte 320		unité domestique	8 (LA-4; LA-5; LA-6; LA-7; LA-8; LA-9; LA-16; LA-20)
Unité 400	D.R.II.	unité domestique	11 (LA-2; LA-3; LA-10; LA-11; LA-12; LA-13; LA-14; LA-15; LA-22; LA-23; LA-24)
Muyuna (Mu)	P.C.T	Résidentiel	8 (Mu-1; Mu-2; Mu-3; Mu-4; Mu-5, Mu-6; Mu-7, Mu-8)
Penas Blancas (PB)		Terrasse domestique	1 (PB-1)
San Jose (SJ)	P.C.T / D.R.I	Résidentiel	1 (SJ-1)
Sarhuaico (Sa)	P.C.T	Terrasse domestique	2 (Sa-1; Sa-2)
Tilcara		Agglomération	10
Flores (F)	P.C.Ini	Sépulture isolée	2 (F-1; F-2)
Huichairas (Hui)	P.C.T	pukara	1 (Hui-1)
Pucara de Tilcara (Til)	P.C.T	pukara	7 (Til-1; Til-2; Til-3; Til-4; Til-5; Til-6; Til-7)

Les données archéologiques et anthropologiques concernant la période céramique tardive dans la Quebrada de Humahuaca sont relativement abondantes (Seldes, 2007). Il se produit à ce moment une éclosion au niveau du développement urbanistique local, avec la construction de très nombreux villages dans toutes les zones de la Quebrada, dont un grand nombre sont situés stratégiquement sur des promontoires rocheux élevés.

L'année 1250 semble marquer un tournant dans l'histoire des populations de la vallée, qui passent d'un mode de vie apparemment égalitaire pendant la période de développements régionaux I (D.R.I), entre 900 et 1250 ap. J.-C., à des sociétés hiérarchisées et inégalitaires pendant la période de développements régionaux II (D.R.II), entre 1250 et jusqu'à l'arrivée des Incas aux alentours de 1480 (Seldes, 2006). Un des arguments en faveur de la hiérarchisation de la société est entre autre la coexistence de nombreux sites de tailles très différentes, qui auraient pu avoir des fonctions distinctes (Nielsen, 2001). En effet, les

installations de cette période ont été classées en trois types différents : les zones résidentielles simples (appelées « primeros poblados » pour premiers villages), les terrasses domestiques et enfin, les agglomérations (Nielsen, 2001; Rivolta, 2007). De plus, cette phase D.R.II, correspond à une période où les sites situés dans les parties les plus basses ont été abandonnés pour peupler des **constructions défensives** ou *pukaras*, signes des tensions importantes existant dans tout le sud des Andes. Cette phase constitue le climax du développement local, interrompu par l'arrivée des Incas, dont l'influence se retrouve dans plusieurs sites de la Quebrada, le plus souvent comme des interventions sur des constructions préexistantes plus que comme des installations typiquement Incas. Lors de la colonisation de la région à partir de 1536, les Européens repèrent l'existence de deux groupes indépendants: les *Omaguacas*, comprenant les populations les plus au nord, et les *Tilcaras* regroupant les populations les plus au sud de la vallée (Figure 6, p. 59).

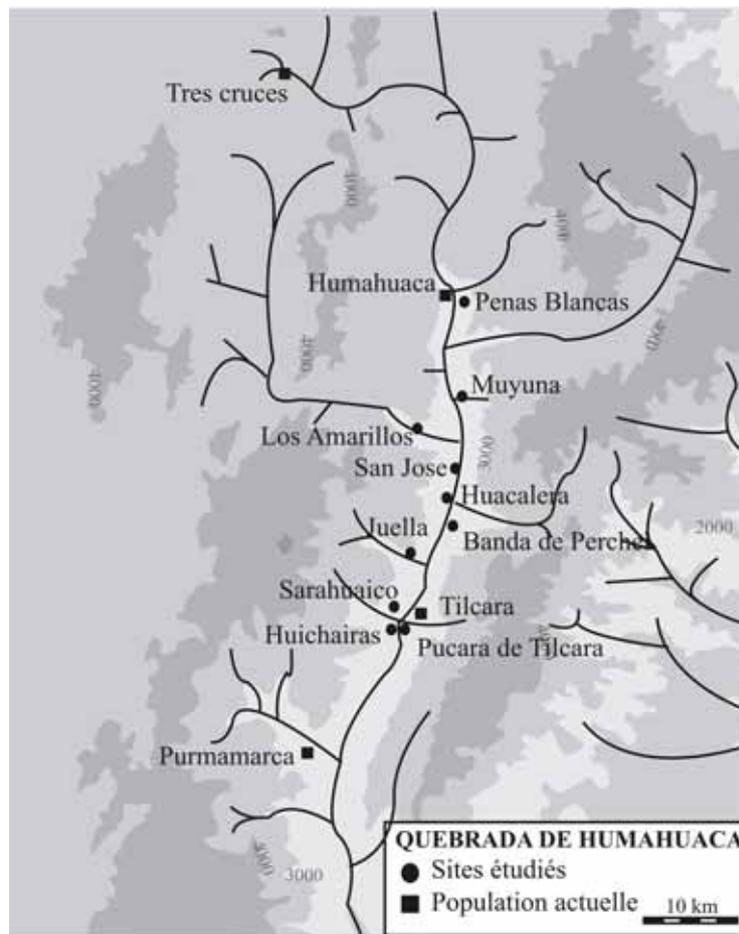


Figure 9 : Localisation des différents sites archéologiques de la Quebrada de Humahuaca étudiés.

2.1.1. Zones résidentielles simples.

Les sites inclus dans la catégorie des « premiers villages » sont le plus souvent situés dans des zones proches des berges du Rio Grande, dans la quebrada principale. L'organisation interne de ces sites comprend : des secteurs constitués d'enceintes, construites avec des murs à double paroi remplis de mortier, organisées autour de patios, des voies de circulation et des secteurs avec des fonctions spécifiques, comme des aires permettant le stockage des déchets, et parfois des enclos pour les animaux (Nielsen, 2001; Nielsen and Rivolta, 1997; Rivolta, 2007). Ces sites possèdent des superficies variant entre 2 et 4 hectares et ont été occupés quasiment exclusivement pendant la phase D.R.I de la période céramique tardive, entre **900 et 1100 ans ap. J.-C.**

Des échantillons appartenant à 10 individus provenant de trois zones résidentielles simples ont été analysés au cours de ce travail : **Huacalera, San José et Muyuna** (Tableau 4).

Les contextes archéologiques des échantillons de Huacalera et San José sont assez mal définis mais en ce qui concerne Muyuna, il s'agit d'un site résidentiel situé sur la berge gauche du Rio Grande. Le Pr. Nielsen a fouillé ce site autour des années 2000 et a récupéré un ensemble d'individus anciens exhumés de 6 structures funéraires. Il a été noté que les échantillons semblaient en mauvais état de conservation (Seldes, 2006).

2.1.2. Terrasses domestiques.

Les sites intégrés dans la catégorie « terrasses domestiques » sont des sites résidentiels qui comme leur nom l'indique sont construits sur des terrasses, avec la même technique de construction que celle utilisée pour les terrasses agricoles (Rivolta, 2007). Ces groupes se sont installés sur des versants abrupts ce qui a conduit à un profil d'établissement et d'organisation complètement différent de celui des zones résidentielles simples. Ces sites sont situés sur les berges du Rio Grande ou au bout des vallées transversales. L'organisation interne de ces sites montre une absence de planification avec l'absence d'espaces communs bien définis et aucune architecture monumentale. Chaque terrasse constitue un espace multifonctionnel, où des activités domestiques aussi bien que de production étaient réalisées (Rivolta, 2007). La grande majorité de ces types de sites ont été occupés entre **1100 et 1300 ans ap. J.-C.**

Des échantillons correspondant à 6 individus et provenant de trois sites de cette catégorie ont été analysés : **Banda de Perchel**, **Penas Blancas** et **Sarahuaico** (Tableau 4).

L'ensemble des restes humains anciens analysés ont été exhumés dans des contextes domestiques, au sein d'habitations ou de patios adjacents à ces habitations.

2.1.3. Agglomérations.

Le terme agglomération a été utilisé pour se référer au troisième type de sites, qui sont des sites de grande envergure que l'on retrouve dans différents endroits du nord-ouest argentin. Il s'agit de sites dont la superficie varie entre 7 et 10 hectares. Ils sont situés dans les zones proches du Rio Grande mais aussi dans les zones hautes des vallées transversales. L'organisation interne de ces sites est complexe, notamment avec une très grande concentration d'édifices, des voies de circulation qui desservent les secteurs résidentiels, des aires fonctionnelles spécifiques (en particulier pour l'artisanat ou pour les rituels). Ces sites ont apparemment été occupés pendant toute la période céramique tardive, mais cette occupation semble s'intensifier pendant la phase D.R.II, quand les sites résidentiels simples ont été abandonnés et que les différents groupes ont migré vers ces agglomérations.

Des échantillons correspondant à 39 individus de trois de ces agglomérations ont été analysés au cours de ce travail : **Juella**, **Los Amarillos** et **Tilcara**.

Les contextes dont sont issus les échantillons de Juella (5 échantillons) et Tilcara (10 échantillons) (Tableau 4) sont assez peu documentés car il s'agit de fouilles très anciennes. Tilcara a été parmi les premiers sites de la vallée fouillés. De plus, dans les années 30, une pyramide a été construite au centre de l'agglomération ancienne empêchant de restituer correctement le contexte général du site. Les deux échantillons de Flores, qui font partie de l'agglomération de Tilcara, sont les seuls datés de la période céramique initiale. Les deux individus ont été exhumés au cours d'une fouille de sauvetage dans l'habitation contemporaine de la famille Flores (Seldes, 2006).

Los Amarillos.

Le site archéologique de Los Amarillos est une agglomération préhispanique complexe, couvrant plus de 10 hectares, ce qui en fait le site le plus étendu de la vallée. Cinq secteurs différents de ce site ont été fouillés au cours de plusieurs campagnes de fouilles, entre les

années 1990 et 1993 puis entre les années 2000 et 2003 (Figure 10).

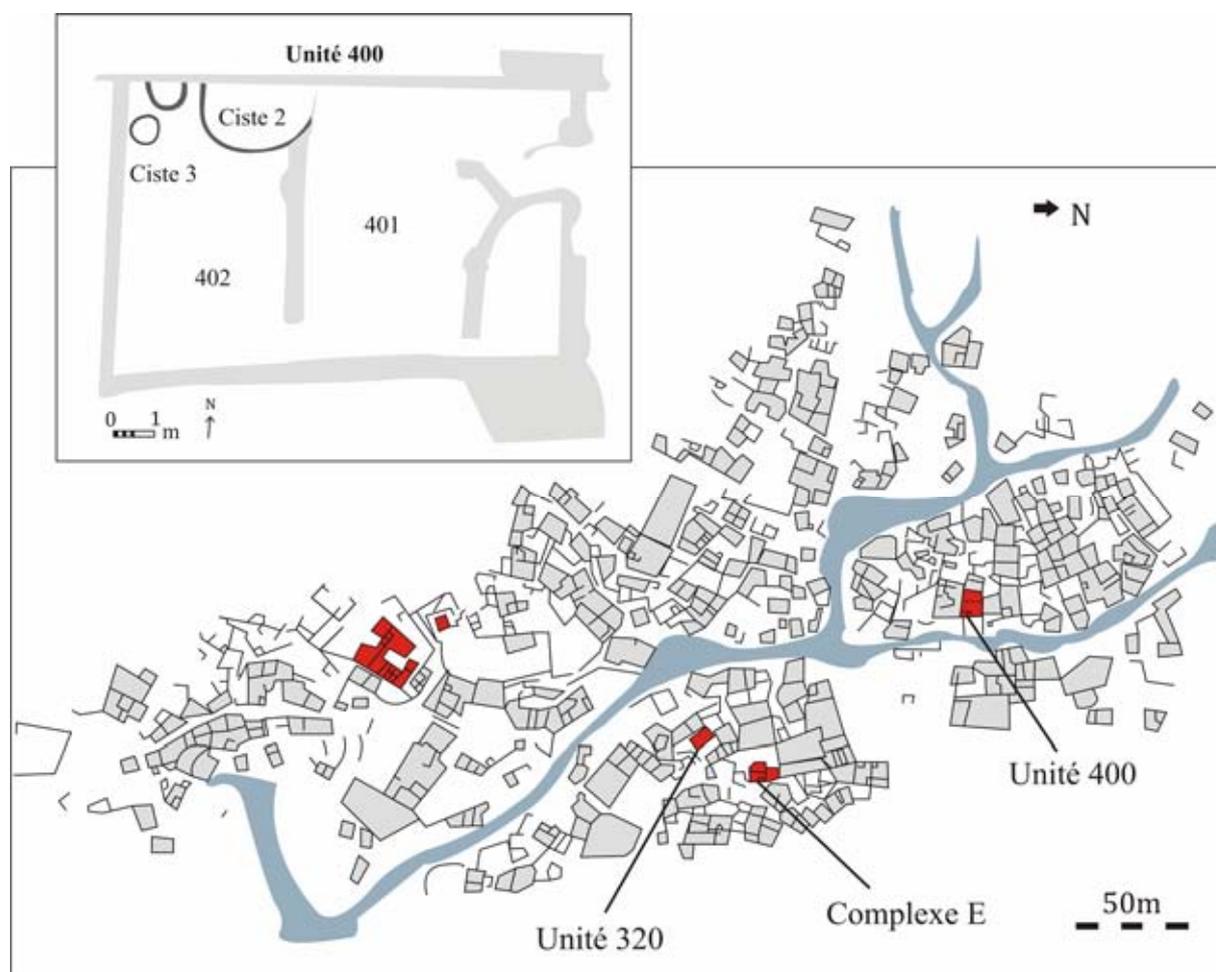


Figure 10 : Plan schématique du site archéologique de Los Amarillos et de l'unité domestique 400.

Complexe E.

Ce complexe correspond à une unité domestique composée de trois habitations accolées, situées au sein d'une aire densément construite. Les fouilles menées indiquent que ce secteur avait un caractère strictement domestique et qu'entre deux moments distincts d'occupation, ce lieu aurait été utilisé pour l'élimination des déchets (Avalos 2002). Une fois abandonné comme lieu d'habitation, différents événements d'inhumation ont eu lieu dans les enceintes d'habitation de ce complexe.

Enceinte n°301 : dans cette enceinte les fouilles ont révélé une inhumation multiple et secondaire dans une urne funéraire. Les restes retrouvés ont pu être attribués à au moins deux

adultes et cinq immatures. Dans une autre partie de l'enceinte, un individu adulte inhumé dans une urne a également été découvert ainsi qu'au moins trois immatures. L'échantillon **LA-1** correspond à un des individus inhumés au sein de cette enceinte.

Enceinte n°302 : dans cette enceinte une inhumation secondaire multiple a également été mise au jour. Cependant, dans ce cas il ne s'agit pas d'une inhumation en urne mais plutôt au sein d'une structure semi-circulaire, dans l'angle nord-ouest de l'enceinte, qui pourrait ressembler à une ciste. Au sein de cette ciste au moins deux individus adultes (**LA-17 et LA-18**) ont été inhumés. Il y avait en plus une inhumation primaire d'un immature au sein de cette structure (Avalos 2002).

Unité 400.

Cette unité est composée d'un ensemble d'enceintes fouillées par Angiorama et Taboada entre les années 1999 et 2001. A partir de leurs travaux il a été déterminé que ce secteur aurait été utilisé comme lieu domestique ainsi que comme lieu de production d'objets métalliques (Angiorama, 2005a; Taboada and Angiorama, 2003). Deux périodes d'occupation ont été détectées : une première période pendant laquelle l'unité aurait servi de lieu de résidence ainsi que d'inhumation (les cistes 1 et 3 sont contemporaines de l'occupation), et une deuxième période pendant laquelle, l'unité n'étant plus habitée, elle a été utilisée pour réaliser plusieurs événements d'inhumations (ciste 2), (Angiorama, 2005b).

Enceinte n°402, ciste 2 : Cette structure a été construite après l'abandon du lieu comme résidence et consiste en un espace semi-ovale délimité par un mur de pierre de forme courbe construit contre les murs d'enceinte nord et est (Figure 10). Dans cette structure les restes d'au moins 6 immatures et 12 adultes ont été mis au jour. Il s'agit donc d'une sépulture multiple, dans laquelle plusieurs événements d'inhumation ont été détectés (Seldes, 2006). Associés à ces restes humains, une grande quantité de matériel varié a été découvert : des pièces céramiques complètes, des pointes de projectile, des restes de vaisselles, de cordes, des artefacts de bois, des éléments de métal, comme un pendentif. Les échantillons **LA-2, LA-3, LA-10, LA-11, LA-12, LA-14, LA-15, LA-22, LA-23, LA-24**, correspondent à des individus exhumés de cette ciste.

Enceinte n°402, ciste 3: il s'agit d'une structure sous-terrainne, un trou de forme quasi-circulaire, mis au jour sous le niveau du seuil de l'enceinte, sans aucun revêtement, ni aucun type de préparation. Dans cette structure, contemporaine de l'utilisation du lieu comme lieu de résidence, les restes de deux individus adultes (**LA-13**) et deux immatures, ont été découverts.

Enceinte n°320.

Dans cette enceinte située dans le secteur est du site, une fouille de sauvetage a été réalisée. Un ensemble de restes humains sans aucune connexion anatomique a été mis au jour, et interprété comme un ossuaire. Ces restes ont été attribués à un minimum de 10 individus adultes et 6 immatures. Aucun accompagnement mortuaire n'a été retrouvé. Les échantillons **LA-4, LA-5, LA-6, LA-7, LA-8, LA-9, LA-16 et LA-20** correspondent à des individus exhumés de cette enceinte.

2.2. Vallée Calchaqui.

L'histoire de la vallée Calchaqui, située dans la province de Salta, est un peu particulière puisqu'à partir de la période céramique intermédiaire, les populations du nord et du sud vont suivre des trajectoires totalement différentes. En effet, les populations les plus au sud vont être intégrées dans le complexe culturel et politique de la Aguada, alors que l'histoire du nord de la vallée Calchaqui est similaire à celle de la Quebrada de Humahuaca. La différence principale avec la Quebrada de Humahuaca réside dans l'intensité des interactions avec les zones voisines. En effet, la Quebrada de Humahuaca a été un point central du réseau de communication de la région, alors que la vallée Calchaqui semble avoir eu une importance moindre dans le transit des ressources.

Des échantillons anciens provenant de trois sites archéologiques du nord de cette vallée, occupés pendant la période céramique tardive ou la phase Inca, ont été étudiés : **Cortaderas, Tero** (S Sal Cac 14) et **Fuerte Alto** (*cf.* n°4 sur la Figure 7 p.70).

La période céramique tardive du nord de la vallée Calchaqui présente les mêmes caractéristiques principales que pour la Quebrada de Humahuaca, c'est-à-dire, une augmentation démographique importante ainsi qu'une augmentation de la complexité sociale (De Marrais, 2001). Les sites de cette période sont beaucoup plus grands et complexes que ceux des périodes précédentes avec l'apparition d'espaces publics, de places, de plateformes, et d'autres structures non résidentielles probablement réservées à des activités cérémonielles ou des réunions publiques (De Marrais, 2001). Cette organisation interne complexe témoigne de l'apparition de différences de niveaux sociaux et d'une hiérarchie politique marquée. Comme dans la Quebrada de Humahuaca ainsi que dans tout le sud des Andes, vers la fin de

la période tardive, les sites commencent à se fortifier, très probablement pour répondre à un besoin croissant de défense.

Tero (S Sal Cac 14).

Tero est une agglomération pré-incaïque, occupée pendant la période tardive, et situé dans les environs de la ville actuelle de Cachi. L'intérêt de ce site provient du fait que des pratiques funéraires minoritaires dans la région y ont été mises au jour (Baldini, 2003; Baldini and Baffi, 2007). En effet, les immatures étaient inhumés dans des urnes, décorées ou brutes, puis recouverts de cendres chaudes (Baldini and Baffi, 2007). Les individus matures, eux, étaient inhumés dans des chambres funéraires en pierres, pouvant s'apparenter à des cistes, dont certaines étaient regroupées à l'extérieur des habitations, suggérant la présence d'une nécropole (Baldini and Baffi, 2007). Cependant, seule une petite partie des investigations menées dans ce site ont été publiées. De plus, comme illustré sur la photographie suivante (Figure 11), des habitations contemporaines ont été construites récemment sur un des secteurs du site, privant les chercheurs d'un grand nombre d'informations et entraînant des perturbations importantes. Ainsi, nous n'avons des informations que sur 19 sépultures ayant révélé 12 individus matures et 7 individus immatures.



Figure 11 : Photographie du site de Tero (S Sal Cac 14) montrant l'étendue des fouilles réalisées, ainsi que les habitations contemporaines (à gauche de la photo) construites sur un des secteurs du site.

Tiré de Soria S. S., La Gestión de recursos arqueológicos en la localidad de Cachi (Salta, Argentina), 2005.

Les échantillons **Sc-1**, **Sc-2**, **Sc-3**, **Sc-4**, **Sc-5**, **Sc-6**, **Sc-7** et **Sc-8**, correspondent à des individus exhumés de ce site.

Fuerte Alto.

Ce site ancien, occupé pendant la période céramique tardive, et situé également aux environs de la localité actuelle de Cachi, a été fouillé au tout début du XX^{ème} siècle, par Debenedetti (Baldini and Baffi, 2007). Le contexte exact des restes humains anciens mis au jour dans ce site est donc très mal connu, voir inconnu, Debenedetti ayant juste fait mention de ce site dans des notes de bas de page de ces travaux de 1908 (Baldini and Baffi, 2007).

Les échantillons notés **FA-1** et **FA-2** correspondent à des restes anciens mis au jour dans ce site.

Cortaderas.

La zone nord de la vallée Calchaqui a été une des régions clé de l'implantation et de la domination des Incas dans le nord-ouest argentin (Acuto, 2001; Acuto *et al.*, 2004). Les deux sites impériaux les plus importants y ont été installés : Potrero de Payogasta et Cortaderas.

Cortaderas était une agglomération complexe, divisée en quatre secteurs différents : les secteurs bas et gauche (*Cortaderas bajo* et *izquierdo*) sont de manufacture typiquement Inca, au sein du secteur haut (*alto*) se dresse une forteresse datant de la période tardive, abandonnée à l'arrivée des Incas et enfin le secteur droit (*derecha*) est un secteur résidentiel de manufacture locale.

L'ensemble des constructions du secteur de Cortaderas derecha, qui sont des unités résidentielles typiquement locales, sont organisées autour d'une place centrale présentant des caractéristiques de construction typiquement Inca. Ce secteur a donc été interprété comme étant un secteur résidentiel de populations locales sous domination des Incas, construit et occupé pendant la phase Inca (Acuto, 2001; Acuto *et al.*, 2004). Une structure funéraire associée à une unité domestique de ce secteur a été découverte et fouillée dans le cadre du Projet Archéologique Calchaqui, entre 1997 et 1998 (Aranda and Luna, 2005). Cette structure, qui a une forme circulaire avec un diamètre minimum de 2.10 mètres, présente des techniques de construction ainsi que des offrandes funéraires typiquement locales (type de céramique *santa-maria*; *cf.* figure 5, p.54). Trois squelettes complets, appartenant apparemment à deux individus féminins et un individu masculin, tous âgés entre 20 et 30 ans, ainsi qu'un crâne isolé (appartenant à un individu immature) ont été mis au jour dans cette

structure (Aranda and Luna, 2005).

Les échantillons **Cor-1**, **Cor-2** et **Cor-3** (fémurs) correspondent à des restes des trois squelettes complets inhumés dans cette structure funéraire.

2.3. Puna : Doncellas.

Trois échantillons provenant du site archéologique de Doncellas, dans la puna de Jujuy (*cf.* figure 7, p.70) ont été analysés. Plus exactement, les échantillons proviennent de la collection Doncellas, constituée par Casanova après les fouilles effectuées entre 1941 et 1943 (Albeck, 2001; de Micou, 2001a). La collection est formée des crânes des individus exhumés de la nécropole de ce site, qui comprend des centaines de sépulture, appelées « casas-tumbas » (maisons-tombe). Toutes les sépultures sont des structures en pierre, similaires, alignées contre une falaise rocheuse les unes à la suite des autres (Albeck, 2001). Les datations carbone 14 effectuées permettent de situer l'occupation du site pendant la période céramique tardive, entre 1 000 et 1 300 ans ap. J.-C. (de Micou, 2001a; de Micou, 2001b).

Les échantillons **Don-1**, **Don-2** et **Don-3** correspondent à trois individus issus de cette collection de Doncellas.

2.4. Sierras : Pampa Grande.

Entre 1969 et 1971, des campagnes de fouilles ont été menées dans les sierras de « Las Pirguas » dans la région de Salta en Argentine, permettant de mettre au jour une centaine de grottes ayant été occupées par les populations locales (Baffi *et al.*, 1996; Salaberry *et al.*, 1999). Six de ces cavernes ont été fouillées exhaustivement (*cf.* Figure 7, p.70), révélant des restes humains anciens correspondant à environ 120 individus, dont 15 étaient naturellement momifiés. Ces restes humains étaient associés à du matériel archéologique affilié à la tradition culturelle Candelaria, et datés par datation radiocarbone entre le V^{ème} et le VIII^{ème} siècle de notre ère (ce qui correspond bien à la phase III de la culture Candelaria, de la fin de la période céramique initiale ; *cf.* Figure 5, p.54) (Baffi *et al.*, 1996; Salaberry *et al.*, 1999; Seldes, 2006). Exceptés quelques vestiges d'occupation quotidienne, ces grottes semblent avoir été utilisées principalement comme lieu d'inhumation (Baffi *et al.*, 1996). La majorité des restes

osseux, aussi bien d'adultes que d'immatures étaient disposés dans des urnes, contenant un seul ou plusieurs individus, et une quantité importante d'offrandes funéraires (Baffi *et al.*, 1996). Quelques inhumations directes en pleine terre ont également été découvertes.

Des échantillons osseux (fémurs) appartenant à **21 individus anciens** exhumés de 5 grottes différentes ont été analysés au cours de ce travail. Les informations principales concernant ces échantillons sont présentées dans le tableau 5 suivant.

Tableau 5 : Informations principales concernant les échantillons du site de Pampa Grande étudiés.

Site	Echantillon	Code laboratoire	Sexe morphologique	Type d'inhumations
El Litro (Li)	17838	Li-1	XY	Tout type d'inhumations
	17863	Li-2	XX	
	17895	Li-3	XX	
	18432	Li-4	XY	
Caverne II (C.II)	17825	C.II-1	XY	Tout type d'inhumations
	17843	C.II-2	XX	
	17864	C.II-3	XX	
	17894	C.II-4	XX	
Caverne V (C.V)	17809	C.V-1	n.d.	Plus d'inhumations en pleine terre qu'en urne.
	17890	C.V-2	n.d.	
	17891	C.V-3	XY	
	18414	C.V-4	XX	
Lampazar IV (Lam)	17824	Lam-1	XY	Quasiment que des inhumations directes en pleine terre.
	17885	Lam-2	XY	
	18365	Lam-3	XY	
	18417	Lam-4	XY	
	18430	Lam-5	XX	
Los Aparejos (Ap)	17842	Ap-1	XX	Autant d'inhumations en urnes qu'en pleine terre
	17886	Ap-2	XY	
	18416	Ap-3	XY	
	18426	Ap-4	XY	

2.5. Désert d'Atacama : Solcor 3.

Un site du désert d'Atacama et plus particulièrement de l'oasis de San Pedro de Atacama a également été analysé : Solcor 3 (*cf.* Figure 7, p.70). Il s'agit d'une nécropole préhispanique

qui se situe dans le rayon de la ville actuelle de San Pedro de Atacama. Les fouilles menées durant l'année 1983, visaient à collecter l'ensemble du matériel pour qu'une route soit ensuite construite dans ce secteur (Bravo and Llagostera, 1986).

Le secteur fouillé a révélé un total de 153 individus distribués au sein de 93 sépultures, qui pour un grand nombre étaient signalées par des morceaux de troncs d'arbres (*chanar* et *algarrobo*) enterrés entre 0.5 et 1.5 mètres de profondeur. Dans le secteur nord de la nécropole les sépultures étaient généralement bien conservées et concernaient généralement des sépultures primaires d'adultes ou d'immatures avec parfois des sépultures primaires multiples d'adultes accompagnées d'un ou plusieurs immatures. Dans ce secteur deux groupes peuvent être différenciés : un premier dont les corps étaient enveloppés dans un tissu léger, et un deuxième composé de *fardo* funéraires, les corps étant enveloppés dans des couvertures épaisses et sécurisés par des cordes, y compris parfois avec des crochets et harnais. La disposition des corps des deux groupes est identique : c'est-à-dire des corps assis, les jambes fléchies vers le thorax, les pieds croisés l'un sur l'autre et les bras fléchis et croisés au niveau du sternum. Les offrandes funéraires étaient disposées en demi-cercle devant le corps, avec une certaine tendance à retrouver les mêmes objets aux mêmes endroits (hallucinogènes sur l'épaule gauche, arc à côté du bras droit...). Sans aucun doute ces similitudes résultent d'un **rituel funéraire bien établi** au sein de cette population.

Le second secteur fouillé présentait des sépultures beaucoup plus perturbées, avec des fosses multiples où les os étaient souvent fracturés et dispersés. Les contours de ces fosses ne sont pas clairs, il pourrait donc s'agir de fosses communes, utilisées à une période plus précoce.

Le matériel retrouvé dans certains secteurs du site, notamment les plus anciens, montre une affinité claire avec les populations de Tiwanaku.

Afin d'estimer si l'influence de la civilisation Tiwanaku a été accompagnée de mouvements de populations, 8 échantillons datés de la période pré-Tiwanaku (III^{ème} –VII^{ème} siècles) et 16 échantillons datés de la période Tiwanaku (VII^{ème} - XI^{ème} siècles) (par datation radiocarbone) ont été prélevés pour des analyses génétiques, ainsi que 29 échantillons supplémentaires encore non datés précisément.

II. Marqueurs moléculaires : à chacun son histoire.

La stratégie d'analyse mise en place pour cette étude est fondée sur l'analyse combinée de plusieurs types de polymorphismes, situés aussi bien sur le génome nucléaire que mitochondrial, afin d'obtenir des informations différentes et complémentaires permettant d'aborder au mieux les problématiques posées.

Un polymorphisme est une modification de la séquence d'ADN dont la fréquence au sein de la population étudiée est présente à plus de 1 %. Comme illustré sur la figure 12 suivante, nous nous intéresserons à différents types de polymorphismes, aussi bien des polymorphismes ponctuels de séquence que des polymorphismes de longueur.

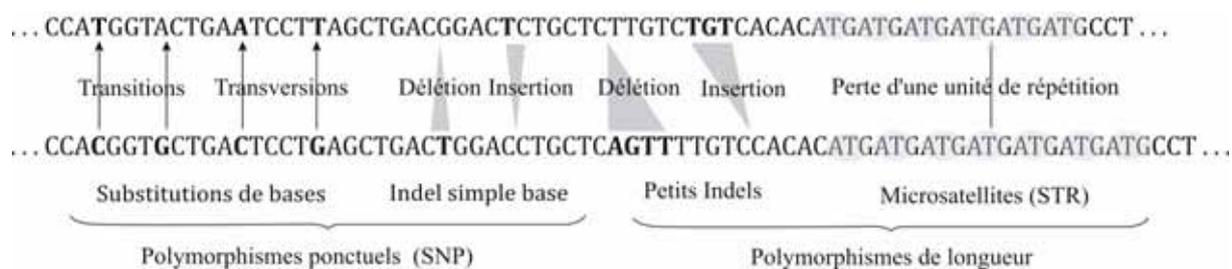


Figure 12 : Présentation des polymorphismes de l'ADN analysés dans le cadre de ce travail.

D'après (Jobling et al., 2004).

Polymorphismes ponctuels de séquence (SNP) : ce type de polymorphisme comprend aussi bien les substitutions ponctuelles (transitions ou transversions) que les insertions ou délétions d'une seule base. Ces polymorphismes sont abondants sur tout le génome humain, nucléaire et mitochondrial, aussi bien sur les régions codantes que non codantes (Jobling *et al.*, 2004). Les SNP présente le grand avantage pour l'ADN ancien de pouvoir être analysés au sein de très courts fragments de séquences, et donc de palier au problème de la fragmentation. De plus, il existe aujourd'hui de très nombreuses bases de données publiques fournissant les informations sur la localisation et les caractéristiques de tous les SNP décrits (International SNP Map Working Group, 2001). Enfin, il existe aujourd'hui de très nombreuses méthodes permettant de typer ces SNP (Gut, 2001), chacune présentant des spécificités propres

concernant le taux de multiplexage, la simplicité et la sensibilité des réactions, le taux de fiabilité, la quantité d'ADN nécessaire et le coût. Ainsi, en fonction de chaque étude, la méthode la plus adaptée devra être trouvée.

Polymorphismes de longueur : les microsatellites (STR pour Short Tandem Repeats) ainsi que les petites insertions délétions sont considérés comme étant des polymorphismes de longueur. Les STR sont présents sur les régions non codantes du génome humain, aussi bien sur les autosomes que sur les chromosomes sexuels. Ces polymorphismes correspondent à la répétition, entre 10 et 30 fois pour les marqueurs génétiques les plus souvent utilisés, d'unités de 2 à 6 paires de bases (Figure 12).

Le typage de ces polymorphismes pourra être fait au sein de courts fragments de séquence. Mais l'intérêt majeur de ces marqueurs vient de leur grande variabilité, dû au brassage génétique, qui en fait des outils de choix, notamment pour l'identification individuelle et les tests de filiation.

1. Les marqueurs à transmission biparentale.

L'ADN étant diploïde, les marqueurs analysés sur ce génome (sauf pour la région NRY) présenteront deux allèles possibles, l'un hérité du père et l'autre de la mère. Pour un locus donné, les individus seront soit homozygotes, si les deux allèles sont identiques, soit hétérozygotes, si les allèles hérités de chacun des parents sont différents.

Pour cette étude, les marqueurs à transmission bi-parentale seront essentiellement étudiés dans le but de mieux comprendre les pratiques funéraires de certaines des populations anciennes analysées.

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

La **diagnose sexuelle** constitue une étape indispensable à toute étude des pratiques funéraires. Cependant, les méthodes d'anthropologie classique ne permettent pas toujours de réaliser cette diagnose sexuelle. En effet, la méthode la plus fiable pour les individus matures étant basée sur le dimorphisme des os coxaux (Bruzek, 2002), un squelette incomplet ou des os

fragmentés empêcheront toute détermination du sexe. De plus, il n'existe pas de méthode fiable pour la détermination du sexe des immatures, privant les anthropologues d'informations pour toute une partie des populations anciennes. La détermination du **sexe moléculaire** par l'analyse d'un marqueur génétique est donc la seule méthode fiable permettant de déterminer le sexe des individus anciens, matures et immatures.

La détermination du sexe génétique, ou sexe moléculaire, se fait grâce à l'analyse du gène de l'amélogénine situé sur les chromosomes sexuels. Ce gène code pour des protéines retrouvées dans l'émail de la dent. La présence d'une délétion de 6 paires de bases dans une région du premier intron de ce gène permet de différencier l'allèle porté par le chromosome X de celui porté par le chromosome Y (Mannucci *et al.*, 1994; Sullivan *et al.*, 1993). Un individu féminin présentera un profil homozygote composé d'un produit d'amplification à la taille de 106 paires de bases alors qu'un individu masculin présentera un profil hétérozygote composé de deux produits d'amplification aux tailles de 106 et 112 paires de bases.

Aujourd'hui, de nombreux kits commerciaux d'amplification PCR permettent d'amplifier simultanément des marqueurs nucléaires ainsi que ce gène de l'amélogénine. Nous avons choisi d'adopter cette stratégie qui permet d'obtenir plusieurs informations à partir d'une seule réaction d'amplification par PCR.

1.2. Détermination des profils génétiques : analyse de marqueurs autosomaux.

Le génotypage d'un ensemble de marqueurs autosomaux (chaque locus présentant une combinaison de deux allèles) permettra de déterminer le profil génétique d'un individu. L'établissement de ce profil génétique a deux objectifs principaux très différents : (1) la comparaison deux à deux de l'ensemble des profils établis permettra d'éliminer les possibles contaminations et ainsi sera un critère supplémentaire permettant d'authentifier les résultats ; (2) les profils génétiques sont les plus informatifs pour les problématiques de détermination de liens de parenté.

L'étude des **relations de parenté** au sein d'ensembles funéraires est un aspect important pour comprendre les pratiques funéraires de populations anciennes. Cependant, les méthodes

classiques fondées sur l'étude de caractères discrets s'avèrent peu fiables. Bien souvent, les liens de parenté ne sont donc que supposés. Dans ce cadre, la paléogénétique possède un certain nombre d'outils (dont certains sont utilisés en criminalistique) permettant de déterminer les liens unissant les échantillons anciens de manière tout à fait fiable et fine. En effet, grâce à une combinaison de marqueurs autosomaux à transmission bi-parentale, permettant de déterminer les profils génétiques et de marqueurs haplotypiques à transmission uni-parentale (région hypervariable de l'ADNmt et portion non recombinante du chromosome Y) permettant d'étudier les lignées maternelles et paternelles (Figure 13), les liens de parenté (proches ou plus éloignés), mais aussi le mode de recrutement funéraire ainsi que les schémas matrimoniaux des populations anciennes pourront être abordés.

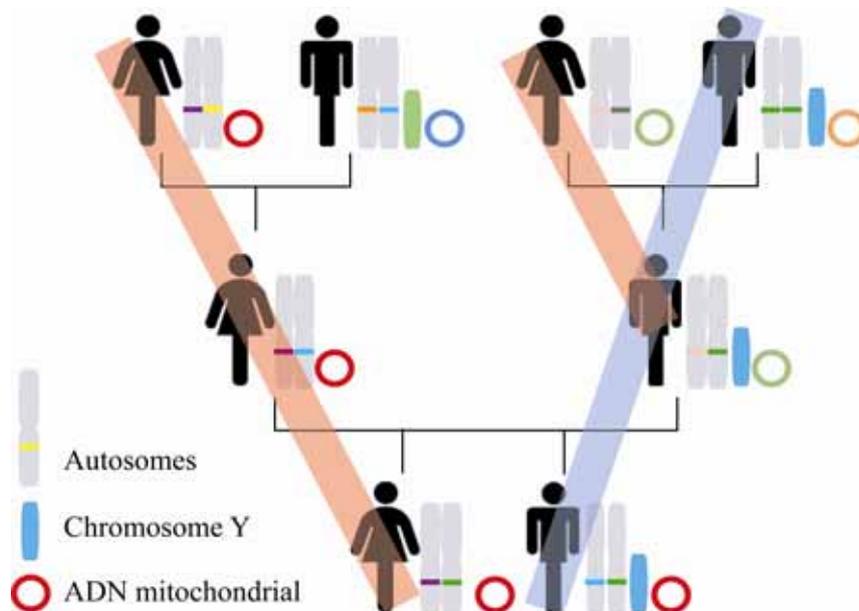


Figure 13 : Illustration de la transmission des marqueurs uni- et bi-parentaux.

Parmi les marqueurs à transmission bi-parentale, les STR autosomaux (STRa), dont le nombre de répétition est très variable en fonction des individus, sont des marqueurs de choix pour ces problématiques. La combinaison d'un certain nombre de ces marqueurs possède le pouvoir discriminant le plus élevé lorsqu'il s'agit d'étudier les liens génétiques unissant deux individus.

Aujourd'hui, il existe de très nombreux kits commerciaux, dédiés pour la plupart à des applications dans le domaine de la criminalistique, qui permettent l'amplification simultanée d'un certain nombre de STRa (entre 9 et 16 suivant les kits utilisés). De plus, il existe de nombreuses bases de données de références publiques compilant les fréquences alléliques des marqueurs analysés dans ces kits, dans différentes populations contemporaines, ce qui permet par la suite de réaliser des comparaisons entre populations anciennes et contemporaines pour étudier l'évolution des populations, comme celle des marqueurs. Nous avons donc choisi d'adapter ces kits à nos applications, puisqu'ils présentent plusieurs avantages essentiels pour les études d'ADNa.

Même si les STR autosomaux apparaissent aujourd'hui comme des marqueurs indispensables pour l'établissement des profils génétiques, ils peuvent présenter quelques inconvénients, notamment lorsque l'on travaille à partir d'extraits d'ADN très fortement dégradé. En effet, certains des marqueurs analysés sont amplifiés au sein de fragments relativement grands (la taille des fragments pouvant aller jusqu'à 400 paires de bases) et peuvent donc s'avérer difficilement amplifiables. Pour certains échantillons anciens fortement dégradés et fragmentés, les profils obtenus pourront être partiels, conduisant à une perte d'information. Pour palier à ce problème, l'utilisation de miniSTR est préconisée. Ces miniSTR sont amplifiables au sein de fragments plus courts et peuvent donc être intéressants pour compléter les profils obtenus avec les STRa.

Ces dernières années, d'autres marqueurs ont été utilisés pour répondre à ces problématiques d'identification individuelle et de parenté. Ainsi, l'intérêt pour les SNP ou les petites insertions délétions (indels) particulièrement adaptés à l'analyse de l'ADN dégradé car analysables au sein de fragments très courts, est grandissant. Si les SNP présentent des caractéristiques très intéressantes pour l'analyse d'ADN dégradé il apparaît que le nombre de marqueurs permettant d'atteindre un degré suffisant de discrimination est relativement élevé (au minimum 52 pour égaler le niveau de discrimination des STR) (Sanchez *et al.*, 2006).

Dans le but de trouver toujours plus de marqueurs intéressants pour les applications d'ADNa, nous avons également testé l'intérêt de l'utilisation d'indels pour l'identification individuelle d'individus humains anciens, et éventuellement pour la détermination de liens de parenté. Ainsi, nous avons utilisé un kit d'amplification nouvellement commercialisé, permettant d'amplifier simultanément 30 indels, pour analyser nos échantillons anciens.

2. Les marqueurs à transmission uni-parentale.

La mitochondrie, qui est un organite cellulaire extranucléaire, possède son propre ADN sous la forme d'une molécule circulaire de 16 569 paires de bases de longueur, qui a déjà été entièrement séquencé (Anderson *et al.*, 1981). Ce génome mitochondrial est transmis uniquement de manière maternelle, et contrairement à l'ADNnu, il ne recombine pas et il est présent en un grand nombre de copies par cellules (plusieurs centaines à plusieurs milliers par cellules). Toutes les variations de séquence observées au niveau de cet ADNmt sont donc le résultat de mutations accumulées et transmises des mères à leur fille (Pakendorf and Stoneking, 2005). Le taux de mutation tout au long de la molécule n'est pas uniforme, avec un taux relativement faible pour la région codante et plus élevé pour les régions non-codantes, dont notamment sur les régions hypervariables de la D-loop (HVR-1, HVR-2).

La portion male-spécifique du chromosome Y (NRY) ne possède pas de contrepartie homologue, n'est pas soumise à la recombinaison et est transmise directement des pères à leur fils. En ce sens, il s'agit en quelque sorte de l'équivalent masculin de l'ADNmt.

Ces deux systèmes génétiques haploïdes permettront de retracer d'une part l'histoire des individus féminins et d'autre part l'histoire des individus masculins, qui sont complémentaires mais peuvent parfois être différentes.

Ainsi, la combinaison de l'état allélique de marqueurs génétiques le long de ces régions du génome permettra de définir des **haplotypes**. En ce qui concerne l'ADNmt, ces haplotypes sont la plupart du temps définis à partir des mutations détectées grâce au séquençage de la première région hypervariable (HVR-1), par comparaison à une séquence de référence [séquence référence de Cambridge révisée (rCRS), (Andrews *et al.*, 1999)]. Les haplotypes du chromosome Y seront déterminés grâce à une combinaison de STR, situés sur la portion non recombinante.

L'histoire de populations anciennes, ou encore la dynamique de peuplement de certaines régions pourront être révélés grâce à l'analyse de ces haplotypes. Les haplotypes partagés entre différents groupes anciens permettront de mettre en évidence des degrés d'affinités et éventuellement des flux génétiques évocateurs de la diffusion ou de la migration de populations. L'évolution des populations (continuité ou discontinuité génétique) pourra être révélée grâce à l'analyse des fréquences des haplotypes au cours du temps.

Pour comprendre des processus évolutifs à plus grande échelle, l'analyse des haplotypes peut s'avérer insuffisante. Dans ces situations il est nécessaire de s'intéresser aux lignées, ou **haplogroupes**. Un haplogroupe peut être défini comme un ensemble de séquences (haplotypes) dérivées d'un même ancêtre commun, informatif de **l'origine biogéographique** d'un individu. Ainsi, ces lignées, aussi bien maternelles que paternelles, seront utiles pour retracer des évènements de migration à grande échelle (comme la colonisation de l'Amérique), permettant de révéler l'origine mais aussi le moment de ces évènements. L'étude de ces lignées permettra également de mettre en évidence des degrés de métissage dans certaines populations, tout comme certains processus évolutifs tels qu'un goulot d'étranglement ou une dérive génétique.

Même si les haplogroupes peuvent être déduits à partir des haplotypes, la détermination de ces lignées est faite grâce à l'analyse de polymorphismes ponctuels situés pour la plupart sur la partie codante de l'ADNmt et sur la région non recombinante du chromosome Y.

III. Préparation et extraction des échantillons d'ADN.

1. Pré-traitement des échantillons.

Avant de procéder à l'extraction de l'ADN des échantillons anciens, ils subissent un traitement préalable dans le but, entre autre, d'éliminer les possibles contaminations de surface. Ce traitement est spécifique à chaque type d'échantillon.

Les dents ont en premier lieu été nettoyées, une par une, et les éventuels résidus (terre, os, colle...) ont été retirés à l'aide d'un scalpel stérile. Ensuite, les dents ont été rincées à l'eau ultra-pure (milli-Q, Millipore) à l'aide d'une compresse stérile, « frottées » à la javel, et enfin à nouveau rincées à l'eau ultra-pure (milli-Q, Millipore). Pour finir, les échantillons disposés dans des grips stérilisés ont été irradiés sous les UV pendant environ 15 minutes de chaque côté. Une fois ces échantillons anciens décontaminés, ils ont été réduits en poudre à l'aide d'un cryobroyeur à azote liquide [6870 SamplePrep Freezer Mill[®] (Fisher Bioblock)], ce qui permet d'éviter les échauffements pouvant induire une dégradation de l'ADN. La poudre de dent ainsi obtenue a été disposée dans un tube 50 mL individuel pour chaque échantillon, ceux-ci ayant été conservés à l'abri de la lumière dans un endroit où seul des échantillons anciens étaient conservés.

Comme pour les dents, les échantillons osseux anciens ont tout d'abord été nettoyés, et les résidus éventuels (terre) éliminés à l'aide d'un scalpel stérile. Par la suite, la surface des fragments d'os longs (pour la plupart des fémurs) a été abrasée mécaniquement sur une épaisseur d'environ 2 mm, à l'aide d'une perceuse de type Dremel[®] (Breda), sous une hotte consacrée à cet usage. Enfin, ces fragments d'os longs abrasés ont été trépanés grâce à une perceuse à colonne équipée d'un trépan chirurgical afin d'obtenir de la « poudre » d'os. Cette poudre a également été disposée dans un tube de 50 mL individuel, puis conservée dans un endroit dédié à la conservation des échantillons anciens.

2. Extraction de l'ADN.

2.1. ADN ancien.

L'extraction est l'une des étapes les plus importantes du processus d'analyse de l'ADN. Depuis la publication des premières études, les protocoles d'extraction ont considérablement évolué et sont encore continuellement améliorés (Campos *et al.*, 2011; Rohland *et al.*, 2010). Au cours de ce travail, différentes techniques d'extraction ont été testées puis la technique présentant le meilleur rendement a été conservée pour le traitement de l'ensemble des échantillons.

La première des deux méthodes testées est fondée sur l'utilisation du kit PrepFiler™ Forensic DNA Extraction Kit (Applied Biosystems). Ce kit est présenté comme permettant d'améliorer la quantité et la qualité de l'ADN extrait à partir d'échantillons d'ADN dégradé et présent en faible quantité, ce qui augmenterait le rendement des analyses réalisées en aval de l'extraction (Brevnov *et al.*, 2009). Le kit PrepFiler™ utilise une combinaison de billes magnétiques et une chimie optimisée permettant de maximiser les performances de chaque étape de l'extraction : (1) la forte capacité de fixation des billes permet de récupérer un maximum d'ADN, (2) la fixation est très stable ce qui permet une étape de lavage efficace et donc de maximiser l'élimination des inhibiteurs de la PCR, tout en minimisant la perte de l'ADN, (3) le rendement de la quantité obtenue après élution est stable et maximal.

La deuxième méthode d'extraction testée est fondée sur l'utilisation d'un autre kit, le kit NucleoSpin® Extract II (Macherey-Nagel). Le principe d'extraction de ce kit est fondé sur l'utilisation d'une colonne avec membrane de silice. Le kit NucleoSpin® Extract II est conçu pour la purification directe de produits PCR et pour la purification de l'ADN sur gels d'agarose. Ce kit, conçu pour éliminer les traces d'amorces inutilisées, et dans le même temps pour purifier les produits de PCR jusqu'à 65 pb, garantit la récupération de tous petits fragments d'ADN purifiés avec un rendement élevé. Même si ce kit n'est pas directement conçu pour l'extraction d'ADN à partir d'échantillons anciens osseux, nous avons choisi d'adapter le protocole à nos applications, car il présente des caractéristiques intéressantes.

Les différentes étapes d'extraction de ces deux kits sont présentées sur la figure 14 suivante (p.95).

Les tests réalisés avec ces deux kits ont montré que le rendement était nettement meilleur avec l'utilisation du kit NucleoSpin[®] Extract II. Ce protocole a donc été utilisé pour l'extraction de l'ensemble des échantillons, suivi par une étape finale de concentration de l'ADN grâce à l'utilisation d'amicons (Y M-30 ; Sigma). Cette étape de concentration permet d'éluer l'ADN dans le volume le mieux adapté, et d'obtenir une concentration optimale pour réaliser différentes analyses par la suite.

2.2. ADN moderne.

Certains échantillons contemporains correspondant à ceux des manipulateurs et fouilleurs ont été analysés. Ces ADN modernes ont été extraits à partir de cellules buccales (cytobrosses). Le kit d'extraction QIAamp[®] DNA Mini kit (Qiagen) a été utilisé pour réaliser ces extractions, selon le protocole décrit par le fabricant.

Avant la réalisation de ces prélèvements, le consentement éclairé de tous les individus a été obtenu par écrit, conformément à la législation française. Ces échantillons n'ont en aucun cas été utilisés pour des recherches mais seulement pour la comparaison des profils génétiques avec les profils des échantillons anciens, afin de vérifier des éventuelles contaminations pas les manipulateurs ou fouilleurs.

3. Quantification de l'ADN extrait.

L'ADN ancien extrait a été quantifié grâce à une méthode de PCR en temps réel utilisant le kit Quantifiler[™] Human Identification (Applied Biosystems) avec le système de détection ABI PRISM[®] 7000 (Applied Biosystems), selon les recommandations du fabricant.

Ce kit est destiné à estimer la quantité totale des molécules d'ADN humaine amplifiables dans un échantillon. Les résultats obtenus permettent d'obtenir plusieurs informations : (1) savoir si la quantité d'ADN humain détectée est suffisante pour procéder à des analyses génétiques telles que le séquençage ou le typage de SNP et/ou STR, (2) permet de savoir la quantité d'extrait à utiliser pour optimiser les réactions effectuée, (3) permet de détecter les potentiels inhibiteurs, et ainsi, d'adapter les protocoles d'analyse en conséquence.

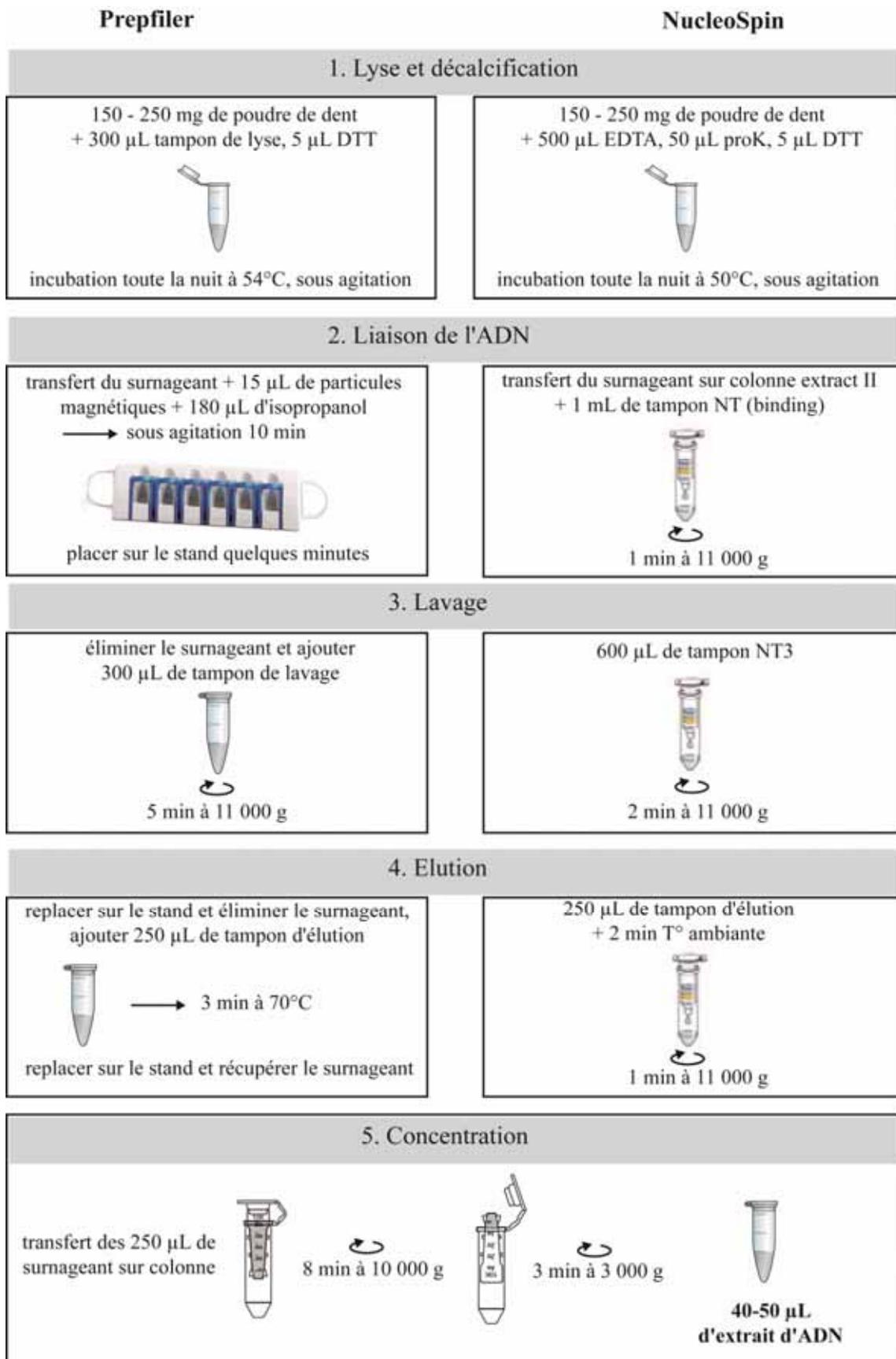


Figure 14 : Présentation des différentes étapes des deux protocoles d'extraction testés au cours de ce travail.

IV. Analyses génétiques menées.

1. Première région hypervariable de l'ADNmt (HVR-1).

1.1. Amplification par PCR.

L'analyse de 359 paires de bases de la première région hypervariable de l'ADNmt (16024-16383) a été faite par l'amplification de deux amplicons se chevauchant, d'une part pour que la longueur des séquences soit compatible avec l'état dégradé de l'ADN analysé, et d'autre part, pour obtenir une zone contrôle au niveau du chevauchement. L'amplification de ces deux fragments a été faite en utilisant deux couples d'amorces, L16589 (Ivanov *et al.*, 1996) et H16239 (Gabriel *et al.*, 2001) pour amplifier la région HVR-1a, et L16190 / H16410 (Gabriel *et al.*, 2001) permettant d'amplifier la région HVR-1b. Dans les cas où une mutation en position 16189 (T→C) entraînait des problèmes d'amplifications, liés à la région polycytosine localisée entre les positions 16184 et 16193 (Szibor and Michael, 1999), empêchant une élongation correcte de l'amorce H16239, une amorce reverse différente a été utilisée : H16167 (Ricaud *et al.*, 2004).

Les réactions d'amplifications ont été effectuées dans un volume réactionnel final de 50µL [10 µL d'ADN, 1 µL de chacune des amorces, 10 µL de mix 5X (5 µL de buffer 10X (Eurogentec), 3 µL de MgCl₂ à 25mM, 0,5 µL de BSA (Fermentas), 1,1 µL d'H₂O et 0,1 µL de chaque dNTP), 0,5 µL de Taq Hot Start (Eurogentec), le volume final étant complété par 27,5 µL d'eau ultra-pure (milli-Q, Millipore)], dans un thermocycleur (T3 ; Biometra). Les programmes PCR suivant ont été utilisés : 13 min d'activation à 94°C, suivies de 38 cycles de 30 sec de dénaturation à 94°C / 30 sec d'hybridation à 48°C pour le fragment HV1a et 51°C pour le fragment HV1b / 45 sec d'élongation à 72°C, avec une phase finale d'élongation de 5 min à 72°C.

1.2. Purification des produits PCR.

La qualité (abondance et spécificité) des produits amplifiés a été contrôlée grâce à la

migration de 10 μL de produit PCR [additionné de 2 μL de bleu de charge (loading dye 6x ; Eurogentec)] sur un gel d'agarose à 1,5 % (1,5 grammes de poudre d'agarose pour 100 mL de tampon TBE 10X et quelques gouttes de BET). Les résultats de cette migration (à 90 V pendant 30 min dans du tampon TBE 1X) sont observés par fluorescence sur une table à UV, puis comparés à un marqueur de taille de référence (DNA ladder, Eurogentec) ce qui permet de vérifier que les produits amplifiés sont à la taille attendue mais également d'estimer la concentration du produit amplifié. Les témoins négatifs et blancs d'extractions sont également systématiquement déposés pour vérifier qu'aucune contamination n'ait eu lieu pendant les étapes d'extraction et d'amplification.

Afin d'éliminer les résidus autres que l'ADN (sels, dNTPs et amorces non incorporés), et pour que tous les produits soient à une concentration homogène, optimale pour les analyses suivantes, chaque produit d'amplification est ensuite purifié à l'aide du kit NucleoSpin[®] extract II (Macherey Nagel). En bref, les 40 μL de produits PCR restants sont mélangés à 100 μL de tampon NT au 1/5, disposés sur une colonne à membrane de silice et centrifugés 1 min à 11 000 g. L'éluât est ensuite éliminé, puis 600 μL de tampon de lavage NT3 sont rajoutés avant une nouvelle centrifugation 2 min à 11 000 g. Enfin, en fonction de la concentration estimée sur gel, le volume adéquat de tampon d'élution NE est ajouté, et après un temps de repos de 1 min à température ambiante, une dernière centrifugation de 1 min à 11 000 g permet de récupérer le produit PCR purifié.

1.3. Séquençage et analyse.

Le kit BigDye[®] Terminator v1.1 Cycle Sequencing (Applied Biosystems) basé sur la méthode appelée "dye terminator sequencing" (Smith *et al.*, 1986), a été utilisé pour amplifier chaque fragment indépendamment. Le mélange réactionnel est constitué de : 2 μL de produits PCR purifiés, 2 μL de BigDye, 1 μL de tampon, 3 μL d' H_2O et 2 μL d'amorce spécifique à 0,8 μM , dans un volume final de 10 μL . les paramètres d'amplification utilisés sont identiques pour tous les fragments : 10 sec de dénaturation à 96°C, suivis de 25 cycles comprenant chacun 10 sec de dénaturation à 96°C, 5 sec d'hybridation à 50°C et 4 min d'élongation à 60°C.

Une précipitation à l'éthanol est ensuite effectuée : les 10 μL de produits sont laissés 15 minutes au congélateur dans 40 μL d'éthanol à 80 %. Après une centrifugation à 12 000 rpm

pendant 30 min, le surnageant est éliminé et le culot re-suspendu et rincé (élimination des amorces et dNTPs résiduels) avec 200 μL d'éthanol à 70%. Après une nouvelle centrifugation de 5 min à 12 000 rpm, le surnageant est à nouveau éliminé.

Les culots sont séchés pendant au minimum 4 heures, puis re-suspendus dans 10 μL d' H_2O . 4 μL de ces produits de séquence seront utilisés pour une électrophorèse sur un séquenceur automatique ABI PRISM[®] 3500. Les fragments seront analysés et une séquence consensus sera obtenue à l'aide du logiciel Sequencher[™] v.4.8 (Gene Codes Corporation).

2. Analyse des STR du chromosome Y.

Le kit AmpFISTR[®] Yfiler[™] (Applied Biosystems) a été utilisé pour analyser les STR du chromosome Y de l'ensemble des individus humains anciens identifiés comme étant masculins, grâce aux analyses des STRa. Ce kit permet l'amplification des loci de l'haplotype européen minimal (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393), ainsi que d'autres loci très polymorphes (DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 et GATA Y H4).

Le protocole fourni par le fabricant a été modifié pour s'adapter au mieux aux spécificités des échantillons anciens analysés. Ainsi, les PCR ont été réalisées dans un volume final de 12,5 μL [5 μL d'extrait d'ADN, 0,4 μL d'AmpliTaq Gold[®] DNA Polymerase (Applied Biosystems), 4,6 μL de Reaction Mix et 2,5 μL de Primer Set]. Les paramètres de PCR recommandés par le fabricant ont également été modifiés, le nombre de cycle PCR ayant été augmenté de 30 à 34 cycles : 11 min d'activation de l'enzyme à 95°C, suivis de 34 cycles comprenant chacun 1 min de dénaturation à 94°C, 1 min d'hybridation à 61°C et 1 min d'élongation à 72°C, suivis d'une élongation finale de 80 min à 60°C.

Les produits amplifiés ainsi obtenus sont ensuite analysés par électrophorèse capillaire sur un séquenceur automatique ABI PRISM 3500 (Applied Biosystems). Pour cela, un mélange réactionnel comprenant 1 μL de produit amplifié, 8,5 μL de Hi-Di Formamide (Sigma) et 0,5 μL de GeneScan[™] – 600 LIZ[®] Size Standard (Applied Biosystems), a été dénaturé 3 min à 95°C avant d'être analysé. Les résultats obtenus ont été analysés avec le logiciel GeneMapper v.4.1 (Applied Biosystems).

3. Analyse des marqueurs autosomaux.

3.1. STR.

Différents kits commerciaux ont été utilisés pour l'analyse des échantillons anciens et contemporains. Pour les échantillons anciens deux kits ont été utilisés : le kit AmpFISTR® Identifiler® Plus (Applied Biosystems) qui permet l'amplification simultanée de 15 STR autosomaux ainsi que le marqueur de l'amélogénine et le kit AmpFISTR® MiniFiler™ (Applied Biosystems) qui permet l'amplification de 9 STR autosomaux ainsi également que le gène de l'amélogénine. Le kit Identifiler® Plus permet l'amplification des treize loci inclus dans le système Combined DNA Index (CODIS) ainsi que deux autres loci, D2S1338 et D19S433.

Les marqueurs amplifiés grâce à ces deux kits sont identiques, mais la taille des amplicons au sein desquels ils sont analysés est différente. En effet, plus la taille des amplicons est réduite plus il y a de chances que ces marqueurs soient amplifiés correctement. La figure 15 suivante illustre la taille des amplicons ainsi que les marqueurs analysés grâce aux différents kits utilisés.

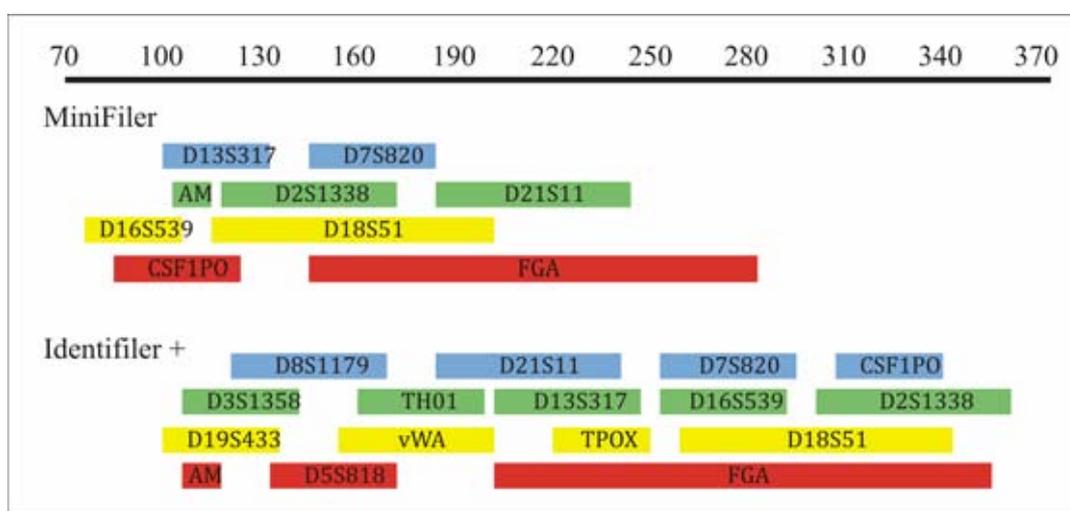


Figure 15 : Illustration représentant la taille des amplicons des différents marqueurs amplifiés grâce aux différents kits utilisés.

Les amplifications ont été réalisées, pour les deux kits, dans un volume final de 12,5 µL [5 µL d'extrait d'ADN, 5 µL de Reaction Mix et 2,5 µL de Primer Mix], avec les paramètres

suivants : 11 min d'activation de l'enzyme à 95°C, suivis pour MiniFiler™ de 30 cycles comprenant 20 sec de dénaturation à 94°C, 2 min d'hybridation à 59°C et 1 min d'élongation à 72°C, suivis d'une élongation finale de 45 min à 60°C, et pour IdentiFiler® Plus de 32 cycles comprenant 20 sec de dénaturation à 94°C, 3 min d'hybridation et d'élongation à 59°C, suivis d'une élongation finale de 10 min à 60°C.

Les produits obtenus ont également été analysés par électrophorèse capillaire, en suivant le même protocole que celui détaillé pour les STR du chromosome Y (IV.2).

Les échantillons du site de Pampa Grande ayant été analysés en partie avant le début de ce travail de thèse, les profils génétiques (réalisés avec le kit Profiler plus permettant l'amplification de 9 marqueurs) avaient été obtenus à l'aide d'autres kits, comme cela est présenté dans l'article Carnese *et al* (2010) (p.247).

3.2. Petites insertions délétions (Indels).

Dans le but d'optimiser au maximum notre stratégie d'analyse de l'ADNa, d'autres marqueurs que les marqueurs classiquement analysés ont été testés au cours de ce travail de thèse. Depuis quelques années l'intérêt des indels est démontré par plusieurs études (da Costa Francez *et al.*, 2011; Pereira *et al.*, 2007). Le caractère bi-allélique de ces marqueurs et la possibilité de les analyser au sein de courts amplicons, entre autres, en font des outils intéressants pour l'analyse d'ADN dégradé.

Le kit Mentype® DIPplex (Qiagen), récemment commercialisé, permet l'analyse de 30 indels ainsi que du gène de l'amélogénine. Les amplicons des loci analysés ne sont pas plus grand que 150 paires de bases, et le pouvoir de discrimination de ce kit est relativement élevé ($2,83 \times 10^{-13}$). Nous avons donc testé ce kit, sur certains des échantillons anciens, afin d'évaluer leur intérêt pour les applications d'anthropologie moléculaire et notamment pour l'identification individuelle et la détermination de liens de parenté.

L'amplification se fait à partir de 5 µL d'ADN dans un volume réactionnel final de 13 µL [2,5 µL de Reaction Mix, 2,5 µL de Primer Mix, 0,3 µL de Multi Taq2 DNA polymérase, et H2O Qsp 8 µL], selon les paramètres d'amplification suivants : 4 min d'activation de l'enzyme à 94°C, suivis de 32 cycles composés de 30 sec de dénaturation à 94°C, 2 min d'hybridation à 61°C et 75 sec d'élongation à 72°C, suivis d'une élongation finale de 60 min à 68°C.

Les produits d'amplification [1 μ L de produit PCR additionné de 12 μ L de Hi-Di Formamide (Sigma) et de 0,5 μ L de BTO DNA size standard (Qiagen)] ont été dénaturés 3 min à 95°C avant d'être analysés par électrophorèse capillaire sur un séquenceur automatique ABI PRISM 3500 (Applied Biosystems). Les résultats ont été visualisés et analysés sur le logiciel GeneMapper v.4.1 (Applied Biosystems).

4. Analyse de SNP par spectrométrie de masse.

Les techniques de génotypage de SNP sont toutes fondées sur deux étapes successives, (1) la discrimination allélique, permettant de générer les produits pour chaque allèle recherché, puis (2) la détection de ces allèles. Pour chacune de ces deux étapes il existe un grand nombre de méthodes variées présentant toutes des caractéristiques différentes. Le choix d'une méthode repose donc sur différents critères. En ce qui concerne les applications pour cette étude, l'objectif était de trouver une technique sensible, permettant un haut degré de multiplexage, un faible coût relatif, et permettant de travailler à partir d'ADN en faible quantité et dégradé. Ainsi, le choix du laboratoire s'est porté sur la spectrométrie de masse et plus particulièrement sur la plateforme MassArray[®] de la compagnie Sequenom. Cette plateforme de spectrométrie de masse est la seule spécialement conçue pour des analyses d'acides nucléiques. De plus, la spectrométrie est la seule méthode fiable de détection des produits alléliques puisqu'elle mesure une propriété intrinsèque des produits à analyser (masse/charge), sans qu'un marquage (fluorochrome ou autre) ne soit nécessaire. Enfin, la technologie de génotypage de SNP, iPLEX[®] Gold, proposée sur cette plateforme est présentée comme fiable, sensible, rapide et permettant un degré de multiplexage important ce qui était intéressant pour nos applications (Bray *et al.*, 2001; Cerezo *et al.*, 2009; Gabriel *et al.*, 2009).

4.1. Principe général et protocole.

Le principe de la plateforme de génotypage de Sequenom est fondé sur une extension simple base de la région d'intérêt puis sur une discrimination des produits alléliques par spectrométrie de masse MALDI-TOF (Matrix Assisted Laser Desorption Ionisation - Time of Flight). Ainsi, comme illustré sur la figure 16, la première étape consiste à l'amplification par PCR des courts fragments d'ADN contenant les SNP d'intérêt. Ensuite, les dNTP non

incorporés sont neutralisés par une solution contenant de la SAP (Shrimp Alkaline Phosphatase). Le mélange réactionnel iPLEX[®] Gold est ajouté aux produits amplifiés à l'étape suivante, ce qui conduit à l'ajout d'un nucléotide de masse modifiée au niveau du site d'intérêt, grâce à une amorce SBE s'hybridant juste en amont de ce site. Avant, l'analyse par spectrométrie de masse, les produits sont dessalés et transférés sur une puce (Spectro Chip, Sequenom) grâce au nanodispenser. Pour finir, les résultats sont obtenus directement sous la forme de spectres.

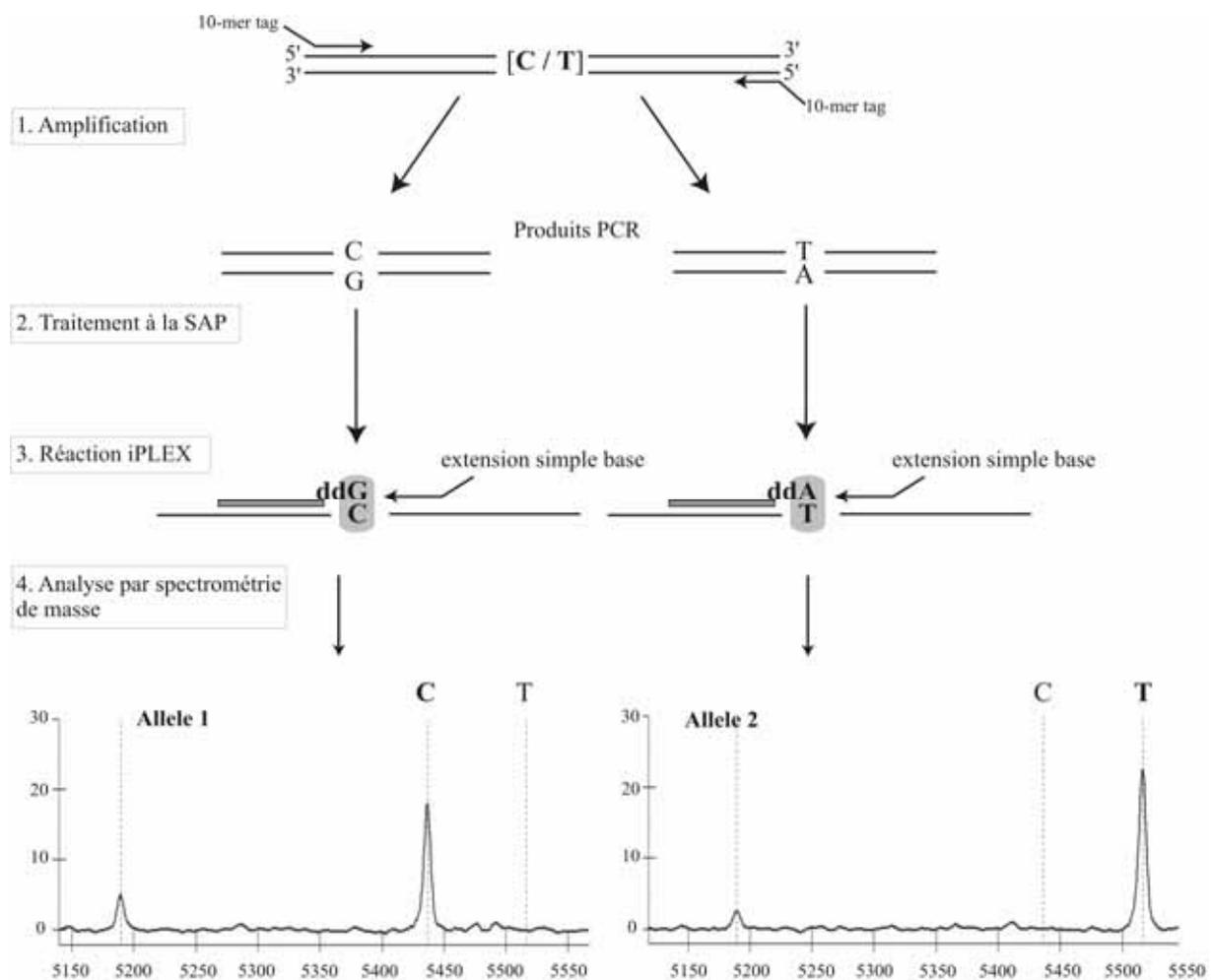


Figure 16 : Principe général de la technologie de génotypage iPLEX Gold[®] de Sequenom.

Le protocole proposé par le fabricant est optimisé pour des quantités d'ADN comprises entre 5 et 10 ng. Dans un premier temps, nous avons donc testé la sensibilité de cette méthode pour connaître sa fiabilité avec des échantillons biologiques dont l'ADN est dégradé et présent en faible quantité. Nous avons donc réalisé une gamme de dilution à partir d'ADN moderne

quantifié (entre 10 ng et 50 pg d'ADN) et testé ces ADN avec une réaction comprenant 17 SNP situés sur le génome nucléaire. Comme décrit dans l'article Mendisco *et al.*, (2011), le protocole iPLEX Gold s'avère effectivement fiable et sensible pour l'analyse d'échantillons anciens, avec cependant quelques modifications du protocole proposé par le fabricant. En effet, pour obtenir une quantité suffisante d'ADN par réaction, le volume d'extrait d'ADN par réaction a été augmenté de 1,4 μ L à 3,6 μ L.

4.2. Choix des SNP d'intérêt.

Dans un premier temps, 25 marqueurs, situés sur le génome mitochondrial ainsi que sur le chromosome Y, caractéristiques des haplogroupes majeurs Amérindiens, et permettant soit de confirmer les haplogroupes déduits des haplotypes, soit d'affiner la détermination de ces haplogroupes, ont été sélectionnés dans la littérature. La lignée maternelle X2a n'étant présente a priori que dans les populations amérindiennes d'Amérique du Nord (Dornelles *et al.*, 2005), seules les lignées maternelles : A2, B2, C1 et D1 ont été intégrées à la réaction (Bandelt *et al.*, 2003; O'Rourke and Raff, 2010; Tamm *et al.*, 2007; van Oven and Kayser, 2009). En ce qui concerne les lignées paternelles, l'ensemble des haplogroupes et sous-haplogroupes décrits à ce jour en Amérique du Sud ont été intégrés à l'analyse (Bisso-Machado *et al.*, 2011; Bortolini *et al.*, 2003; Karafet *et al.*, 2008; Zegura *et al.*, 2004).

L'objectif lorsque l'on travaille à partir d'ADN en faible quantité et dégradé est de pouvoir obtenir un maximum d'informations à partir d'une faible quantité d'extrait d'ADN, ainsi, le but était de pouvoir typer l'ensemble des marqueurs sélectionnés en une seule réaction. La conception des amorces et des réactions a été réalisée à l'aide du logiciel associé à la plateforme de génotypage MassArray[®] : Assay designer 4.1. Ce logiciel nous a permis dans un premier temps de créer deux réactions de 11 SNP chacune (Mendisco *et al.*, 2011), rejetant donc trois SNP qui ne pouvaient être typés par cette méthode.

Par la suite, une nouvelle lignée paternelle (Q1a3a4) ayant été décrite comme étant typique des populations andines (Jota *et al.*, 2011), nous avons voulu l'intégrer à nos analyses. Parmi les deux réactions précédemment créées, nous avons intégré certains marqueurs qui caractérisent la même lignée, dans le but d'optimiser les chances d'obtenir un résultat exploitable. A la vue des premiers tests réalisés, nous avons donc réorganisé les réactions manuellement, en ne gardant que les marqueurs indispensables, qui avaient fonctionné, et en

rajoutant le marqueur (SA01) caractéristique de la nouvelle lignée Q1a3a4. La réaction finale obtenue, qui a permis d'analyser l'ensemble des échantillons anciens, est donc composée de 15 SNP mitochondriaux et du chromosome Y (Figure 17), permettant de définir les lignées suivantes : A2, B2, C1b, C1c, C1d, D1, C, Q1a3*, Q1a3a*, Q1a3a1, Q1a3a2, Q1a3a3 et Q1a3a4.

L'association au sein de la même réaction de marqueurs mitochondriaux et du chromosome Y permet d'optimiser au maximum le nombre d'informations obtenues à partir d'une quantité minimale d'extrait utilisée, ce qui est un réel avantage pour les études d'ADNa et n'avait pas été décrit auparavant.

L'ensemble des séquences des des amorces PCR utilisées pour amplifier les fragments d'intérêt ainsi que les amorces SBE des 15 marqueurs de la réaction finale sont présentées en annexe 1 (p.232).

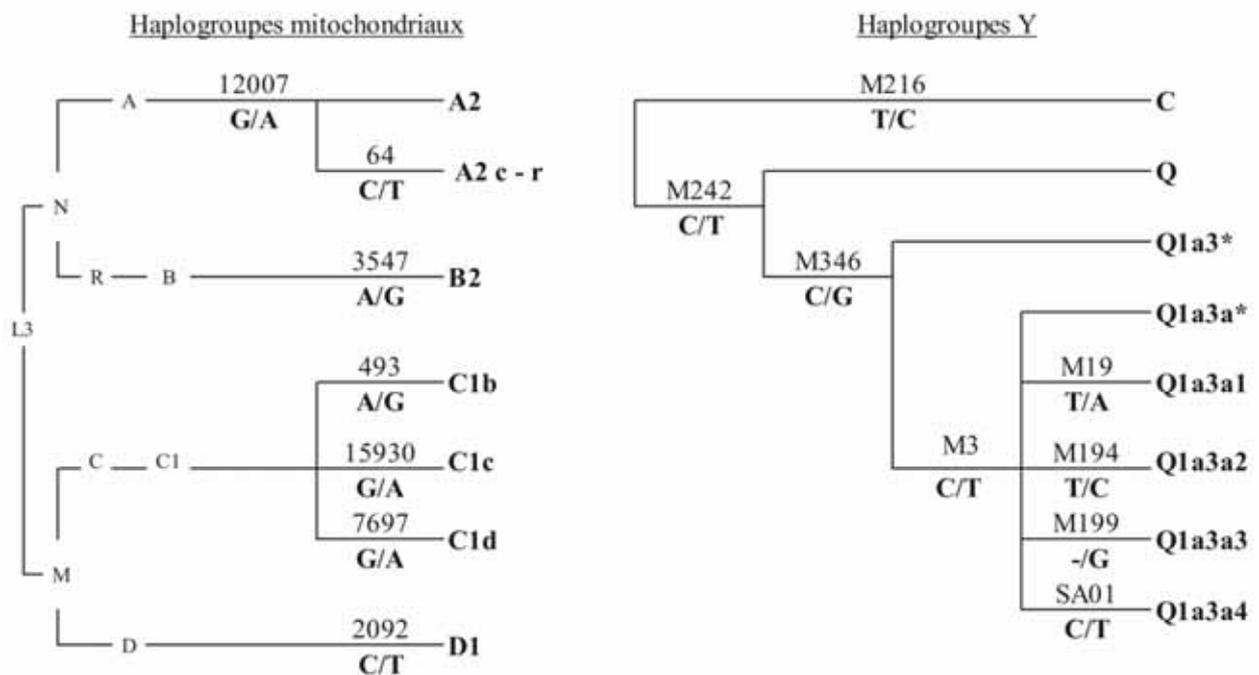


Figure 17 : Arbres phylogénétiques illustrant les haplogroupes mitochondriaux et Y déterminés, les marqueurs typés ainsi que les polymorphismes associés.

V. Authentification des résultats.

Pour les études d'ADNa, l'authentification des résultats est une étape cruciale mais qui peut parfois s'avérer délicate. Depuis l'apparition des premières études de paléogénétique et la mise en évidence que ces premiers résultats étaient pour la plupart des contaminations (Paabo *et al.*, 2004), de nombreux auteurs ont proposés différents critères permettant d'assurer l'authenticité des résultats (Cooper and Poinar, 2000; Gilbert *et al.*, 2005; Hofreiter *et al.*, 2001; Paabo *et al.*, 2004). Le suivi exact des critères peut s'avérer impossible. En effet, la quantité de matériel disponible ne permet pas toujours la réplication des analyses dans deux laboratoires indépendants, ou encore, des restes de faune associés aux échantillons analysés ne sont pas toujours disponibles. Il semble donc raisonnable d'adopter une stratégie adaptée à chaque situation et permettant un contrôle à plusieurs étapes du processus d'analyse, ainsi qu'en amont et en aval de cette analyse.

1. Préventions contre les contaminations.

La contamination est le problème le plus sérieux pour tous les chercheurs travaillant avec la technique de la PCR est plus particulièrement pour les analyses d'ADNa humain. Ceci est dû à la sensibilité de cette technique, qui tout en étant un réel avantage pour les études d'ADNa peut s'avérer être un désavantage, la PCR permettant de générer une très grande quantité de produits amplifiés à partir d'une quantité de produit aussi petite qu'une cellule unique (Higuchi *et al.*, 1984; Zhang *et al.*, 1992). Généralement, il subsiste peu de molécules d'ADN dans les restes humains anciens, et elles peuvent être sérieusement endommagées (fragmentées et/ou dégradées) et difficiles à amplifier. Au contraire, les molécules d'ADN contemporain sont omniprésentes dans notre environnement, non endommagées, et seront donc préférentiellement amplifiées par rapport à l'ADNa. Les précautions à suivre tout au long du processus d'analyse d'échantillons anciens doivent donc être constantes et rigoureuses.

1.1. Prélèvement des échantillons.

Le prélèvement direct des échantillons anciens sur le site de fouille est la meilleure garantie contre les contaminations de contact. Le prélèvement, par une seule personne (équipée de gants, d'un masque et d'une blouse), des échantillons anciens qui seront disposés directement dans des sachets uniques et stériles, permet d'éviter les contaminations par des fouilleurs ou des personnes en contact avec les échantillons avant l'analyse en laboratoire. Cependant, cette situation idéale n'est que rarement possible, et la grande majorité des études d'ADNa sont réalisées à partir d'échantillons issus de collections archéologiques conservées dans des musées, laboratoires ou organismes divers. Le risque que les échantillons soient contaminés par des fouilleurs et des manipulateurs divers (chercheurs, responsables des collections...) est beaucoup plus important dans ces cas.

Les échantillons anciens analysés dans le cadre de cette étude provenaient tous de collections archéologiques fouillées quelques dizaines d'années auparavant, et conservées dans les laboratoires de l'université de Buenos Aires et dans le Musée Gustavo le Paige de l'université catholique du Nord du Chili à San Pedro de Atacama. Les collections argentines étant relativement anciennes, et conservées par partie anatomique, il n'était pas possible de reconstituer des squelettes complets. Ainsi, pour avoir un maximum d'informations sur la situation des restes étudiés, et pour éviter d'analyser plusieurs fois le même individu, les prélèvements ont été faits uniquement sur les crânes. Ces collections ayant déjà été étudiées depuis leur découverte, nous avons choisi de prélever des dents intactes et en place dans la mandibule ou le maxillaire. Il semble en effet que les molécules d'ADN soient relativement bien conservées dans la chambre pulpaire, de plus, lorsque la dent est intacte, il est possible d'imaginer que peu de contamination sont « entrées » dans cette cavité. Enfin, comme présenté précédemment (Partie 2, III.1), un protocole de décontamination de la surface des dents est effectué avant toute analyse. En ce qui concerne les échantillons osseux (sites de Cortaderas et Pampa Grande), des fragments d'os longs (fémurs) dont la surface peut être facilement abrasée, ont été privilégiés.

Concernant, les prélèvements de la collection du Chili, nous avons eu l'occasion de nous rendre sur place afin d'étudier les ossements anciens et de pouvoir « choisir » les échantillons qui nous semblaient les plus adaptés pour des analyses génétiques. Dans ce cas nous avons choisi la même stratégie que précédemment, en privilégiant les dents intactes et en place dans

la mandibule ou le maxillaire. Les dents ainsi désignées ont été moulées, afin de garder une copie au musée et conserver les informations morphologiques de la dent, puis envoyées dans des sachets individuels et stériles en France.

1.2. Précautions lors de l'analyse.

1.2.1. *Bonnes pratiques de laboratoire.*

L'ensemble des analyses réalisées au cours de ce travail l'on été dans des locaux adaptés à l'analyse d'ADN humain ancien.

Les étapes pré- et post-PCR ont été réalisées dans des laboratoires physiquement séparés et isolés, où le sens de circulation des personnes, du matériel et des produits se fait toujours depuis le pré-PCR vers le post-PCR. De plus, les extractions et la préparation des mélanges réactionnels pour les amplifications ont été faites sous des hottes à flux laminaire dans des pièces séparées. L'entrée dans la partie pré-PCR est restreinte et ne se fait qu'avec un équipement adapté : sur-chaussures, blouse, masque, charlotte et gants stériles.

L'ensemble du matériel utilisé était stérile et à usage unique ou autoclavable. Enfin, les salles et paillasse de travail étaient régulièrement décontaminées (javel et DNA Away) et irradiées sous ultra-violets.

1.2.2. *Contrôles d'extraction et de réaction PCR.*

Chaque série d'extraction réalisée a été accompagnée d'un blanc d'extraction et toutes les amplifications associées à des témoins PCR (positifs et négatifs). Chaque blanc d'extraction et témoin PCR a été vérifié sur gel d'agarose, permettant de contrôler la survenue de contamination dans les réactifs utilisés, ou de contaminations croisées entre échantillons.

L'ensemble de ces précautions permettent de limiter les contaminations et sont essentielles, mais elles ne permettent pas à elles seules de garantir l'authenticité des résultats. Ces précautions doivent donc être associées à l'analyse de critères complémentaires afin de pouvoir valider les résultats obtenus.

2. Stratégies pour garantir l'authenticité des résultats.

2.1. Réplication des analyses.

Pour chacun des individus analysés au cours de ce travail, les analyses effectuées ont été répliquées à la fois à partir d'extractions indépendantes (avec un minimum de deux extractions par échantillons), mais également à partir d'échantillons indépendants provenant du même individu (un minimum de deux dents ou de deux fragments d'os par individu ont été analysées, lorsque cela était possible¹). Cette multiplication d'extraits obtenus à des moments différents permet par la suite de réaliser un minimum de quatre amplifications de la région HVR-1, ce qui permettra de construire des séquences consensus à partir des différentes séquences obtenues.

Les analyses de STR autosomaux ou du chromosome Y ont également été répliquées à partir d'extraits indépendants, pour chacun des échantillons. Dans ce cas également, un profil consensus est obtenu à partir des diverses amplifications réalisées à des moments différents. De plus, en ce qui concerne les STR autosomaux des kits commerciaux complémentaires ont été utilisés permettant de confirmer les résultats obtenus pour certains loci.

2.2. Stratégie multi-marqueurs.

Comme présenté au cours de ce chapitre, nous avons mis en place une stratégie multi-marqueurs, qui en plus d'apporter des informations différentes permettant de caractériser et comprendre les populations anciennes étudiées, permet également de valider les résultats obtenus. La cohérence entre les résultats obtenus par le biais de différents marqueurs sera évaluée. Par exemple, l'analyse de SNP mitochondriaux, en plus d'affiner la détermination des lignées maternelles, permet également, de confirmer la déduction faite grâce au séquençage direct. La sensibilité de cette technique de génotypage étant plus élevée que celle de la méthode de séquençage, certaines contaminations non visibles par séquençage pourront être révélées par le génotypage.

¹ Pour la collection de d'échantillons du site Solcor 3 du Chili, seulement une dent par individu a été obtenue, et par conséquent la réplication sur des échantillons indépendants n'a pu être réalisée.

2.3. Cohérence phylogénétique.

La cohérence entre les résultats obtenus et la région de provenance des échantillons peut être une indication supplémentaire pour la validation de l'authenticité des résultats. Les lignées aussi bien maternelles que paternelles amérindiennes sont peu nombreuses et facilement identifiables, soit par des mutations spécifiques sur la région HV-1, ou par des SNP caractéristiques.

Les populations contemporaines argentine et chilienne sont aujourd'hui largement métissées et présentent majoritairement des lignées européennes. Les contaminations exogènes par des européens seront très facilement détectables.

2.4. Comparaison des profils génétiques.

Afin de s'assurer qu'aucune personne ayant eu accès au laboratoire d'ADNa n'ait contaminé les échantillons, leur profil génétique est établi puis comparé à l'ensemble des profils de chaque échantillon ancien.

Cette stratégie peut également être utilisée pour les fouilleurs, cependant, dans le cas qui nous intéresse, les fouilles ayant été réalisées il y a un grand nombre d'années et les collections ayant été intensément étudiées, il nous a été impossible de retrouver l'ensemble des personnes ayant eu accès à ces échantillons. Nous avons pu obtenir, avec leur consentement éclairé, les profils des personnes ayant fouillé un certain nombre des sites argentins, mais pour le site chilien, cela n'a pas pu être le cas.

En plus de la comparaison, l'établissement de profils génétiques permet dans certains cas de mettre en évidence le comportement particulier de l'ADNa. En effet, une décroissance de l'intensité du signal pour les amplicons de plus grande taille est parfois observée ce qui traduit l'état de dégradation de l'ADNa qui n'est présent que sous la forme de courts fragments. Ces profils sont donc typiques de l'ADNa et peuvent être une indication sur l'état de dégradation de l'ADN. Il ne s'agit cependant pas d'un critère d'authentification direct, étant donné qu'une contamination ancienne, par exemple lors de rituels particuliers dans des sépultures secondaires, ou lors de pillages, peut conduire au même type de profils.

VI. Analyses des données génétiques.

Les données récoltées seront analysées sur deux plans principaux : (1) dans un premier temps, ces données seront analysées sur un plan individuel dans le but de caractériser chacun des individus anciens, en révélant leurs affinités génétiques (liens de parenté), mais aussi leur origine ethnique ; (2) dans un second temps, les données pourront être analysées à un niveau populationnel, dans le but cette fois-ci de comprendre les relations et affinités entre les populations de la région mais aussi d'estimer leur dynamique d'évolution, ceci afin d'apporter de nouvelles données permettant d'appréhender le peuplement des Andes méridionales et de l'aire andine en général.

Dans le cadre de ce travail nous entendrons par population un ensemble d'individus non apparentés vivant dans un même environnement à une période donnée. Nous sommes conscients que constituer une « population » à partir d'une « population de cimetière » peut induire certains biais, ainsi, tous les résultats obtenus seront analysés avec précautions en gardant à l'esprit qu'il s'agit de tendances générales qui permettent de se faire une première idée sur l'histoire de ces groupes anciens.

1. Populations de comparaison : établissement de bases de données de référence.

Pour déceler les affinités ethniques des individus anciens, mais aussi pour décrypter l'histoire évolutive des populations anciennes étudiées, des comparaisons avec des populations anciennes et/ou contemporaines d'Amérique du Sud sont indispensables. Les bases de données en ligne ne disposant pas toujours des informations souhaitées, nous avons choisi de constituer des bases de données internes (GeneData v.1), en compilant les données disponibles au laboratoire et dans la littérature (haplotypes et haplogroupes). Cette démarche a été suivie pour les données concernant l'ADNmt mais aussi pour les données sur le chromosome Y.

Par comparaison avec d'autres régions du monde (Asie ou Europe par exemple) les données sur les populations sud-américaines sont relativement peu nombreuses, surtout en ce qui

concerne les lignées paternelles. En ce qui concerne l'ADNmt, nous avons pu compiler les données concernant 75 populations ou ethnies sud-américaine (natives ou métissées), et obtenir 2269 séquences HVR-1 de comparaison (16024-16383 pb). Les données de référence du chromosome Y concernent 27 populations sud-américaines, et nous ont permis de constituer une base de 3732 haplotypes Y. La base de données en ligne Y-HRD [Y-STR haplotype reference database, (<http://www.ystr.org/index.html>)] a également été utilisée en complément de notre base de données personnelle.

Les informations principales concernant ces populations de comparaison ainsi que leur localisation sont présentés dans les tableaux 6 et 7 suivants et sur la figure 19.

L'histoire des populations amérindiennes étant largement liée à leur environnement, les populations, ethnies ou échantillons de comparaison ont été classés en fonction de différentes aires géographiques et culturelles illustrées sur la figure suivante 18.



Figure 18 : Représentation des principales aires écologiques d'Amérique du Sud.

Tableau 6 : Tableau récapitulatif des populations anciennes et contemporaines sud-américaines dont les données ont été utilisées pour réaliser les analyses statistiques concernant le génome mitochondrial.

Populations / Sites	n ^{o1}	Rég ²	N ³	Données ⁴	Référence
Colombie					
Arawak	1	1	50 (30)	Hg-Ht	(Melton <i>et al.</i> , 2007)
Chibchan	2	1	138 (80)	Hg-Ht	(Melton <i>et al.</i> , 2007)
Pool Nord Colombie	3	1	379	Hg	(Keyeux <i>et al.</i> , 2002; Torres <i>et al.</i> , 2006)
Zenu - Embera	4	2	59	Hg	(Mesa <i>et al.</i> , 2000)
Pool Andes Colombie	5	2	189	Hg	(Keyeux <i>et al.</i> , 2002; Torres <i>et al.</i> , 2006)
Ingano	6	2	27	Hg	(Mesa <i>et al.</i> , 2000)
Pool Est Colombie	7	5	433	Hg	Keyeux <i>et al.</i> , 2002; Torres <i>et al.</i> , 2006)
Ticuna	8	5	82	Hg	(Torrioni <i>et al.</i> , 1993; Mesa <i>et al.</i> , 2000)
Venezuela					
Caracas	9	1	100	Hg-Ht	(Lander <i>et al.</i> , 2008)
Guahibo	10	1	59	Hg-Ht	(Vona <i>et al.</i> , 2005)
Yanomami.1	11	5	155	Hg-Ht	(Williams <i>et al.</i> , 2002)
Equateur					
Cayapa	12	2	62 (32)	Hg-Ht	(Rickards <i>et al.</i> , 1999; Tamm <i>et al.</i> , 2007)
Guyane					
Palikur	13	5	48	Hg-Ht	(Mazieres <i>et al.</i> , 2008)
Emerillon	14	5	30	Hg-Ht	(Mazieres <i>et al.</i> , 2008)
Kali'na	15	5	29	Hg-Ht	(Mazieres <i>et al.</i> , 2008)
Wayampi	16	5	53	Hg-Ht	(Mazieres <i>et al.</i> , 2008)
Brésil					
Apalai	17	5	102	Hg-Ht	(Mazieres <i>et al.</i> , 2008)
Yanomami.2	18	5	83	Hg-Ht	(Easton <i>et al.</i> , 1996)
Pool Nord Brésil	19	5	48	Hg-Ht	(Alves-Silva <i>et al.</i> , 2000)
Gaviao	20	5	27	Hg-Ht	(Ward <i>et al.</i> , 1996)
Zoro	21	5	30	Hg-Ht	(Ward <i>et al.</i> , 1996)
Xavante	22	5	28 (25)	Hg-Ht	(Ward <i>et al.</i> , 1996)
Pool Nord Est Brésil	23	5	50	Hg-Ht	(Alves-Silva <i>et al.</i> , 2000)
Pool Sud Est Brésil	24	8	100	Hg-Ht	(Alves-Silva <i>et al.</i> , 2000)
Guarani	25	8	200	Hg-Ht	(Marrero <i>et al.</i> , 2007)
Pool Sud Brésil	26	8	50	Hg-Ht	(Alves-Silva <i>et al.</i> , 2000)
Kaingang	27	8	78	Hg-Ht	(Marrero <i>et al.</i> , 2007)
Amazonie*	28	8	18	Hg-Ht	(Ribeiro-do-Santos <i>et al.</i> , 1996)
Pérou					
Yungay	29	3	36	Hg-Ht	(Lewis <i>et al.</i> , 2007b)
Ancash	30	3	33	Hg-Ht	(Lewis <i>et al.</i> , 2005)
Lima / Pasco	31	3	52	Hg	(Rodriguez-Delfin <i>et al.</i> , 2001)
San Martin	32	3	22	Hg-Ht	(Fuselli <i>et al.</i> , 2003)
Tupe	33	3	16	Hg-Ht	(Lewis <i>et al.</i> , 2007b)
Tayacaja	34	3	61	Hg-Ht	(Fuselli <i>et al.</i> , 2003)
Quechua.1	35	3	19	Hg	(Merriwether <i>et al.</i> , 1995)
Puno	36	3	44	Hg-Ht	(Lewis <i>et al.</i> , 2007b)
Arequipa	37	3	22	Hg-Ht	(Fuselli <i>et al.</i> , 2003)
Titicaca	38	3	64	Hg-Ht	(Barbieri <i>et al.</i> , 2011)
Matsiguenga	39	5	38	Hg-Ht	(Mazières <i>et al.</i> , 2008)
Sican* ⁵	40	4	28	Hg-Ht	(Shimada <i>et al.</i> , 2004)
Ayacucho*	41	3	32 (28)	Hg-Ht	(Kemp <i>et al.</i> , 2009)
Palpa*	42	4	130 (104)	Hg-Ht	(Fehren-Schmitz <i>et al.</i> , 2010a)
Laramate*	43	3	72 (65)	Hg-Ht	(Fehren-Schmitz <i>et al.</i> , 2010b)
Ollantaytambo* ⁵	44	3	35	Hg-Ht	(Shinoda <i>et al.</i> , 2006)
ChenChen*	45	3	23	Hg	(Lewis <i>et al.</i> , 2007a)

Populations / Sites	n ^{o1}	Rég ²	N ³	Données ⁴	Référence
Bolivie					
Aymara.1	46	3	301 (96)	Hg-Ht	(Bert <i>et al.</i> , 2001; Gaya-Vidal <i>et al.</i> , 2011; Merriwether <i>et al.</i> , 1995)
La Paz	47	3	110	Hg-Ht	(Afonso Costa <i>et al.</i> , 2010)
Piedmont andin	48	3	46	Hg-Ht	(Corella <i>et al.</i> , 2007)
Quechua.2	49	3	125 (93)	Hg-Ht	(Bert <i>et al.</i> , 2001 ; Gaya-Vidal <i>et al.</i> , 2011)
Plaines Moxos	50	5	54	Hg-Ht	(Bert <i>et al.</i> , 2004)
Llanos pool	51	5	168	Hg	(Bert <i>et al.</i> , 2001)
Ayoreo	52	7	57	Hg	(Dornelles <i>et al.</i> , 2004)
Tiwanaku*	53	3	13	Hg	(Rothhammer <i>et al.</i> , 2003)
Paraguay					
Ayoreo	54	7	34	Hg	(Dornelles <i>et al.</i> , 2004)
Ache	55	7	63	Hg-Ht	(Schmitt <i>et al.</i> , 2004)
Uruguay					
Tuacarembos	56	8	100	Hg-Ht	(Bonilla <i>et al.</i> , 2004)
Montevideo	57	8	108	Hg	(Gascue <i>et al.</i> , 2005)
Chili					
Aymara.2	58	3	120	Hg	(Rocco <i>et al.</i> , 2002)
Atacama	59	6	73	Hg	(Merriwether <i>et al.</i> , 1995 ; Rocco <i>et al.</i> , 2002)
Pehuenche	60	10	205 (24)	Hg-Ht	(Merriwether <i>et al.</i> , 1995 ; Moraga <i>et al.</i> , 2000)
Huilliche - Mapuche	61	10	191 (34)	Hg-Ht	(Merriwether <i>et al.</i> , 1995 ; Moraga <i>et al.</i> , 2000)
Chiloe	62	10	158	Hg	(Garcia <i>et al.</i> , 2006)
Yaghan	63	10	21 (15)	Hg-Ht	(Moraga <i>et al.</i> , 2000)
Arica*	64	3	83 (30)	Hg-Ht	(Moraga <i>et al.</i> , 2005)
Fuegian* ⁵	65	10	24	Hg-Ht	(Garcia-Bour <i>et al.</i> , 2004)
Argentine					
Jujuy	66	6	100	Hg	(Cardoso <i>et al.</i> , 2009)
Puna	67	6	65	Hg	(Dipierri <i>et al.</i> , 1998)
Coyas	68	6	61	Hg-Ht	(Alvarez-Iglesias <i>et al.</i> , 2007)
Salta	69	6	18	Hg-Ht	(Tamm <i>et al.</i> , 2007)
Nord Ouest Argentine	70	6	98	Hg	(Bobillo <i>et al.</i> , 2009)
Catamarca/Tucuman	71	6	34	Hg-Ht	(Tamm <i>et al.</i> , 2007)
Centre Ouest Argentine	72	6	358 (193)	Hg	(Bobillo <i>et al.</i> , 2009; Motti <i>et al.</i> , 2009)
Gran Chaco	73	7	345 (204)	Hg-Ht	(Cabana <i>et al.</i> , 2006; Demarchi <i>et al.</i> , 2001b)
Centre Argentine	74	9	475	Hg	(Corach <i>et al.</i> , 2010; Garcia and Demarchi, 2009)
Sud Argentine	75	10	47	Hg	(Bobillo <i>et al.</i> , 2009)

¹ le numéro se rapporte à la figure 19, illustrant la situation des populations étudiées.

² le numéro indiqué correspond à la région géographique dans laquelle se trouve la population, figure 18.

³ les données entre parenthèses représentent le nombre de séquences HVR-1 disponibles.

⁴ Hg = seul les fréquences des haplogroupes sont disponibles; Hg-Ht = les fréquences des haplogroupes ainsi que les séquences HVR-1 sont disponibles.

⁵ la séquence HVR-1 est partielle : seulement entre les positions 16209-16402

*Populations anciennes.

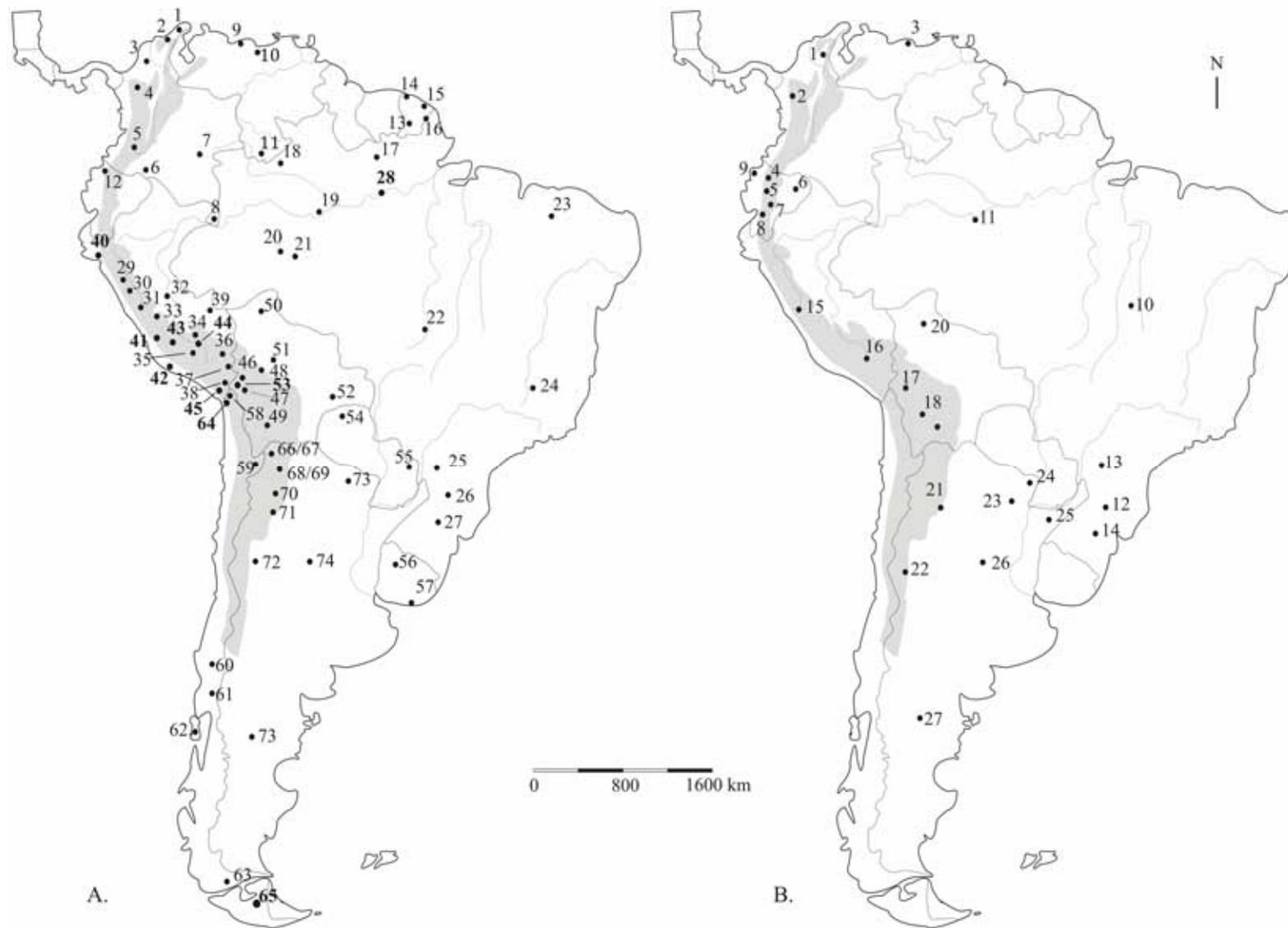


Figure 19 : Localisation des populations de comparaison utilisées dans ce travail : (A) populations dont les données ADNmt ont été utilisées, (B) populations dont les données Y ont été utilisées

Tableau 7 : Récapitulatif des populations amérindiennes contemporaines dont les données sur le chromosome Y ont été utilisées pour les analyses statistiques.

Population	n ^{o1}	Reg ²	N ³	Données ⁴	Références
Colombie					
Pool Nord Colombie	1	1	305	Ht	(Romero <i>et al.</i> , 2008)
Pool Antioquia	2	2	1177	Ht	(Builes <i>et al.</i> , 2006; Gaviria <i>et al.</i> , 2005)
Venezuela					
Caracas	3	1	62	Ht	(Tovar <i>et al.</i> , 2006)
Equateur					
Quito	4	2	120	Ht	(Baeza <i>et al.</i> , 2007)
Cayapa	5	2	26	Ht	(Tarazona-Santos <i>et al.</i> , 2001)
Waorani	6	2	35	Ht	(Gonzalez-Andrade <i>et al.</i> , 2009)
Kichwa	7	2	102	Ht	(Gonzalez-Andrade <i>et al.</i> , 2009)
Metisses	8	2	102	Ht	(Gonzalez-Andrade <i>et al.</i> , 2009)
Afro-equatoriens	9	2	102	Ht	(Gonzalez-Andrade <i>et al.</i> , 2009)
Brésil					
Pool Brésil	10	5	500	Ht	(Pereira <i>et al.</i> , 2007)
Pool Nord Brésil (Am)	11	5	84	Ht	(Tarazona-Santos <i>et al.</i> , 2001)
Pool Sud Brésil	12	8	249	Ht	(Leite <i>et al.</i> , 2008)
Guarani	13	8	78	Hg	(Bortolini <i>et al.</i> , 2003; Marrero <i>et al.</i> , 2007)
Kaingang	14	8	58	Hg	(Bortolini <i>et al.</i> , 2003; Marrero <i>et al.</i> , 2007)
Pérou					
Pérou.1	15	3	79	Ht	(Iannacone <i>et al.</i> , 2005)
Arequipa - Tayacaja	16	3	59	Ht	(Tarazona-Santos <i>et al.</i> , 2001)
Bolivie					
Aymara	17	3	59	Hg-Ht	(Gaya-Vidal <i>et al.</i> , 2011)
Altiplano	18	3	59	Ht	(Lee <i>et al.</i> , 2007)
Quechua	19	3	55	Hg-Ht	(Gaya-Vidal <i>et al.</i> , 2011)
Beni.1	20	5	229	Ht	(Lee <i>et al.</i> , 2007; Tirado <i>et al.</i> , 2009)
Argentine					
Pool Colla	21	6	63 (34)	Hg-Ht	(Blanco-Verea <i>et al.</i> , 2010; Toscanini <i>et al.</i> , 2008)
Diaguita	22	6	24	Hg-Ht	(Blanco-Verea <i>et al.</i> , 2010)
Pool Toba	23	7	89	Ht	(Berardi <i>et al.</i> , 2003; Toscanini <i>et al.</i> , 2008)
Gran Chaco	24	7	78 (60)	Ht	(Demarchi and Mitchell, 2004; Toscanini <i>et al.</i> , 2008)
Guarani	25	7	33	Hg-Ht	(Altuna <i>et al.</i> , 2006)
Cordoba	26	9	100	Ht	(Fondevila <i>et al.</i> , 2003)
Mapuche	27	10	76	Hg-Ht	(Blanco-Verea <i>et al.</i> , 2010)

¹ le numéro se rapporte à la figure 19, illustrant la situation des populations étudiées.

² le numéro indiqué correspond à la région géographique dans laquelle se trouve la population, figure 18.

³ les données entre parenthèses représentent le nombre d'haplotypes disponibles.

⁴ Hg = seul les fréquences des haplogroupes sont disponibles; Ht = seuls les haplotypes sont disponibles ; Hg-Ht = les fréquences des haplogroupes ainsi que les haplotypes sont disponibles.

2. Identification individuelle et détermination des liens de parenté.

Le recrutement funéraire est une problématique qui intéresse de nombreux archéologues et anthropologues. La génétique est à ce jour le seul outil permettant de déterminer avec fiabilité les liens génétiques unissant les individus inhumés au sein d'un même espace. De plus, l'anthropologie moléculaire bénéficie aujourd'hui des avancées technologiques faites dans le domaine de la criminalistique.

L'analyse des haplotypes mitochondriaux et du chromosome Y fournissent une première indication permettant de supposer des potentiels liens de proche parenté. En effet, le partage d'un même haplotype par deux individus, d'autant plus s'ils sont inhumés très proches l'un de l'autre conduit à étudier plus en détail la relation les unissant. Pour cela, les profils génétiques seront étudiés et les possibles relations déterminées à l'aide de plusieurs logiciels : ML-Relate (Kalinowski *et al.*, 2006), Familias v.1.97 (Egeland *et al.*, 2000) et DNA•VIEW™ (Brenner, 1997). Les relations de type parent/enfant et de type fratrie pourront être testées. Des probabilités [rapports de vraisemblance ou LR (Likelihood Ratio)] associées à chacune des hypothèses émises seront calculées à partir des algorithmes propres de chaque logiciel, et à partir d'une base de données de référence de fréquences alléliques.

Les données sur les fréquences alléliques de marqueurs autosomaux de populations anciennes étant très rares, nous avons utilisé une base de données de comparaison publique, disponible sur internet, et référençant les fréquences alléliques de STR autosomaux d'individus contemporains de toutes les provinces de l'Argentine (Facultad de Bioquímica, huellas digitales : <http://www.ffyb.uba.ar/gxpsites/hgxpp001.aspx?2>).

La mise en évidence d'individus apparentés au sein des groupes anciens étudiés permettra d'exclure ces individus pour les analyses suivantes et ainsi de constituer des « populations archéologiques » ne présentant aucuns individus apparentés.

3. Analyses statistiques.

Avant toute analyse statistique, les séquences consensus HVR-1, haplotypes consensus Y et les génotypes consensus déterminés à partir des STRa ont été construits à partir d'au moins quatre profils ou séquences obtenus à partir d'extraits et d'amplifications différents. La

détermination des haplogroupes par analyse des SNP a été faite au cours d'au moins deux génotypages indépendants.

Dans un premier temps, les fréquences des haplotypes et haplogroupes observés au sein de nos populations anciennes ont été calculées par comptage direct. Un réseau phylogénétique a été créé à partir des séquences HVR-1 (16024-16383 pb) des populations étudiées. Ce réseau, tenant compte des poids évolutifs des différentes mutations (Hasegawa *et al.*, 1993) permet de révéler les mutations détectées et leur répartition entre les différents groupes.

Les haplotypes mitochondriaux et Y détectés dans notre échantillon ont par la suite été comparés à nos bases de données personnelles afin de révéler des **haplotypes partagés** et ainsi estimer l'origine des individus. La distribution de ces haplotypes a également pu être illustrée grâce à des **analyses phylogénétiques**, et plus particulièrement la construction de **réseaux phylogénétiques** par la méthode **Median Joining** (MJ) grâce au logiciel Network v.4.6 (Fluxus Technology : <http://www.fluxus-engineering.com/sharenet.htm>) (Bandelt *et al.*, 1999). Cette méthode d'analyse, qui permet de prendre en compte l'ensemble de la variabilité présente (le poids moléculaire de chaque mutation HVR-1 et le taux de mutation de chaque STR Y ont été fixés à partir des données de la littérature), a été utilisée aussi bien pour les données des lignées paternelles que maternelles. L'objectif étant d'avoir un aperçu de la distribution des haplotypes entre les différentes régions d'Amérique du Sud et de révéler les affinités de nos échantillons anciens par rapport à la diversité connue, nous avons choisi d'appliquer l'option dite de « star contraction », dans le but de simplifier les réseaux obtenus (Forster *et al.*, 2001). Avec le même objectif, l'option « MP calculation » a permis d'enlever les liens et vecteurs médians superflus du réseau (Polzin and Daneschmand, 2003).

La comparaison des fréquences d'haplotypes peut également permettre de révéler des évènements évolutifs à l'échelle de la population. Cependant, les données concernant les haplotypes de populations amérindiennes étant encore relativement peu nombreuses, il est probable de ne pas trouver de correspondances. Pour palier à ce problème, nous nous sommes également intéressés à la distribution des fréquences d'haplogroupes, notamment mitochondriaux, que nous avons illustrée grâce à une analyse en composantes principales (ACP) réalisée avec le logiciel XLSTAT v.5.7.2.

Une fois les populations archéologiques correctement définies, les **indices de diversité** classiques ont été calculés : diversité nucléotidique π , diversité haplotypique H_d , le nombre

moyen de substitutions entre paires d'haplotypes P_w , à l'aide des logiciels DnaSP v.5.10.01 (Rozas *et al.*, 2003) et Arlequin v.3.11 (Excoffier *et al.*, 2005). L'indice de neutralité F_S de Fu a également été calculé à l'aide du logiciel Arlequin v.3.11, permettant de déterminer si une population est stationnaire (valeur nulle ou positive de l'indice F_S) ou en expansion (valeur négative significative de l'indice F_S). Ces paramètres standards de diversité ont par la suite pu être comparés à ceux des populations de référence anciennes et contemporaines.

Pour finir, la **différentiation génétique** entre les différentes populations a été estimée grâce à l'obtention de matrices de distances F_{ST} , avec le logiciel Arlequin v.3.11 (Excoffier *et al.*, 2005). L'indice de fixation F_{ST} permet d'indiquer le degré de divergence entre deux groupe [entre 0 et 0,05 la divergence est faible ; entre 0,05 et 0,15 la divergence est modérée ; entre 0,15 et 0,25 la divergence est importante, et au dessus de 0,25 la divergence est très importante]. La significativité de ces distances, au seuil de 0,05, a été testée grâce aux 1 000 permutations réalisées.

Les distances entre populations calculées sur la base des haplotypes mitochondriaux ont été représentées graphiquement par une **analyse multidimensionnelle** (Multi Dimensional Scaling, MDS), réalisée avec le logiciel XLSTAT v.5.7.2. Les MDS sont intéressantes car elles permettent de comparer un grand nombre de populations avec un effectif relativement important. Cependant, il ne s'agit que d'une représentation graphique et non pas du reflet exact des valeurs de F_{ST} obtenues, ainsi, ces MDS seront toujours interprétés en tenant compte des matrices de distances.

Les distances génétiques moyennes entre les groupes, obtenues à partir des fréquences de lignées mitochondriales et des haplotypes Y (haplotype minimal moins le loci DYS385) ont été représentées par des **arbres phylogénétiques** (Neighbor Joining (NJ) et UPGMA), en utilisant le programme MEGA v.5 (Tamura *et al.*, 2011).

Troisième partie.

Résultats et analyses génétiques.

I. Caractérisation des échantillons anciens étudiés.

Dans ce premier chapitre, l'ensemble des résultats bruts obtenus à partir des différents marqueurs analysés pour tous les échantillons anciens étudiés seront présentés. Certaines particularités observées seront interprétées au sein même de ce chapitre dans le but de ne pas alourdir la discussion. De plus, ces résultats bruts seront interprétés afin de révéler les potentiels liens de parenté entre les échantillons anciens, et pouvoir définir des populations archéologiques.

1. Taux de génotypage et conservation de l'ADN.

Les analyses que nous avons réalisées dans le cadre de ce travail démontrent que l'ADN était relativement bien conservé au sein des échantillons anciens auxquels nous avons eu accès. En effet, comme le montre le tableau 8 suivant, le taux de génotypage des marqueurs mitochondriaux est relativement élevé, autour de 70%. Seuls les échantillons anciens chiliens, du site de Solcor 3 étaient beaucoup trop dégradés pour en tirer une quelconque information. Le taux de génotypage des marqueurs autosomaux, qui est de 70%, est particulièrement intéressant, et montre que même les marqueurs nucléaires peuvent être étudiés sur des échantillons anciens avec un taux de réussite important. Il faut cependant noter que ce pourcentage comprend tous les échantillons pour lesquels au moins un marqueur nucléaire a été amplifié de manière reproductible. Le taux de génotypage des marqueurs du chromosome Y est un peu plus faible (autour de 50 %) mais reste satisfaisant étant donné la nature des échantillons analysés.

Les résultats de la quantification des extraits d'ADNa sont présentés en annexe 2 (p.233). Les valeurs varient très fortement en fonction des sites et des échantillons mais aussi en fonction des extraits d'ADN quantifiés. Les concentrations d'ADNnu obtenues s'échelonnent de quelques picogrammes par μL à $1,68 \text{ ng}/\mu\text{L}$, avec une moyenne de $131,3 \text{ pg}/\mu\text{L}$. Il est important de rappeler que cette quantification cible un fragment d'ADNnu et que par conséquent nous ne pouvons pas connaître la quantité d'ADNmt présente dans l'extrait, qui *a priori* est plus importante.

Sans surprise, certains des échantillons pour lesquels nous n'avons pas obtenu de résultats exploitables présentaient une quantité tellement infime d'ADN qu'elle ne pouvait être quantifiée. La faible quantité d'ADN n'est pourtant pas la seule raison de l'absence de résultats interprétables. En effet, dans certains cas, même si l'ADN semblait présent dans des quantités raisonnables, il apparaissait par la suite trop dégradé ou contaminé, empêchant toute analyse.

Tableau 8 : Taux de génotypage obtenu pour chacun des sites anciens étudiés et chacun des marqueurs génétiques analysés.

	N ¹	Mt Ht ²	Mt Hg ²	Y-STR ²	Y Hg ²	aSTR
Banda de Perchel	3	1	2	0	0	1
Huacalera	1	1	1	1	1	1
Juella	5	3	3	2	2	3
Los amarillos	24	17	20	12	12	20
Muyuna	8	6	6	3	3	5
Penas Blancas	1	0	0	0	0	0
San Jose	1	1	1	1	1	1
Sarahuaico	2	2	2	1	1	2
Tilcara	10	8	10	4	4	7
Quebrada de Humahuaca total	55	39	45	24	24	40
Cortaderas	3	3	2	2	1	3
Tero (SSal Cac 14)	8	5	5	3	3	5
Fuerte Alto	2	1	2	1	1	1
Calchaqui total	13	9	9	6	5	9
Doncellas	3	2	3	1	1	2
El Litro	4	4	4	3	3	4
Caverna II	4	3	4	1	1	4
Lampazar 4	5	4	5	3	3	4
Caverna V	4	4	4	2	3	4
Los Aparejos	4	4	4	2	2	4
Pampa Grande total	21	19	21	11	12	20
Solcor 3	11	0	0	0	0	0
TOTAL (N / %)	103	69 / 67%	78 / 76%	42 / 52%	42 / 52%	71 / 69%

¹ Effectif absolu du nombre d'échantillons.

² Mt = ADN mitochondrial ; Y = chromosome Y ; Ht = haplotypes; Hg = haplogroupes.

2. STR autosomaux et amélogénine.

L'établissement des profils génétiques est très important pour plusieurs raisons, déjà évoquées précédemment.

Comme présenté dans le tableau 8, le taux de génotypage des STR autosomaux est assez élevé pour des échantillons anciens, puisqu'il atteint les 70%. Ce taux correspond au pourcentage d'échantillons pour lesquels nous avons obtenu des résultats reproductibles pour au moins un marqueur autosomal. Le pourcentage d'échantillons pour lesquels nous avons obtenu au moins la moitié des loci amplifiés est d'environ 54% ce qui reste tout à fait exceptionnel pour des échantillons d'ADN dégradé et présent en faible quantité.

L'ensemble des génotypes consensus obtenus à partir de l'analyse des STR autosomaux sont présentés dans le tableau 9.

La comparaison de ces génotypes aux profils des fouilleurs et manipulateurs nous a permis de confirmer qu'aucun échantillon n'a été contaminé par une personne ayant eu accès au laboratoire d'ADN ancien, ni aucune personne ayant manipulé les échantillons anciens récemment. Les profils complets des fouilleurs et manipulateurs sont présentés en annexe 3 (p.234).

La comparaison deux à deux de l'ensemble des profils génétiques a mis en évidence que certains échantillons partageaient exactement le même génotype (génotypes grisés dans le tableau 9). Cette situation se répète pour quatre sites différents : LA-9 et LA-16 du site Los Amarillos, Til-5 et Til-7 du site de Tilcara, Mu-5 et Mu-7 du site de Muyuna et Sc-2 et Sc-8 du site de Tero (S Sal Cac 14). Dans trois de ces cas (Los Amarillos, Tilcara et Tero), l'hypothèse privilégiée est celle d'une erreur lors de l'identification et l'individualisation des échantillons par les archéologues. En effet, la plupart de ces restes anciens proviennent de sépultures multiples, quelquefois secondaires, et ont été collectés il y a de nombreuses années, puis entreposés dans des musées. Le prélèvement de ces échantillons issus de sites archéologiques d'Argentine a été réalisé sur place, par une archéologue de l'Université de Buenos Aires. Des dents encore en place ont été prélevées, aussi bien sur des mandibules que des maxillaires, attribués *a priori* à des individus différents.

En ce qui concerne les échantillons LA-9 et LA-16, Mu-5 et Mu-7 ainsi que Til-5 et Til-7 par

exemple, comme nous pouvons le voir sur la figure 20, les échantillons ont été prélevés à partir de mandibule et maxillaires qui pourraient être associées à un seul et même individu. L'hypothèse que ces individus soient de vrais jumeaux est extrêmement difficile à prouver, et la fréquence élevée observée au sein de notre échantillon tend à appuyer l'hypothèse d'un biais lors du prélèvement des échantillons. Pour la suite de l'étude, les échantillons correspondant à des doublons ont été enlevés (LA-16, Til-7, Sc-8, Mu-7). Ainsi, la cohorte analysée est composée de 99 individus et non 103.

En ce qui concerne le site de Muyuna, la situation semble quelque peu différente. En effet, comme indiqué par les archéologues, les échantillons étaient en très mauvais état de conservation (Seldes, 2006), certains ayant même subi une crémation. L'ensemble des analyses menées sur ces individus ont montrés qu'ils présentaient tous des profils génétiques, ainsi que des haplotypes mitochondriaux et Y très proches, voir identiques. Même si les haplogroupes déterminés sont cohérents avec la région étudiée, le doute d'une contamination étant trop important, nous avons décidé de ne pas utiliser les résultats obtenus à partir des échantillons de ce site pour les analyses statistiques.



Figure 20 : Echantillons pour lesquels les prélèvements réalisés ont révélé un profil génétique identique.

Grâce à l'établissement de ces profils génétique, nous avons pu déterminer le sexe génétique de 66 des 99 échantillons anciens analysés. Parmi ces 66 individus 44 ont été identifiés de manière fiable et reproductible comme étant des individus masculins et 22 comme étant féminins. Il apparaît donc un certain déséquilibre en faveur des hommes dans cette cohorte, avec pratiquement deux fois plus d'individus masculins que féminins.

La détermination du sexe morphologique, par des méthodes anthropologiques « classiques », avait déjà été faite pour les échantillons des sites de Cortaderas, Pampa Grande et Solcor 3. Si pour le site de Solcor 3 la mauvaise conservation de l'ADN n'a pas permis de comparer les résultats obtenus grâce aux différentes méthodes, cela a été possible pour les deux autres sites. Cette comparaison a été possible pour 21 échantillons pour lesquels les deux informations étaient disponibles, révélant des résultats concordants pour 16 d'entre eux. Ainsi, pour 4 échantillons du site de Pampa Grande (Li-2, Li-3, C.II-2 et Ap-3) et l'échantillon Cor-2 du site de Cortaderas, le sexe morphologique et génétique ne correspondent pas. Pour quatre de ces échantillons (Li-2, Li-3, C.II-2 et Cor-2), le sexe morphologique avait été désigné comme féminin et s'est avéré être masculin. Ces résultats pourraient être dus à un faible dimorphisme sexuel au sein de ces populations, ou aux méthodes de détermination du sexe morphologique employées.

Tableau 9 : Profils génétiques consensus obtenus, pour les individus anciens de l'ensemble des sites archéologiques analysés, grâce à l'utilisation de deux kits complémentaires : Minifiler et Identifier + (Applied Biosystems).

	AMEL	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Doncellas																
Don-1	XY										14/15					
Don-2	XY					15/15					14/14					
Pampa Grande																
Li-1	XY	9/14	29/32,2	11/13		15/15		9/13				16/17			11/11	21/23
Li-2	XY	13/13				15/15		9/12				17/17			11/11	
Li-3	XY	10/13	31,2/32,2			17/17		11/13				16/18		12/12	7/9	23/25
Li-4	XY	13/15	30/31	11/12		15/15		12/13				16/17		15/18	9/11	21/22
C.II-1	XY	13/13	31,2/33,2	11/14		15/15		9/13				16/17		14/14	10/11	20/25
C.II-2	XY		30/38													23/24
C.II-3	XX	10/14	32,2/33,2			15/15		9/13				16/18		13/13	10/11	21/25
C.II-4	XX	12/13	30/30	11/11		14/16		10/13				15/16		12/15	11/11	21/22
C.V-1	XY		24,2/35			14/15						19/19				
C.V-2	XY	13/13	30/31,2			15/17		9/13				15/17		13/13	9/9	21/21
C.V-3	XY	12/13	30/30	11/11		16/17		9/13				16/16			8/9	20/24
C.V-4	XX	11/15				15/15		9/12				16/17		14/15	9/11	22/25
Lam-2	XY	12/14	28/30			15/15		12/13				17/19		15/15	11/11	21/25
Lam-3	XY	11/12	31,2/31,2			15/15		9/12				15/17		12/14	11/11	21/24
Lam-4	XY	14/14				15/15		12/12				17/19		18/18	11/11	22/26
Lam-5	XX	14/14	31,2/31,2			15/17		12/14				16/17		12/16	9/12	20/24
Ap-1	XX	13/15	30/32,2			16/16		9/10				16/20		16/16	9/12	20/26
Ap-2	XY	11/15	31/31,2	11/12		15/18		12/13				16/16		13/15	9/11	21/25
Ap-3	XX					15/15						17/17				
Ap-4	XY	13/14	28/31	11/11		15/16		11/12				16/17		14/15	9/12	25/26
Vallée Calchaqui																
Cor-1	XY	13/14	29/31,2	10/11	10/10	15/15	6/7	10/12	9/13	19/22	13,2/14	16/16	12/12	17/17	11/11	20/24
Cor-2	XY	14/14	29/31,2	11/11	12/12	15/15	6/6	9/12	10/10	19/22	13,2/14	16/17	8/12	15/17	9/11	20/24
Cor-3	XX	13/14	31,2/31,2	11/11	10/12	15/15	7/7	9/12	9/10	22/23	13,2/15	16/17	12/12	17/18	9/11	24/25
FA-1	XY	13/15	29/32,2	10/12		15/15	6/9,3	13/14	9/10	23/23	12,2/14	16/16	8/8	13/14	7/13	19/20
Sc-1	XX	12/12				17/17	7/7	12/12			13,2/13,2	16/16				24/25
Sc-2	XY	12/14	29/30			15/17	7/7	11/13	11/11	17/17	12,2/13,2	17/18	11/12	14/14	9/11	24/27
Sc-3	XY	13/13				15/15	7/7				13/14	15/15			10/10	
Sc-5	XY	13/14	32,2/32,2		9/9	15/17	7/7		10/10		14/16	16/16	8/8		9/11	26/26
Sc-8	XY	12/14	29/30	10/12	12/12	15/17	7/7	11/13	11/11	17/19	12,2/13,2	17/18	11/12	14/16	9/11	24/27

	AMEL	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Quebrada de Humahuaca																
BP-1	XX	13/13	28/31,2	10/10	10/10	15/15	7/9,3	9/9	9/12	17/22	14/15	16/17	8/12	13/17	11/12	21/21
Hu	XY	13/13	29/31,2	10/11	11/12	15/15	7/9,3	10/12	9/10	19/23	13/13,2	16/18	8/8	13/13	11/12	19/25
Ju-1	XX	15/15	30/32,2	10/13	11/12	15/17	7/7	11/11	9/11	17/20	15/16	14/16	11/11	14/17	11/12	21/25
Ju-4	XY	12/17	30/30	11/12	12/12	15/16	7/7	11/12	9/9	22/23	14/14	16/18		13/13	11/11	23/27
Ju-5	XY	11/12	32,2/32,2	10/10	11/12	15/15	7/7	11/14	9/9	22/23	14/15	16/16	8/8	17/19	7/11	19/24
LA-1	XX										12,2/14,2					
LA-2			28/28							20/20				13/13		23/23
LA-3	XY	13/13	30/30	10/12	12/12	15/16	7/7	12/13	11/12	17/19	12,2/15	17/21	8/11	13/16	11/11	19/26
LA-4	XY	12/13	31,2/34,2	10/11	11/14	15/17	7/7	9/11	9/9	17/22	14/15	17/19	8/12	13/17	7/13	23/26
LA-6	XX				10/12			10/10		19/19				13/13		
LA-7	XY	13/15				15/15	7/7	10/14	9/9		14/15	16/18	8/8		11/12	25/25
LA-8	XX	12/15	30/31,2	10/11	10/14	15/17	7/7	11/13	9/9	22/23	12,2/15	17/19	8/8	13/13	11/13	25/26
LA-9	XY	12/15	30/31,2	11/12	10/12	15/15	7/7	13/13	9/10	22/23	13/13	16/16	8/8	12/13	11/13	19/25
LA-10	XY	13/13	31,2/31,2	10/10	10/11	15/16	7/7	12/13	10/12	19/20	12,2/15	17/21	8/8	13/16	11/13	19/21
LA-12	XY	13/14	31,2/31,2	10/11	11/11	15/15	7/7	12/13	9/10	17/17	13,2/15	15/18	8/8	16/16	7/11	19/24
LA-13	XX	13/14	29/31,2	10/10	11/12	14/15	7/7	8/12	9/10	17/21	13,2/14	15/18	8/11	16/17	11/11	19/24
LA-14	XX	11/13	31,2/32,2	10/11	10/11	15/15	7/7	13/13	11/12	17/20	12,2/12,2	16/17	11/11	13/14	9/13	25/25
LA-16	XY	12/15	30/31,2	11/12	10/12	15/15	7/7	13/13	9/10	22/23	13/13	16/16	8/8	12/13	11/13	19/25
LA-17	XY	14/15	30/31	10/10	10/10	15/15	7/7	9/13	9/12	20/23	13,2/14	16/19	14/14	14/17	11/11	22/24
LA-18	XX	15/15	29/31,2	10/11	11/12	17/18	7/9,3	10/10	9/9	19/19	14/15,2	16/16	8/11	12/17	7/11	25/26
LA-20	XY	12/14	30/32,2	11/12	10/12	15/17	7/9,3	13/13	9/9	19/23	13/15	16/18	8/11	12/13	12/13	25/26
LA-21	XY	12/13	30/31,2	11/13	11/12	15/16	7/7	9/11	9/12	20/22	12,2/13	16/20	8/8	15/16	11/11	23/24
LA-22	XY	13/14	30/30		12/12	15/15	7/7			19/22	12,2/15	16/16		13/13	9/11	26/26
LA-23	XX	13/13	31,2/31,2	10/12	10/11	15/15	6/7	12/13	9/12	17/19	12,2/13	16/17	11/11	13/16	9/11	25/26
LA-24	XY	13/14	31,2/32,2	10/12	10/11	15/15	6/7	9/12	9/9	19/20	13/14	16/16	8/8	13/16	9/11	24/25
Mu-3	XX			11/11	11/11		7/7	9/9	11/12	20/22	14/14	15/15	8/8	15/17		25/25
Mu-4	XY	13/13	31,2/31,2	10/11	11/11			9/12	11/12	20/21		16/17		13/17	12/12	25/28
Mu-5	XY	13/14	31,2/31,2	10/11	11/11	15/16	7/9,3	9/12	11/12	20/21	14/15	15/16	8/12	13/17	11/14	25/28
Mu-7	XY	13/14	31,2/31,2	10/11	11/11	15/16	7/9,3	9/12	11/12	20/21	15/15	15/16	8/12	13/17	11/14	25/28
Mu-8	XX	13/13	31,2/31,2	11/11	11/11	18/18	7/7	9/14	11/12	20/20	15/15			17/17		25/25
SJ	XY	12/14	31,2/31,2		12/12	16/18	7/7	9/11	10/10	17/20	14/14,2	15/16	8/8	13/16	11/11	19/19
Sa-1	XY	13/14	29/32,2	10/11	12/12	15/16	6/6	11/13	9/10	22/23	13,2/15	14/16		12/12	11/12	23/23
Sa-2	XX	12/13	31,2/32,2	10/12	11/11	15/16		9/11	10/10	23/23	13,2/14			17/17	11/12	21/21
Til-1	XY	13/16	30/30	11/11	11/12	15/15	6/9,3	12/13	9/13	19/23	15/15	17/17	8/8	13/15	11/11	24/26
Til-3	XX	14/16	31,2/31,2	11/11	12/12	15/15	7/7	12/12	9/12	17/19	13,2/14	16/16	8/11	13/15	11/11	25/25
Til-4	XY	12/14	31,2/32,2	11/11	10/12	15/16	7/9,3	9/9	10/10	17/21	14/15	17/19	8/12	13/17	13/13	21/24
Til-5	XY	10/13	31,2/31,2	11/12	12/12	15/15	7/7	10/12	9/9	19/23	13,2/14	16/16	8/8	14/17	11/11	22/24
Til-7	XY	10/13	31,2/31,2	11/12	12/12	15/15	7/7	10/12	9/9	19/23	13,2/14	16/16	8/8	14/17	11/11	22/24
F-1	XY	15/15	31,2/32,2	11/11	10/12			9/9								
Hui	XX	13/14	29/32,2	11/11	11/12	15/16	6/7	10/12	10/10	20/20	13,2/15	16/17	8/8	13/17	11/11	19/22

3. Détermination des haplotypes.

3.1. Variations de séquences de la région HVR-1 et distribution des haplotypes anciens.

Une séquence HVR-1 de 359 paires de bases a été obtenue de manière reproductible¹ pour 64 des individus anciens étudiés (Tableau 10). L'alignement des séquences par rapport à la séquence de référence de Cambridge révisée (Anderson *et al.*, 1981; Andrews *et al.*, 1999), a permis de détecter 32 sites nucléotidiques variables, dont la majorité sont des substitutions T – C. Sur ces 32 sites variables, seules trois transversions sont observées, en positions 16111 (C→G), 16182 (A→C) et 16183 (A→C), toutes les autres substitutions étant des transitions. Les séquences sont réparties en 25 haplotypes différents (Tableau 10), dont 13 sont uniques alors que les 12 autres sont partagés par au moins 2 individus (NOA-B1, NOA-B2, NOA-B5, NOA-B9 et NOA-D1) et au maximum 16 individus (NOA-A1).

La détection de mutations caractéristiques des lignées maternelles amérindiennes a permis d'associer chaque haplotype à un haplogroupe fondateur. Ainsi, dans l'échantillon analysé, l'haplogroupe B est le plus représenté (21 individus) et le plus diversifié avec 12 haplotypes différents qui présentent les mutations caractéristiques de cette lignée en positions : 16189 (T→C) et 16217 (T→C). La substitution en position 16189 peut être accompagnée d'un stretch de C qui peut rendre la lecture des résultats difficile entre les positions 16182 et 16194, c'est pourquoi de nombreux auteurs ne considèrent pas ces positions, notamment les mutations ponctuelles en positions 16182 et 16183 qui sont fortement dépendantes de la présence de la transition en position 16189. Cependant, les séquences ayant été obtenues à partir de nombreux extraits et produits PCR différents, et les mutations aux positions 16182, 16183 et 16189 étant facilement identifiables, nous avons conservé ces mutations dans nos résultats.

La deuxième lignée la plus représentée, caractérisée par les mutations en positions 16111 (C→T), 16223 (C→T), 16290 (C→T), 16319 (G→A) et 16362 (T→C), est la lignée A2 avec 19 individus, dont 11 qui proviennent du même site de Los Amarillos dans la Quebrada de Humahuaca. Seize individus, partagés entre 5 haplotypes différents, présentent les mutations caractéristiques de la lignée D1 : 16223 (C→T), 16325 (T→C) et 16362 (T→C). Pour finir,

¹ Les séquences consensus ont été obtenues à partir d'au moins quatre séquences d'extraits et/ou de produit PCR différents.

les 8 individus restants, distribués au sein de 4 haplotypes différents, ont pu être rattachés à l'haplogroupe C1 car ils possèdent les mutations caractéristiques de cette lignée en positions 16223 (C→T), 16298 (T→C), 16325 (T→C) et 16362 (T→C).

La distribution des haplotypes de notre échantillon a été étudiée grâce à une analyse phylogénétique, présentée en figure 21. Les 64 individus sont répartis en quatre groupes différents, correspondant aux quatre lignées maternelles amérindiennes identifiées. Seuls les haplotypes fondateurs des haplogroupes A2, B2 et D1, respectivement NOA-A1, NOA-B10 et NOA-D3, ainsi que l'haplogroupe NOA-C2 sont partagés par des individus provenant de sites localisés dans des aires géographiques différentes. Tous les autres haplotypes sont partagés par des individus exhumés d'un même site archéologique, ou de sites de la même zone.

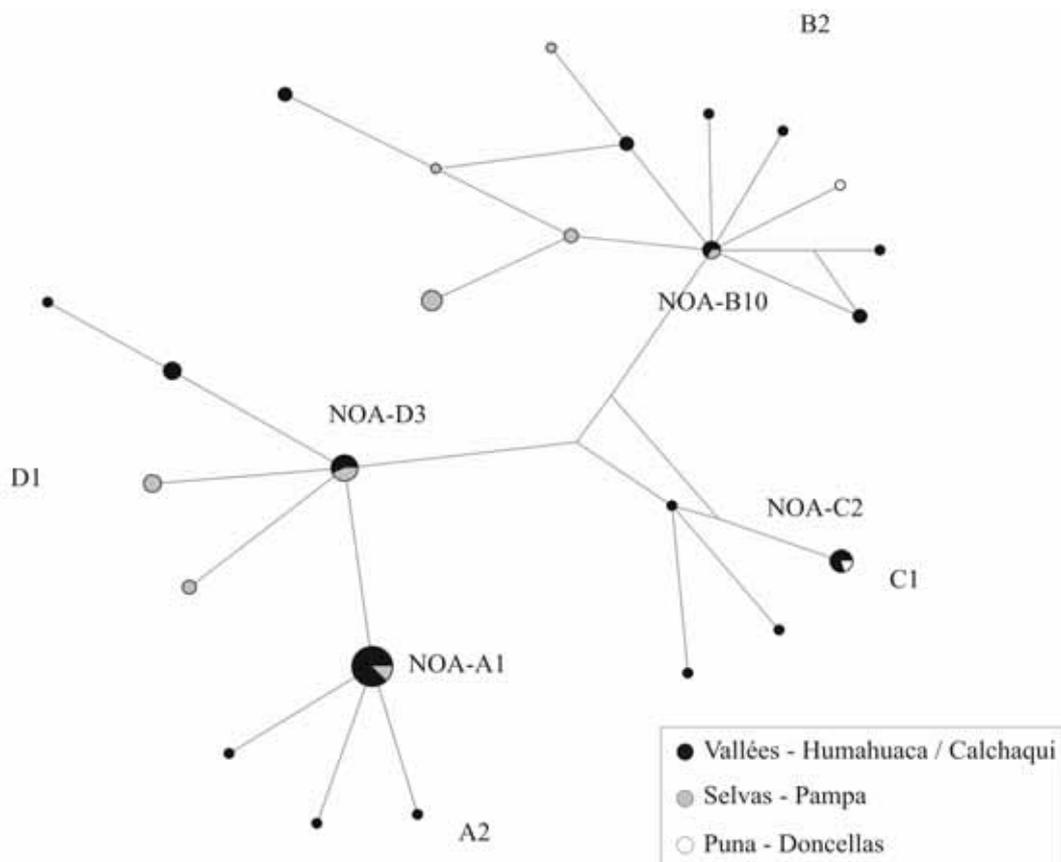


Figure 21 : Analyse phylogénétique des 64 séquences HVR-1 obtenues à partir des échantillons anciens analysés.

La taille des cercles est proportionnelle à la fréquence de chaque haplotype dans l'échantillon.

3.2. Y-STR et haplotypes du chromosome Y.

Nous avons pu déterminer les haplotypes de 35 des 44 individus identifiés comme étant masculins. L'ensemble de ces haplotypes est présenté dans le tableau 11. Parmi ces 35 échantillons masculins, 31 haplotypes uniques ont été détectés, et quatre haplotypes sont partagés par deux individus.

Tableau 11 : Haplotypes Y obtenus pour les individus anciens masculins.

Ht	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS437	DYS438	DYS439	GA-H4TA	DYS448	DYS456	DYS458	DYS635	Code échantillon
H1	13	12	28	25	11	14	13	13,2	17	14	11	11	11	20	15	15	22	Li-4
H2	13	12	29	23	10	15	13	15	15	14	11	10		19	14	15		Cor-1
H3	13	13	29	24	10	14	13	14	14	14	11	11	12	20	14	19	22	SC-2
H4	13	13	29	24	11	14	13	14	18	14	11	11	11	20	15	15	22	Li-1
H5	13	13	30	23	9	14	14	13	16	14	11	11	11	20	15	16	22	Til-5
H6	13	13	30	23	10	14	13	14	16	14	11	11	12	20	17	16		LA-24
H7	13	13	30	24	10	14	13	14	18	14	11	11	11	20	15	17	23	Ju-4
H8	13	13	30	25	10	11	13	13	18	14	11	11	11	19	14	14	22	Ap-4
H9	13	13	30	25	10	14	12	14	17	14	12	12	12		15	15	22	Ju-5
H10	13	13	30	25	10	14	13	14	14	14	11	12	13	20	15	20	22	FA-1, LA-4
H11	13	14	31	25	10	14	12	15	17	14	11	11	11	20	15	17	22	LA-21
H12	13	14	32	23	9	14	13	13	18	14	11	11	12	20	15	15	22	LA-3, LA-10
H13	13	14	32	24	11	14	13	15	16	14	11	11	12	20	15	17	22	LA-9, LA-20
H14	13	14	33	25	10	14	13	14	19	14	11	11	11	20	16	17	23	Til-1
H15	14	13	31	25	10	14	13	15	16	14	11	12	13		15	17	22	C.II-1
H16	14	14	32	25	10	14	13	15	16	14	11	12	13		15	17		C.V-2
H17*	13	14	32	26	10	14	13	14	17	14	11	11	11	19	16	17	23	Mu-4, Mu-5
H18		13	29	24	10	14	13	15	18	14	11							Lam-4
H19		14	33	25	10	14	13	14	18	14		11	11		16	18		LA-7
H20	13	13		25	10	14	12	14	15	14	12	13	12	20	16	15	22	Hu
H21	13	13	30	23	10		13	14	17	14	11		10		15	15	22	Til-4
H22	14	13	29	24	11		13	15	17	14	11	12	11	19	15	14	22	Ap-2
H23	14	13	29	24	11		13	15	18	14	11	13	11		14	15	22	Lam-2
H24	14	14	30	24	10		13	16	17	14	11	11	11	19	16	17	22	LA-17
H25	14	14	32	24	10		13	15	16	14	11	12	13		15	18		Li-3
H26		13	29	24	11		13	14	18	14	11	11	11	19	15	15	22	Lam-3
H27	12	13	31		10		13	14	14	14		11	11		16	16		SJ
H28	14	13			11		13	15	17	14	11	12	11	19	15	14		C.V-3
H29	13	12		23			13	15	15	14		10		19	14	15		Cor-2
H30	13	12		24	10		13	15	15	14		10	12		14	15		SC-5
H31	13	13	29		10		13	16	17	14	11	11	11		17	15	22	Sa-1
H32		13					9	14	15						15	16		Don-1
H33		13		25	10		13	14	14	14	11		13		15	20	22	LA-12
H34		13			10		13	13	18	13			12		16			LA-22

* Les résultats obtenus pour les échantillons de Muyuna sont présentés mais n'ont pas été pris en compte pour les analyses ultérieures.

Chacun des haplotypes ainsi détecté a été analysé grâce à la base de données en ligne Haplogroup Predictor (<http://www.hprg.com/hapest5/>) afin d'essayer de déduire l'haplogroupe auxquels ils correspondent. Par la suite, les haplotypes H1 à H16 ont été comparés à la base de données disponible en ligne Y-HRD, sur la base de sept marqueurs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393). Les résultats de ces analyses sont présentés dans le tableau 12 suivant.

Tableau 12 : Résultat de la déduction des haplogroupes et de la comparaison avec la base de données en ligne Y-HRD.

HG	%	Y-HRD		
		Native américaine	Métissée	Eurasiatique / Européenne / Africaine
H1	Q 99,9	-	-	-
H2	Q 94,6		Quindío (Colombie)	Ecuador, (African American)
H3	Q 99,8	Mapuche (Argentine); Quichua (Equateur); Zoé (Brésil)	Antioquia, Bogota (Colombie); Pérou; Quito (Equateur); Sao Paulo, Santarem, Manaus, Belem (Bresil)	Misiones, Buenos Aires, Cordoba, Salta (Argentine)
H4	Q 99,9		Sao Paulo, Alagoas, Manaus, Belem, Parana (Bresil); Pérou; Santander, Cundinamarca, Bogota (Colombie)	Mendoza, Santa Fe (Argentine); Sao Paulo, Rio Grande do Sul (Bresil)
H5	Q 99,5			Rio Negro (Argentine)
H6	Q 99,4	Beni (Bolivie); Tiriyo, Yanomami (Bresil); Quichua (Equateur); Toba (Argentine)	Pérou; Cesar, Bogota, Atlantico, Santander, Caldas, Antioquia (Colombie); Belem, Parana (Bresil); Quito (Equateur)	Cordoba, Mendoza (Argentine)
H7	Q 100	Toba, Pilaga, Coya, Diaguaita, Wichi (Argentine); Awa-Guaja, Urubu-Kaapor, Zoé (Bresil); Quichua (Equateur)	Pérou; Belem, Bahia, Santarem, Manaus (Bresil); Antioquia, Narino, Cordoba, Valle de Cauca, Cauca (Colombie); Quito (Equateur); Caracas (Venezuela)	Mendoza, Buenos Aires, Cordoba (Argentine); Rio Grande do Sul (Bresil)
H8	n.d.		Goias, Maranhao, Minas Gerais, Rio Grande do Sul (Bresil); Antioquia (Colombie)	Rio de Janeiro, Sao Paulo, Rio Grande do Sul (Bresil); Mendoza, Buenos Aires, Cordoba (Argentine)
H9	Q 100	Beni (Bolivie)	Santarém (Bresil)	
H10	Q 100	-	-	-
H11	Q 100	-	-	-
H12	Q 99,5	-	-	-
H13	Q 97,8	Pilaga, (Argentine); Quichua (Equateur)	Perou; Antioquia, Nariño, (Colombie)	
H14	Q 100			Chaco (Argentine)
H15	Q 99,9	Quichua (Equateur)	Quito (Equateur)	
H16	Q 99,5	-	-	-

4. Détermination des haplogroupes par spectrométrie de masse MALDI-TOF.

Le typage simultané de 15 SNP mitochondriaux et du chromosome Y a permis d'obtenir des résultats reproductibles et concordants pour chacun des différents extraits d'ADN¹. Grâce à l'analyse combinée des allèles amplifiés nous avons pu déterminer les haplogroupes mitochondriaux de 70 échantillons anciens et les haplogroupes Y de 37 des 44 individus masculins, comme cela est présenté dans le tableau 13.

Les haplogroupes mitochondriaux ainsi déterminés sont concordants avec la déduction faite à partir du séquençage de la région HV-1, dans 100 % des cas. De plus, l'étude de polymorphismes de la région codante de l'ADNmt a permis la détermination des lignées maternelles pour 10 échantillons (BP-2, LA-1, LA-2, LA-6, Til-2, Til-6, FA-2, C.II-2, Lam-1 et Don-3), pour lesquels aucun résultat exploitable n'avait été obtenu à partir du séquençage classique. Enfin, en plus de la confirmation de la détermination des lignées, cette analyse a également permis d'affiner certaines de ces déterminations. Ainsi, il apparaît que l'ensemble des échantillons ayant été rattachés à l'haplogroupe C1, présentent une mutation en position 493 (A→G), caractéristique du sous-haplogroupe C1b. En ce qui concerne les échantillons identifiés comme appartenant à la lignée maternelle A2, ils possèdent tous la mutation caractéristique de cette lignée (position 12007, G→A), et mis à part les deux échantillons LA-7 et C.V-4, ils possèdent une mutation supplémentaire en position 64 (C→T), qui permet de dire qu'ils appartiennent à un sous-haplogroupe de la lignée A2, allant de A2c à A2r. Les échantillons LA-7 et C.V-4 qui ne possèdent pas cette mutation appartiennent probablement à l'haplogroupe ancestral A2. Il est intéressant de noter que la présence ou non de cette mutation 64T a également permis de montrer que malgré le partage d'un même haplotype des individus pouvaient appartenir à une lignée différente. En effet, l'échantillon C.V-4 présente le même haplotype que 15 autres individus : 16111T – 16223T – 16290T – 16319A – 16362C, associé à la lignée A2. Cependant, il ne possède pas de mutation en position 64, et appartient donc à une autre lignée, contrairement aux 15 autres échantillons (mutés en position 64) qui appartiennent à un des sous-haplogroupes A2c – A2r.

¹ L'ensemble des génotypes consensus obtenus pour chacun des individus sont présentés en annexe 4 (p.235-236).

Tableau 13 : Haplogroupes déterminés par l'analyse de SNP des régions codantes de l'ADN mitochondrial et de la portion non recombinante du chromosome Y, par spectrométrie de masse MALDI-TOF.

ID	Sexe	Hg-HV1	Hg-SNP	Hg Y	ID	Sexe	Hg-HV1	Hg-SNP	Hg Y
Quebrada de Humahuaca					Vallée Calchaqui				
BP-1	XX	A2	A2 c-r		Cor-1	XY	C1	C1b	Q1a3*
BP-2*	n.d.		C1b		Cor-2	XY	C1		
Hu	XY	D1	D1	Q1a3a*	Cor-3	XX	C1	C1b	
Ju-1	XX	B2	B2		FA-1	XY	A2	A2 c-r	Q1a3a*
Ju-4	XY	D1	D1	Q1a3a*	FA-2	n.d.		A2 c-r	
Ju-5	XY	C1	C1b	Q1a3a*	SC-1	XX	C1	C1b	
LA-1	XX		A2 c-r		SC-2	XY	B2	B2	Q1a3a*
LA-2	n.d.		A2 c-r		SC-3	XY	A2	A2 c-r	
LA-3	XY	A2	A2 c-r	Q1a3a*	SC-5	XY	B2	B2	Q1a3*
LA-4	XY	D1	D1	Q1a3a*	Pampa Grande				
LA-6	XX		A2 c-r		Li-1	XY	B2	B2	Q1a3*
LA-7	XY	A2	A2	Q1a3a*	Li-2	XY	D1	D1	
LA-8	XX	A2	A2 c-r		Li-3	XY	D1	D1	Q1a3a*
LA-9	XY	D1	D1	Q1a3a*	Li-4	XY	D1	D1	Q1a3a*
LA-10	XY	A2	A2 c-r	Q1a3a*	C.II-1	XY	B2	B2	Q1a3a*
LA-12	XY	A2	A2 c-r	Q1a3a*	C.II-2	XY		B2	
LA-13	XX	A2	A2 c-r		C.II-3	XX	B2	B2	
LA-14	XX	A2	A2 c-r		C.II-4	XX	B2	B2	
LA-17	XY	B2	B2	Q1a3*	C.V-1	XY	D1	D1	Q1a3
LA-18	XX	C1	C1b		C.V-2	XY	B2	B2	Q1a3a*
LA-20	XY	A2	A2 c-r	Q1a3a*	C.V-3	XY	B2	B2	Q1a3*
LA-21	XY	A2	A2 c-r	Q1a3a*	C.V-4	XX	A2	A2	
LA-22	XY	A2	A2 c-r	Q1a3*	Lam-1	n.d.		A2 c-r	
LA-23	XX	A2	A2 c-r		Lam-2	XY	D1	D1	Q1a3*
LA-24	XY	A2	A2 c-r	Q1a3a*	Lam-3	XY	A2	A2 c-r	Q1a3*
SJ	XY	C1	C1b	Q1a3a*	Lam-4	XY	D1	D1	Q1a3
Sa-1	XY	B2	B2	Q1a3*	Lam-5	XX	B2	B2	
Sa-2	XX	B2	B2		Ap-1	XX	D1	D1	
Til-1	XY	B2	B2	Q1a3a*	Ap-2	XY	B2	B2	Q1a3*
Til-2	n.d.		D1		Ap-3	XX	B2	B2	
Til-3	XX	B2	B2		Ap-4	XY	D1	D1	Q1a3*
Til-4	XY	A2	A2 c-r	Q1a3a*	Doncellas				
Til-5	XY	B2	B2	Q1a3a*	Don-1	XY	B2	B2	Q1a3a*
Til-6	n.d.		B2		Don-2	XY	C1	C1b	
F-1	XY	B2	B2	Q1a3	Don-3	n.d.		B2	
F-2	n.d.	B2	B2						
Hui	XX	A2	A2 c-r						

* Les échantillons en gras sont ceux pour lesquels seuls le génotypage a permis de déterminer les haplogroupes.

Concernant les lignées paternelles, l'étude de ces SNP a permis d'affiner au maximum la détermination des haplogroupes. Ainsi, il est possible de dire de manière certaine que 23 individus présentant l'allèle muté pour les marqueurs M346 (A → G) et M3 (C → T) et l'allèle ancestral aux marqueurs M19, M194, M199 et SA01, appartiennent au paragroupe Q1a3a* (aujourd'hui appelée Q1a3a1*). Le génotype M346G - M3C partagé par 11 individus indique qu'ils se situent au sein du paragroupe Q1a3*. Pour trois des individus (F-1, C.V-1 et Lam-4), l'absence d'amplification de certains des marqueurs ne permet pas de les rattacher de manière fiable à une lignée paternelle et permet seulement de les inclure dans les lignées Q1a3, sans plus de précision, car ils présentent les allèles mutés pour les marqueurs M242 et M346.

5. Authenticité des résultats.

L'authentification des résultats est une étape cruciale pour toute étude portant sur l'ADNa. L'ensemble des critères suivis, tout au long du processus d'analyse, nous permettent de garantir l'authenticité des résultats présentés dans ce travail, même si nous n'avons pas suivi l'ensemble des 10 critères proposés classiquement pour l'authentification.

- L'ensemble des résultats (profils génétiques, séquences HVR-1 et typages de SNP et STR) a été obtenu après plusieurs répliques indépendantes des analyses à partir d'extraits et d'amplifications différents. Seuls les échantillons pour lesquels les résultats des différentes répliques se sont avérés concordants ont été conservés pour les analyses.
- Les échantillons pour lesquels les doutes d'une contamination croisée entre échantillons étaient trop forts ont été éliminés pour les analyses statistiques (échantillons du site de Muyuna). De plus, tous les extraits ou amplifications pour lesquels les témoins négatifs ou blancs d'extractions présentaient des signes de contaminations ont été retirés de toutes les analyses.
- Une attention particulière a été portée à ce que les résultats obtenus à partir de marqueurs différents soient concordants. Une concordance de 100 % a été obtenue pour la détermination des lignées à partir des deux méthodes différentes utilisées (génotypage et

séquençage). De plus, la cohérence des résultats avec la région géographique étudiée a été vérifiée. Aussi bien les lignées maternelles que paternelles décrites, sont toutes bien spécifiques des populations natives américaines.

- La comparaison des profils génétiques des échantillons anciens avec ceux des manipulateurs confirme l'absence de contamination par un manipulateur ayant eu accès aux échantillons récemment.

6. Recrutement funéraire et détermination des liens de parenté.

La détermination de liens de parenté potentiels entre individus anciens, inhumés au sein d'un même ensemble funéraire, est une problématique qui intéresse depuis longtemps les anthropologues et archéologues étant donné qu'elle permet : (1) de mieux comprendre le recrutement funéraire (identification des individus inhumés, détermination de liens génétiques et détermination du sexe), (2) de mieux appréhender les pratiques funéraires et sociales des populations du passé.

Afin de comprendre au mieux les liens unissant les individus anciens de différents sites, l'analyse combinée de plusieurs marqueurs, situés sur des portions différentes du génome, a été utilisée : des marqueurs à transmission uni-parentale permettant de définir les haplotypes mitochondriaux et Y des individus, mais aussi des marqueurs à transmission biparentale permettant de déterminer les profils génétiques de ces mêmes individus. Il arrive cependant que même avec tous ces différents marqueurs, le type de parenté entre les individus ne puisse pas être défini. C'est pour cela que de nouveaux marqueurs pouvant être utiles pour ces problématiques ont été recherchés et étudiés au cours de ce travail, telles que les petites insertions et délétions.

6.1. Recrutement funéraire du site de Pampa Grande.

Le site archéologique de Pampa Grande regroupe plusieurs cavités sépulcrales dont l'occupation n'a semble-t-il pas dépassé quelques centaines d'années, à la fin de la période céramique initiale (Baffi *et al.*, 1996; Salaberry *et al.*, 1999). Les études anthropologiques

menées sur les individus exhumés des cinq cavités étudiées au cours de ce travail montrent qu'il n'y avait apparemment pas d'inhumations préférentielles ; c'est-à-dire que les restes squelettiques identifiés appartenaient aussi bien à des individus matures qu'immatures, de sexe féminin ou masculin (Salaberry *et al.*, 1999). Le recrutement funéraire semble donc se rapprocher de ce que l'on peut appeler une « population normale » (individus des deux sexes et de tout âge). La question des liens unissant ces individus anciens s'est donc posée et des investigations en ce sens ont été menées.

Dans un premier temps, les individus présentant un haplotype mitochondrial et/ou Y identique ont été regroupés puis analysés plus en détail. Le tableau 14 suivant résume de manière plus synthétique les couples d'échantillons pour lesquels des liens de parenté pouvaient être supposés au vue des haplotypes. Même si de nombreux individus partagent la même lignée maternelle, aucun lien de proche parenté n'a pu être mis en évidence. En effet, même si les profils génétiques établis sont pour la plupart partiels, aucune paire de profils comparés deux à deux ne partage un allèle en commun pour chacun des marqueurs amplifiés. Il est donc évident que parmi les individus que nous avons pu analyser aucune relation de type parent/enfant n'est avérée. Les deux individus Ap-2 et C.V-3 possèdent un haplotype Y apparemment identique (les profils n'étant pas complets, il apparaît que les 12 marqueurs qui ont pu être amplifiés sont identiques) ce qui laisse imaginer une relation de type père/fils ou frères. Cependant, ces deux hypothèses peuvent être rejetées car ils ne possèdent pas le même haplotype mitochondrial (excluant une relation de frères), et ne partagent pas d'allèles en commun pour tous les marqueurs qui ont pu être amplifiés (excluant une relation de type père/fils).

Des liens de type fratrie ou demi-frère/demi-sœur ont été testés pour tous les individus à l'aide de différents logiciels (DNA•VIEWTM, ML-Relate). Aucune relation de ce type ne s'est révélée être significativement probable.

Tableau 14 : Paires d'échantillons présentant un haplotype mitochondrial et/ou Y identique pour le site de Pampa Grande.

	Ht-Mt	Ht-Y	nombre d'allèles partagés (STRa)
Li-2 / Li-4	Identique	/	6/8
C.II-1 / C.II-4	Identique	/	5/9
C.V-2 / C.V-3	Identique	Différent	5/7
Lam-2 / Lam-4	Identique	Différent	5/7
Ap-1 / Ap-4	Identique	/	5/8
Li-1 / Ap-3	Identique	/	4/8
Li-2 / C.V-1	Identique	/	
Li-3 / Ap-1	Identique	/	4/8
Li-3 / Ap-4	Identique	Différent	5/8
Li-4 / C.V-1	Identique	/	
C.II-1 / C.V-2	Identique	Différent	5/8
C.II-1 / C.V-3	Identique	Différent	5/8
C.II-4 / C.V-2	Identique	/	5/8
C.II-4 / C.V-3	Identique	/	6/8
Ap-2 / C.V-3	Différent	Identique (12)	4/8

6.2. Cortaderas.

Une structure funéraire particulière a été étudiée au sein du site de Cortaderas, dans la vallée Calchaqui, datant de la période Inca. L'étude de cette sépulture, au sein de laquelle trois jeunes individus matures ont été inhumés de manière simultanée, a conduit les archéologues à émettre une hypothèse de lien de parenté entre les trois individus, sans pour autant qu'ils puissent l'affirmer avec certitude.

De façon similaire au cas précédent, plusieurs marqueurs ont été analysés de manière combinée afin d'obtenir un pouvoir de discrimination suffisant. Ainsi, 30 petites insertions/délétions¹ ont été analysées afin d'essayer de trancher entre plusieurs hypothèses émises sur la base de l'analyse des STR autosomaux.

¹ L'ensemble des profils génétiques consensus obtenus à partir de l'amplification des insertions et délétions sont présentés en annexe 5 (p.237).

Tableau 15 : Résultats des différents marqueurs analysés pour les trois individus de Cortaderas, et valeurs des LR et pourcentages des différentes hypothèses de parenté testées.

	Nombre de marqueurs pour lesquels 1 allèle est partagé				LR / P (%)	
	Ht-Mt	Ht-Y	STRa	Indels	Fratrie	Parent/enfant
Cor-1 / Cor-2	Identique	Identique (10)	13/15	30/30	592 / 99,83	0,211
Cor-1 / Cor-3	Identique	/	15/15	30/30	695 / 99,86	6180 / 99,98
Cor-2 / Cor-3	Identique	/	14/15	30/30	40,3.10 ⁶ / 99,99	250.10 ⁶ / 99,99
Cor-1/Cor-2/Cor-3					111.10 ³ / 99,99	717.10 ³ / 99,99

Il semble évident à la vue des résultats des différents marqueurs analysés, présentés dans le tableau 15, que ces trois squelettes anciens partagent un lien de très proche parenté. Ces trois individus possèdent le même haplotype mitochondrial, et les deux individus masculins (Cor-1 et Cor-2) partagent probablement le même haplotype Y (seuls 10 STR-Y ont pu être amplifiés pour l'échantillon Cor-2, mais qui sont tous identiques à ceux de Cor-1). A partir de ces premières données deux hypothèses ont été émises : (1) les trois individus pouvaient appartenir à une même fratrie ; (2) une relation de type parent/enfant aurait également pu être cohérente avec les données obtenues.

Pour tester ces différentes hypothèses nous nous sommes donc intéressés à des marqueurs nucléaires transmis par les deux parents comme les STR autosomaux mais aussi les indels. A la vue des résultats il apparaît beaucoup plus probable que les deux individus masculins soient frères (99.83 %) plutôt que père/fils. En ce qui concerne l'individu féminin (Cor-3), il n'est pas possible de discriminer les deux hypothèses à partir de l'analyse des STR autosomaux, même si des probabilités légèrement supérieures sont obtenues pour un lien de type mère/fils. Pour essayer de trancher entre ces deux hypothèses nous avons donc analysé de petites insertions/délétions, également transmises aux enfants par les deux parents. Malheureusement, ces marqueurs n'ont permis ni d'exclure ni de favoriser une des deux hypothèses. Les trois individus possèdent un allèle en commun pour chacun des 30 indels analysés, indiquant un probable lien de proche parenté, mais ne permettant pas de différencier un lien de type parent/enfant ou frères/sœur.

Les indels semblent très intéressants pour l'identification individuelle à partir d'échantillons

d'ADN dégradé, car ils peuvent être analysés au sein de courts fragments d'ADN, mais ne sont apparemment pas assez polymorphes pour obtenir un pouvoir de discrimination suffisant pour la détermination de liens de parenté (Hollard *et al.*, in press).

6.3. Los Amarillos.

Le site archéologique de Los Amarillos est une des agglomérations préhispaniques les plus importantes de la Quebrada de Humahuaca. Des échantillons correspondant à 23 individus anciens, ayant occupé ce site pendant la fin de la période céramique tardive, ont été prélevés, puis analysés au cours de cette étude, entre autre dans le but de déterminer si les liens de parenté étaient pris en compte lors des inhumations.

La grande majorité des activités de la population ancienne peuplant l'agglomération de Los Amarillos étaient réalisées au sein de la sphère domestique. Ainsi, les deux unités domestiques étudiées dans le cadre de ce travail ont mis au jour un grand nombre de vestiges anciens, aussi bien d'usage domestique, de production ou liés aux pratiques funéraires. En effet, au sein des unités 320 et 400, plusieurs structures funéraires pouvant s'apparenter à des cistes ont été mises au jour. Dans ces structures, plusieurs évènements d'inhumation se sont succédé conduisant à l'accumulation de plus d'une dizaine d'individus anciens. Deux des trois cistes analysées ont été créées et utilisées à une période où les unités domestiques avaient été abandonnées comme lieu d'habitation, il est donc légitime de se demander si les individus inhumés au sein de ces structures appartenaient à une même lignée, ou si des individus provenant de tout le site ont été inhumés à cet endroit, qui aurait pu être utilisé comme petite nécropole à un moment donné.

6.3.1. Unité 320.

Parmi les 8 individus analysés au sein de cette structure, nous avons pu mettre en évidence des liens de proche parenté pour 4 d'entre eux, comme présenté dans le tableau 16. En effet, les individus LA-4 et LA-8 présentent des profils génétiques avec un allèle partagé pour les 15 STR autosomaux analysés, ce qui est le cas également pour les individus LA-9 et LA-20. Ceci implique une forte probabilité pour que ces individus présentent un lien de type parent-

enfant. Ainsi, LA-4 et LA-8 n'ayant pas le même haplotype mitochondrial, l'hypothèse privilégiée est celle d'une relation père (LA-4) / fille (LA-8), confirmée par les analyses faites grâce au logiciel DNA•VIEW™: probabilité de plus de 99.99 %. En ce qui concerne les individus masculins LA-9 et LA-20, qui possèdent un haplotype Y identique mais un haplotype mitochondrial différent, il existe également une très forte probabilité pour qu'ils partagent une relation de type père/fils (99.97 %), sans qu'il soit pour autant possible de déterminer lequel de ces deux individus était le père.

Ces relations sont confirmées par les indels analysés, car comme pour les STR autosomaux, LA-9 et LA-20 partagent un allèle en commun pour les 30 marqueurs, alors que LA-4 et LA-8 partagent un allèle en commun pour les 16 marqueurs qui ont pu être amplifiés de manière reproductible (annexe 5).

Tableau 16 : LR (likelihood ratio) et probabilités calculées pour les liens de parentés supposés entre les individus inhumés au sein de l'unité domestique 320 du site de Los Amarillos.

	Lien supposé	Ht-Mt	Ht-Y	N allèles partagés (STRa)	LR	P (%)
LA-4/LA-8	parent/enfant	Différent	/	15/15	130000	99,999
LA-9/LA-20	parent/enfant	Différent	identique	15/15	4200	99,976
LA-9/LA-8	Fratrie	Différent	/	13/15	212	99,528
LA-20/LA-8	Fratrie	Identique	/	14/15	35	97,142

L'analyse réalisée à partir des profils génétiques (STRa) a également révélé de fortes probabilités quant à des liens de type frère/sœur entre les individus LA-8 et LA-9 ainsi que LA-8 et LA-20. Cependant, comme nous avons montré que LA-9 et LA-20 sont très probablement père/fils, il apparaît impossible que LA-8 soit la sœur de ces deux individus. De plus, LA-9 et LA-20 possèdent un haplotype Y différent de celui de LA-4, confirmant que LA-9 et LA-20 ne peuvent être les frères de LA-8. Des relations plus éloignées (demi-frère/demi-sœur, neveu...) sont donc à rechercher.

LA-9 possède un haplotype mitochondrial différent de LA-8, le lien entre ces deux individus reste donc difficile à établir. Par contre, LA-20 possède lui un haplotype mitochondrial identique à celui de LA-8, il est donc possible qu'ils soient demi-frère (LA-20) / demi-sœur

(LA-8). Les liens probables unissant les individus anciens de cette structure funéraire sont illustrés sur la figure 22 suivante.

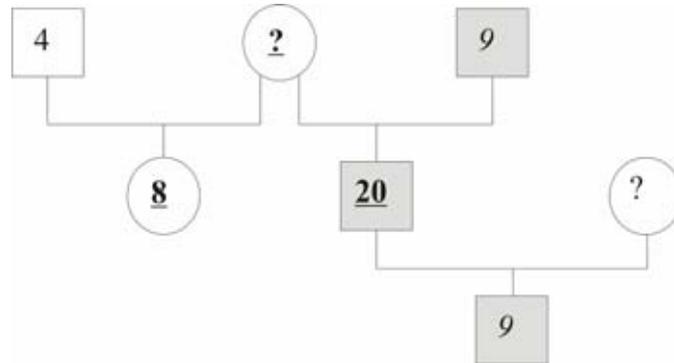


Figure 22 : Représentation schématique des liens de parenté supposés au sein de l'unité domestique 320 du site archéologique de Los Amarillos.

6.3.2. Unité 400.

Des liens de proche parenté ont pu être mis en évidence parmi les 11 individus (LA-2, LA-3, LA-10, LA-11, LA-12, LA-13, LA-14, LA-15, LA-22, LA-23, LA-24) exhumés de l'unité domestique 400 du site de Los Amarillos.

L'ensemble de ces individus partage la même lignée maternelle (A2) mais surtout le même haplotype mitochondrial, comme présenté dans le tableau 17. Même si il s'agit de l'haplotype fondateur de la lignée A2, distribué dans toute l'Amérique, il n'est pas aussi fréquent habituellement dans les populations andines. Cette forte représentation au sein d'une si petite structure laisse supposer des relations de parenté étroites entre les individus.

A la vue des différents marqueurs analysés, nous pouvons clairement distinguer certains liens de type parent/enfant, notamment entre LA-13 (mère) et LA-12 (fils), qui partagent un allèle en commun pour les 15 STR autosomaux et 29 des 30 indels analysés, et entre LA-14 et LA-23 (mère/fille), qui partagent un allèle en commun pour tous les marqueurs autosomaux amplifiés (45). Un lien de type frère/frère semble également évident (99.99 %) entre les échantillons LA-3 et LA-10 qui présentent des haplotypes mitochondriaux et Y identiques.

Tableau 17 : LR (likelihood ratio) et probabilités calculées pour les liens de parentés supposés entre les individus inhumés au sein de l'unité domestiques 400 du site de Los Amarillos.

	Lien supposé	Mt Ht	Y Ht	N allèles partagés (STRa)	LR	P (%)
LA-3/LA-10	Frères	identique	identique	13/15	19000	99,995
LA-3/LA-23	Fratrie	identique	/	13/15	359	99,721
LA-3/LA-22	Fratrie	identique	différent	10/11	232	99,569
LA-12/LA-13	parent/enfant	identique	/	15/15	18500	99,995
LA-14/LA-23	parent/enfant	identique	/	15/15	33400	99,997
LA-23/LA-24	Fratrie	identique	/	14/15	535	99,813

D'autres liens sont également supposés après l'analyse des profils génétiques par le logiciel DNA•VIEW™. Cependant, comme en ce qui concerne l'unité 320, ces liens qui semblent être plus éloignés sont plus difficiles à détecter et confirmer. Pour exemple, l'individu LA-23 semble présenter un lien de type frère/sœur avec LA-3, cependant la probabilité que LA-23 soit la sœur de LA-10 est quasiment nulle. LA-23 pourrait être la demi-sœur de LA-3, mais en n'ayant pas accès aux profils génétiques des parents de ces individus, cette hypothèse est très difficile à confirmer. Dans le cas où LA-23 serait la demi-sœur de LA-3, il serait plus probable que LA-14 soit la fille de LA-23, car aucun lien de proche parenté n'a été détecté entre LA-14 et LA-3 ou LA-10.

7. Constitution des populations archéologiques.

La détermination des liens de proche parenté au sein de groupes anciens est une étape indispensable permettant de constituer des groupes d'individus, sans redondance des informations. Ainsi, les liens que nous avons pu mettre en évidence dans les différents groupes étudiés vont nous permettre de constituer des « populations archéologiques » que nous pourrions par la suite étudier de manière intra-populationnelle mais aussi inter-populationnelle.

7.1. Population des vallées tempérées.

L'ensemble des échantillons, issus de l'aire écologique appelée vallées tempérées a été

regroupé au sein d'une même population (que nous appellerons vallées dans la suite de ce travail) qui comprend donc les échantillons issus de la vallée Calchaqui ainsi que les échantillons de la Quebrada de Humahuaca. L'histoire de ces deux vallées a semble t-il suivi la même trajectoire, de plus, l'occupation des différents sites de ces deux vallées est contemporaine, ce qui permet de constituer un groupe relativement homogène.

Les échantillons Cor-2 et Cor-3 du site de Cortaderas, LA-3, LA-13 et LA-23 du site de Los Amarillos qui partagent des liens de proche parenté avec d'autres échantillons de ces sites ont été retirés de l'échantillon pour les analyses statistiques. De plus, pour les analyses concernant la portion non recombinante du chromosome Y, en plus des échantillons précités, l'échantillon LA-9 du site de Los Amarillos a été retiré. Ainsi cette population se compose de 41 individus (34 de la Quebrada de Humahuaca et 7 de la vallée Calchaqui), dont 11 individus féminins, 24 individus masculins (seulement 22 seront pris en compte pour les analyses) et 6 individus dont le sexe n'a pas pu être déterminé.

7.2. Population des sierras.

L'aire écologique des sierras n'est représentée que par le site de Pampa Grande. Aucune relation de proche parenté n'ayant été mise en évidence entre les individus étudiés, ils ont tous été conservés pour constituer la population appelée Pampa Grande. Cette population, dont les restes sont datés de la fin de la période céramique initiale (V^{ème} – VIII^{ème} siècles), est donc composée de 21 individus, dont 7 individus féminins et 14 individus masculins.

7.3. Individus de la puna.

Les échantillons issus du site archéologique de Doncellas, situé dans l'aire écologique de la puna, ont été traités à part étant donné qu'il s'agit d'une zone bien différenciée des deux autres, ayant eu sa propre histoire. Ces trois échantillons ne pouvant constituer une population, mis à part la recherche d'haplotypes partagés avec d'autres individus amérindiens, aucune analyse statistique n'a été effectuée à partir de ces échantillons.

II. Diversité génétique des populations des Andes méridionales.

1. Diversité des lignées maternelles et paternelles.

La répartition des lignées maternelles et paternelles déterminées à partir des SNP mitochondriaux et du chromosome Y est présentée dans le tableau 18 suivant, pour chacune des deux populations archéologiques constituées.

Tableau 18 : Pourcentage des lignées maternelles et paternelles détectées grâce à l'analyse des SNP mitochondriaux et du chromosome Y dans les deux populations analysées.

	Hg-Mt (% / N)				Hg-Y (% / N)				
	N	A2	B2	C1b	D1	N	Q1a3*	Q1a3a*	Q1a ?
Vallées	41	43,9 (18)	29,3 (12)	14,6 (6)	12,2 (5)	22	22,7 (5)	72,8 (16)	4,5 (1)
Pampa Grande	21	14,3 (3)	47,6 (10)	0	38,1 (8)	12	50 (6)	33,3 (4)	16,7 (2)

La lignée maternelle A2 est la plus représentée au sein de la population des vallées (43,9 %), alors que pour la population de Pampa Grande l'haplogroupe mitochondrial le plus fréquent est le B2 (47,6 %). Cette lignée B2, souvent présente à des fréquences élevées dans les populations andines, est également assez fréquente dans la population des vallées (29,3 %). Alors que l'haplogroupe D1 est le moins fréquent dans les vallées tempérées (12,2 %), il a été retrouvé à une très forte fréquence dans le site de Pampa Grande (38,1 %). Enfin, la lignée maternelle C1 n'a été détectée pour aucun des individus de Pampa Grande, mais en ce qui concerne les groupes des vallées, elle est retrouvée avec une fréquence non négligeable de 14,6 %.

La lignée paternelle Q1a3a* a été détectée pour plus de 70 % des individus masculins de la population des vallées, les individus restants ayant été rattachés au paragroupe Q1a3* (22,7 %). Pour la population de Pampa Grande, les fréquences observées des lignées paternelles

sont quasiment inversées. En effet, pour ce groupe la lignée la plus fréquente, détectée pour la moitié des individus, est le paragroupe Q1a3*, alors que la lignée Q1a3a* représente 33,3 % de la population. Il faut noter que pour 3 individus (1 des Vallées et 2 de Pampa Grande), nous avons seulement pu dire qu'ils appartenaient à une des lignées Q1a3, sans pouvoir déterminer laquelle. Les fréquences réelles des lignées paternelles dans ces deux populations peuvent donc être légèrement différentes.

Il apparaît donc, à la vue de ces résultats, que la structure de ces deux populations semble différente aussi bien en ce qui concerne les lignées maternelles que paternelles. En effet, les fréquences des lignées amérindiennes sont distribuées de manière différente pour ces deux groupes. Il faut cependant garder en mémoire que l'effectif faible de ces populations peut entraîner un biais significatif.

2. Diversité des haplotypes mitochondriaux et du chromosome Y.

Parmi les 41 individus de la population des Vallées, une séquence HVR-1 a été obtenue pour 34 d'entre eux, et un haplotype Y au moins partiel a été obtenu pour 21 des 24 individus masculins (dont 12 complets pour l'haplotype minimal). Concernant le groupe de Pampa Grande, sur les 21 individus pour lesquels les haplogroupes ont été déterminés, les haplotypes mitochondriaux de 19 d'entre eux ont été déterminés de manière reproductible, ainsi que 11 haplotypes du chromosome Y (dont 5 complets pour l'haplotype minimal).

A partir de ces données haplotypiques, les paramètres classiques de diversité, présentés dans le tableau 19, ont été calculés à l'aide du logiciel Arlequin, pour chacune des deux populations.

En ce qui concerne l'ADNmt, alors que la diversité nucléotidique et le nombre moyen de différences par paires de séquences sont similaires, il apparaît que la diversité haplotypique est plus faible pour la population des Vallées que pour la population de Pampa Grande. Cette faible diversité génétique en ce qui concerne la population des Vallées pourrait être due au nombre important d'individus présentant le même haplotype NOA-A1 (11 dont 8 du site de Los Amarillos). Même si les individus du site de Los Amarillos présentant des liens de proche parenté ont été enlevés de l'échantillon, cet haplotype reste surreprésenté par rapport aux

autres haplotypes mitochondriaux. Des tests complémentaires ont donc été réalisés en enlevant de l'échantillon certains individus du site de Los Amarillos qui présentent cet haplotype particulier (cinq individus enlevés). Avec un effectif de 29 séquences, la diversité haplotypique augmente à 0,930, mais les autres indices (indice de Fu, nombre moyen de différences par paires de séquences et diversité nucléotidique) restent quasiment identiques.

Pour ce qui est des paramètres de diversité calculés à partir des haplotypes Y, l'effectif semble trop faible pour que ces indices de diversité soient réellement représentatifs.

Tableau 19 : Paramètres de diversité calculés à l'aide du logiciel arlequin pour les populations anciennes des vallées et de Pampa Grande.

	Ht-Mt*							Ht-Y*			
	N	h ¹	k ¹	Hd ¹	π ¹	Pw ¹	F _s ¹	N	h	Hd	Pw
Vallées	34	17	26	0,884	0,0174	6,242	-3,153	12	11	0,985	3,09
Pampa Grande	19	9	17	0,912	0,017	6,35	0,535	5	5	1,00	3,7

* Les séquences HVR-1 ont été analysées entre les positions 16024-16380 pb. Seuls les haplotypes Y minimum ont été utilisés pour cette analyse.

¹ h = nombre d'haplotypes différents ; k = nombre de sites polymorphes ; Hd = diversité haplotypique ; π = diversité nucléotidique ; Pw = nombre moyen de différences entre deux paires de séquences ; F_s = indice de neutralité de Fu.

III. Comparaisons inter-populationnelles.

1. Distribution des haplogroupes mitochondriaux et Y en Amérique du Sud.

1.1. Comparaisons de fréquences.

La grande majorité des informations disponibles dans la littérature sur les populations amérindiennes étant les fréquences des haplogroupes mitochondriaux, nous avons dans un premier temps compilé les données disponibles concernant les haplogroupes de populations sud-américaines anciennes et contemporaines afin de réaliser des comparaisons avec nos populations (l'ensemble des fréquences de lignées maternelles de toutes les populations utilisées au cours de ce travail sont présentées en annexe 6, p.238).

Comme cela est présenté dans le tableau 20 suivant, une fréquence importante d'haplogroupe B se retrouve dans de nombreuses populations anciennes des Andes centrales (sauf pour la population de Tiwanaku). En ce qui concerne les populations contemporaines, les régions des Andes centrales, Andes méridionales et du Chaco sont caractérisées par une majorité de lignée B au sein de leurs groupes, avec des fréquences respectivement de 62 %, 55.6 % et 41.5%. La fréquence élevée des lignées A, dans la population ancienne des vallées (42.5 %) que nous avons étudiée, conduit donc à différencier cette population des populations andines contemporaines qui pour la plupart présentent d'assez faibles fréquences de cette lignée. Il faut cependant noter, que les populations anciennes d'Ayacucho, de Sican, de Chen Chen et d'Arica (aux périodes céramiques initiale et intermédiaire) présentent des fréquences notables d'haplogroupe A (entre 21,4 % et 50 %). Actuellement, les fréquences les plus élevées de cette lignée sont détectées au sein de l'aire caribéenne (41,5 %) et tendent à décroître en allant vers le sud, notamment vers l'aire andine et la Patagonie. Il faut cependant, noter que l'est de l'Amérique du Sud, notamment la partie amazonienne présente des fréquences non négligeables de cet haplogroupe A, avec un pic de fréquence dans la région des hauts-plateaux brésiliens (54,5 %). En ce qui concerne la population de Pampa Grande, si elle présente des similitudes avec les populations andines de part sa fréquence élevée en haplogroupe B (47,6 %), elle se démarque de ces dernières par la fréquence élevée de l'haplogroupe D (38,1 %). Les populations anciennes de la côte Pacifique péruvienne, ainsi que la population ancienne

de Terre de Feu présentent également des fréquences élevées de cette lignée maternelle D (entre 25 % et 58,7 %), alors qu'au sein des populations anciennes des Andes centrales cette lignée est très peu représentée, voire absente. Des fréquences élevées d'haplogroupe D ne sont retrouvées actuellement que dans les populations de l'extrême sud du sous-continent américain, notamment en Patagonie, mais surtout en Terre de Feu, et sont généralement plutôt faibles dans les autres régions d'Amérique du Sud. En ce qui concerne la lignée maternelle C, aucun gradient de fréquence ne se dégage clairement, il n'est donc pas associé à une région particulière. A l'image de la différence de fréquence observée entre nos deux populations des Andes méridionales, les fréquences de la lignée C peuvent varier assez fortement entre différentes populations d'une même zone géographique.

Tableau 20 : Fréquences des lignées amérindiennes dans les populations amérindiennes anciennes et dans les différentes régions géographiques d'Amérique du Sud.

Populations / Sites		N	A	B	C	D	Autre
Vallées		41	43,9	29,3	14,6	12,2	0
Pampa Grande		21	14,3	47,6	0	38,1	0
Populations anciennes							
	Amazonie	18	50	16,7	33,3	0	0
Côte Pérou	Sican	28	21,4	25	3,6	25	25
	Palpa	130	2,9	11,5	26,9	58,7	0
Andes centrales	Ayacucho	32	21,9	34,4	37,5	6,2	0
	Laramate P.C.Int	32	6	57	37	0	0
	Laramate P.C.T	40	5	55	31	9	0
	ChenChen	23	39	39	17	4	0
	Tiwanaku	13	8	15	23	23	31
	Arica P.C.I	14	50	35,7	7,1	7,1	0
	Arica P.C.Int	19	31,6	42,1	26,3	0	0
	Arica P.C.T	15	7	68	11,9	13,1	0
Terre de Feu	Fueguian	24	0	0	58,2	41,8	0
Régions géographiques d'Amérique du Sud							
	Aire caribéenne	726	41,5	23	26,6	3,9	5
	Andes septentrionales	389	27	29,3	33,7	6,9	3,1
	Andes centrales	1040	9,7	62	15,5	11,2	1,6
	Amazonie	1060	19	27	33,4	15	5,6
	Andes méridionales	351	15,1	55,6	17,7	11,4	0,2
	Chaco	499	11	41,5	25,5	22	0
	Hauts-plateaux brésiliens	486	54,5	14,8	22,3	8,4	0
	Pampa	310	18,5	12	42,6	26,9	0
	Patagonie / Terre de Feu	622	2,9	14,3	35,2	44,2	3,4

A l'inverse des données concernant le génome mitochondrial, les données sur les fréquences des haplogroupes du chromosome Y sont relativement rares en ce qui concerne les populations sud-américaines. De plus, la phylogénie de la lignée Q, majoritaire en Amérique, évolue constamment et de nouveaux marqueurs ou de nouvelles lignées sont découverts très régulièrement [l'haplogroupe Q1a3a4 a été décrit en *octobre* 2011, (Jota *et al.*, 2011)], ce qui rend les comparaisons difficiles et peu informatives.

1.2. Distances génétiques et analyses multidimensionnelles.

Afin de mieux comprendre les relations entre les populations anciennes d'Amérique du Sud, une matrice de distance¹, basée sur les F_{ST} a été générée grâce au logiciel arlequin, à partir des comparaisons de populations faites sur la base des fréquences des haplogroupes mitochondriaux. Un arbre neighbor joining (NJ) a été construit à partir de la matrice de distance afin de mieux visualiser les relations entre ces populations anciennes.

Cette analyse, dont le résultat est présenté sur la figure 23, souligne que la population ancienne de Terre de Feu (Fueguian) ainsi que la population de la côte péruvienne de Palpa sont clairement différenciées des autres populations anciennes. La population des Vallées semble assez proche des populations andines d'Arica (aux périodes céramique initiale et intermédiaire) et de Chen Chen, ainsi que de la population amazonienne, ce qui est essentiellement dû aux fréquences relativement similaires des lignées A qu'elles possèdent. Nous pouvons également remarquer que si le reste des populations anciennes des Andes centrales sont regroupées, la population de Pampa Grande apparaît relativement isolée. Probablement dû à la fréquence élevée d'haplogroupe D, Pampa Grande présente des distances modérées à importantes avec toutes les populations andines (entre 0,11 avec Laramate à la période tardive et 0,21 avec Laramate à la période intermédiaire), mis à part avec les populations d'Arica (période tardive) et de Sican avec lesquelles la divergence est faible mais non significative. Ainsi, même si globalement la divergence entre populations andines reste faible ou modérée, deux groupes distincts semblent se dégager parmi ces populations.

Nous avons également voulu savoir, à partir des fréquences de lignées maternelles comment

¹ Les valeurs de F_{ST} et les p-values correspondantes pour les deux populations étudiées sont présentées en annexe 7 (p.239).

se situent les deux populations anciennes analysées par rapport à la diversité mitochondriale sud-américaine contemporaine. Comme nous pouvons l'observer sur la figure 24, l'ACP générée à partir des fréquences de lignées maternelles permet de différencier les populations andines (Andes centrales et méridionales essentiellement) par rapport aux populations des autres régions géographiques. En effet, ces populations andines sont très homogènes et regroupées entre elles contrairement aux populations amazoniennes par exemple. Les populations de la Terre de Feu se distinguent également étant donné qu'elles ne possèdent quasiment que des lignées maternelles C et D.

Si la population de Pampa Grande semble présenter des affinités avec le groupe homogène de populations andines, la population des Vallées au contraire apparaît un peu plus divergente. Le groupe des vallées présente des affinités avec un groupe de populations hétérogènes regroupant des populations anciennes (Arica du nord du Chili et Chen Chen de l'extrême sud du Pérou), une population contemporaine des Andes septentrionales (Zenu/Embera), deux populations amazoniennes (Wayampi et Llanos) et une population de l'aire caribéenne (Arawak).

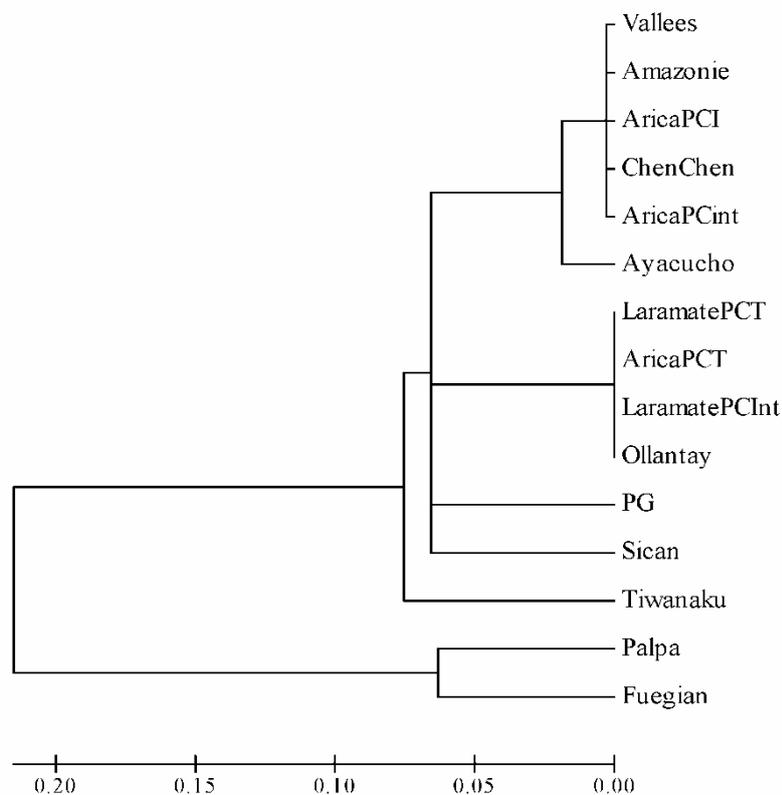


Figure 23 : Arbre Neighbor Joining généré à partir d'une matrice de distances F_{ST} obtenues en comparant les fréquences de lignées maternelles de 15 populations amérindiennes anciennes.

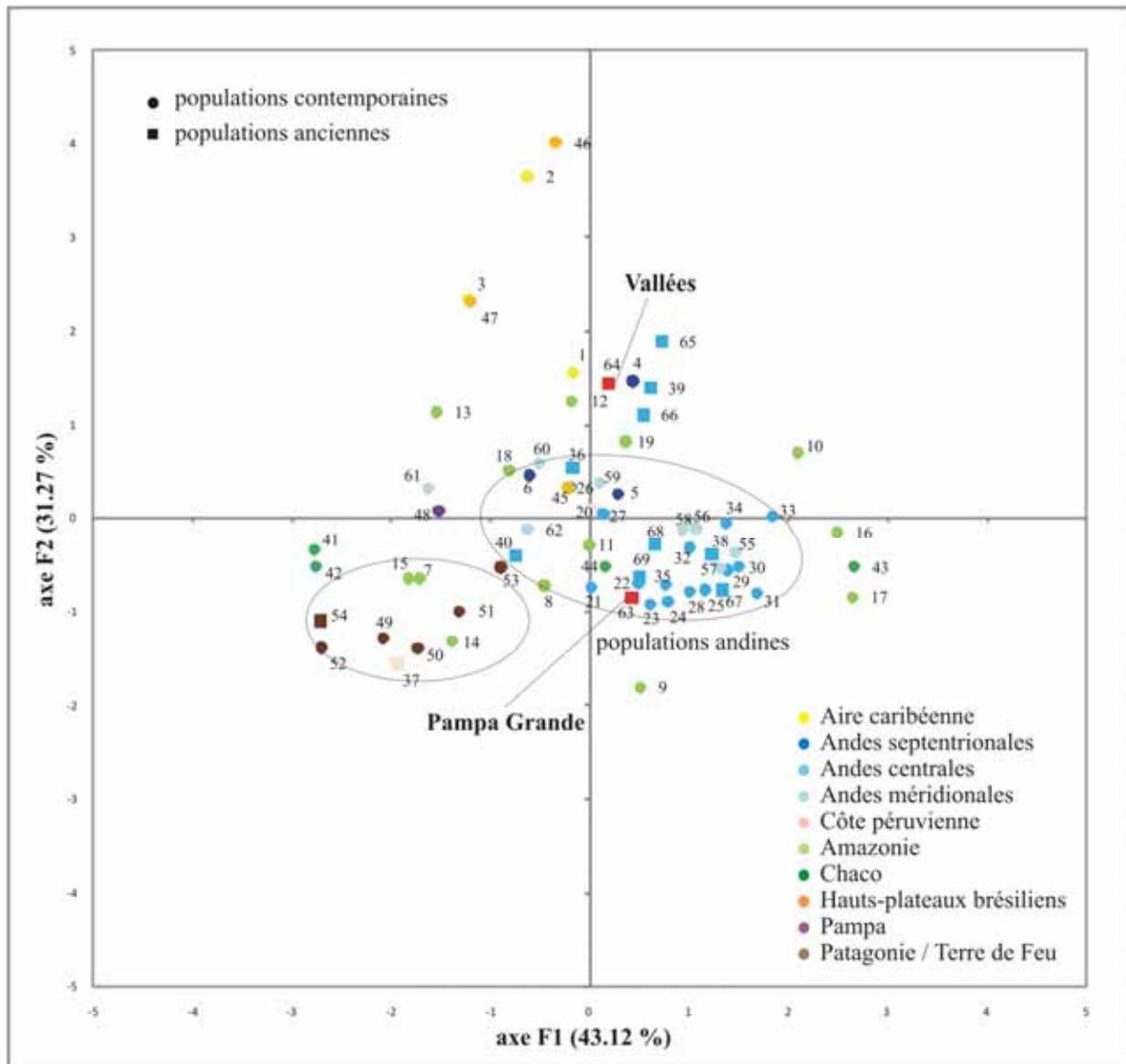


Figure 24 : Analyse en composante principale (ACP) générée à partir des fréquences de lignées maternelles de 69 populations anciennes et contemporaines amérindiennes.

1 : Arawak ; 2 : Chibchan ; 3 : Guahibo ; 4 : Zenu/Embera ; 5 : Ingano ; 6 : Cayapa ; 7 : Ticuna ; 8 : Yanomami ; 9 : Palikur ; 10 : Emerillon ; 11 : Kali'na ; 12 : Wayampi ; 13 : Apalai ; 14 : Gaviao ; 15 : Zoro ; 16 : Xavante ; 17 : Matsiguenga ; 18 : Moxos ; 19 : Llanos ; 20 : Sican ; 21 : Yungay ; 22 : Ancash ; 23 : Lima/Pasco ; 24 : San Martin ; 25 : Tupe ; 26 : Tayacaja ; 27 : Quechua (Pérou) ; 28 : Puno ; 29 : Arequipa ; 30 : Titicaca ; 31 : Aymara (Bolivie) ; 32 : La Paz ; 33 : Piedmont ; 34 : Quechua (Bolivie) ; 35 : Aymara (Chili) ; 36 : Ayacucho ; 37 : Palpa ; 38 : Ollantaytambo ; 39 : Chen Chen ; 40 : Tiwanaku ; 41 : Ayoreo (Bolivie) ; 42 : Ayoreo (Paraguay) ; 43 : Ache ; 44 : Gran Chaco ; 45 : Tuacarembo ; 46 : Guarani ; 47 : Kaingang ; 48 : Centre Argentine ; 49 : Pehuenche ; 50 : Huilliche/Mapuche ; 51 : Chiloe ; 52 : Yaghan ; 53 : Sud Argentine ; 54 : Fueguian ; 55 : Atacama ; 56 : Jujuy ; 57 : Puna ; 58 : Coyas ; 59 : Salta ; 60 : Nord-ouest Argentine ; 61 : Catamarca/Tucuman ; 62 : Centre ouest Argentine ; 63 : Pampa Grande ; 64 : Vallées ; 65 : Arica P.C.I ; 66 : Arica P.C.Int ; 67 : Arica P.C.T ; 68 : Laramate P.C.Int ; 69 : Laramate P.C.T.

2. Distribution et comparaison des haplotypes mitochondriaux et du chromosome Y en Amérique du Sud.

2.1. Comparaison de diversités génétiques.

Les indices de diversité génétique calculés à partir de variations de séquence HVR-1 pour les deux populations anciennes, ont été comparés à ceux des populations amérindiennes anciennes et contemporaines de notre base de données.

Les résultats, présentés dans le tableau 21, montrent que les deux populations des Andes méridionales possèdent une diversité génétique (haplotypique) plus faible¹ que celle des 4 populations anciennes. De plus ces populations anciennes montrent toutes des signes d'expansion démographique, avec des valeurs significatives et très largement négatives de l'indice F_S de F_u , ce qui n'est pas le cas des deux populations étudiées de Pampa Grande et des Vallées.

Les populations andines contemporaines présentent généralement une diversité génétique plus élevée que celle de nos deux populations analysées, et un grand nombre montre également des signes d'une expansion démographique. Il s'agit des seules régions, où quasiment toutes les populations présentent de telles caractéristiques. En effet, au sein des autres aires géographiques sud-américaines, les paramètres de diversité génétique sont beaucoup plus variés, avec des valeurs allant de 0.204 (Aché) jusqu'à 0.996 (Tuacarembó). De plus, globalement, les populations de ces régions ne montrent aucun signe évident d'expansion (mis à part les populations du Gran Chaco, Tuacarembó et Moxos).

Cette comparaison est cohérente avec les analyses précédentes et montre que même si les deux populations que nous avons étudiées ne sont pas significativement différentes des populations andines, elles divergent tout de même par certains points.

Les calculs de diversité réalisés à partir des haplotypes Y n'ont pas été comparés à des données contemporaines car le nombre d'haplotype par population est trop faible pour pouvoir être représentatif.

¹ Même en considérant une diversité haplotypique de 0,930 pour la population des Vallées (lorsque l'on enlève les individus de Los Amarillos présentant le même haplotype), cette diversité reste plus faible que celle de la majorité des populations anciennes et contemporaines des Andes centrales et méridionales.

Tableau 21 : Indices de diversité génétique et indice de neutralité, calculés à l'aide du logiciel Arlequin pour 4 populations anciennes et 39 populations contemporaines sud-américaines, à partir des variations de séquence HVR-1.

Populations / Sites	N	h	K	Hd	π	Pw	Fs*	Référence
Pampa Grande	19	9	17	0,912	0,017	6,35	0,535	présente étude
Vallées	34	17	26	0,884	0,0174	6,242	-3,153	présente étude
Populations anciennes								
Ayacucho	28	20	31	0,968	0,018	6,4	-9,641	Kemp et al., 2009
Laramate	65	29	35	0,95	0,016	5,78	-12,049	Ferhen-Schmitz et al., 2010
Palpa	92	51	38	0,977	0,013	4,65	-25,649	Ferhen-Schmitz et al., 2010
Arica	30	29	41	0,998	0,017	6,1	-25,229	Moraga et al., 2005
Aire caribéenne								
Arawak	30	5	17	0,789	0,016	5,8	6,49	Melton et al. 2007
Chibchan	80	7	15	0,681	0,008	2,89	3,294	Melton et al. 2007
Guahibo	59	12	13	0,858	0,014	4,98	0,934	Vona et al. 2005
Amazonie								
Yanomami.1	155	6	15	0,657	0,011	3,99	7,82	Williams et al. 2002
Emerillon	31	3	10	0,471	0,01	3,535	70,6	Mazières et al., 2008
Kalin'a	29	14	30	0,938	0,021	7,32	-0,935	Mazières et al., 2008
Palikur	47	9	14	0,806	0,014	5,125	2,584	Mazières et al., 2008
Wayampi	53	12	24	0,882	0,019	6,831	2,233	Mazières et al., 2008
Apalai	103	16	31	0,972	0,0173	6,174	1,28	Mazières et al., 2008
Yanomami.2	50	20	31	0,828	0,011	4,1	-6,77	Easton et al. 1996
Gaviao	27	7	16	0,866	0,012	4,42	1,935	Ward et al. 1996
Zoro	30	8	16	0,772	0,01	3,92	1,067	Ward et al. 1996
Xavante	25	4	9	0,677	0,007	0,52	3,719	Ward et al. 1996
Matsiguenga	38	11	16	0,777	0,0074	2,654	-2,15	Mazières et al., 2008
Plaines Moxos	54	33	40	0,969	0,017	6	-19,326	Bert et al. 2004
Hauts-plateaux brésiliens								
Guarani	200	9	13	0,764	0,007	2,5	1,738	Marrero et al. 2007
Kaingang	77	11	24	0,767	0,018	6,34	3,969	Marrero et al. 2007
Tuacarembo	24	23	39	0,996	0,02	7,177	-16,021	Bonilla et al., 2004
Andes septentrionales								
Cayapa	32	8	20	0,853	0,018	6,48	3,477	Tamm et al. 2007
Andes centrales								
Yungay	36	20	27	0,954	0,014	5,11	-7,38	Lewis et al. 2007
Ancash	33	27	43	0,981	0,0171	6,09	-19,01	Lewis et al. 2004
San Martin	22	15	21	0,939	0,014	5,1	-5,486	Fuselli et al. 2003
Tupe	16	9	17	0,867	0,014	5	-0,765	Lewis et al. 2007
Tayacaja	61	42	41	0,968	0,017	5,98	-25,273	Fuselli et al. 2003
Puno	44	32	45	0,975	0,014	5	-24,589	Lewis et al. 2007
Arequipa	22	18	25	0,978	0,013	4,93	-10,72	Fuselli et al. 2003
Titicaca	64	31	38	0,958	0,015	5,51	-14,028	Barbieri et al. 2011
Aymara	96	43	53	0,968	0,013	4,78	-25,538	Gaya-Vidal et al. 2011
La Paz	110	66	67	0,956	0,018	6,4	-25,026	Alfonso Costa et al. 2010
Piedmont andin	46	21	37	0,903	0,013	4,75	-7,001	Corella et al., 2007
Quechua	93	46	55	0,961	0,018	6,26	-25,099	Gaya-Vidal et al. 2011
Andes méridionales								
Coyas	61	38	44	0,97	0,016	5,7	-25,129	Alvarez-Iglesias et al. 2007
Salta	18	14	22	0,967	0,017	6,2	-4,844	Tamm et al. 2007
Catamarca/Tucuman	34	16	25	0,907	0,016	5,66	-3,791	Tamm et al. 2007
Chaco								
Gran Chaco	204	46	57	0,938	0,019	6,63	-14,439	Cabana et al. 2006
Ache	63	3	6	0,204	0,003	1,08	2,546	Schmitt et al. 2004
Patagonie								
Pehuenche	24	13	22	0,902	0,016	5,63	-1,993	Moraga et al. 2000
Huilliche - Mapuche	34	9	17	0,838	0,015	5,31	1,967	Moraga et al. 2000
Yaghan	15	7	15	0,886	0,015	5,26	0,91	Moraga et al. 2000

2.2. Haplotypes partagés.

Tous les haplotypes mitochondriaux et Y que nous avons détecté au sein des populations de Pampa Grande, des Vallées et pour les individus de Doncellas, ont été comparés à notre base de données, afin de voir dans quelles populations anciennes et/ou contemporaines amérindiennes ils avaient déjà été observés. Les résultats obtenus sont résumés dans les tableaux 22 et 23.

Seuls 10 des haplotypes mitochondriaux détectés au sein des échantillons anciens analysés ont été retrouvés dans des populations contemporaines ou anciennes amérindiennes. Ainsi, 13 haplotypes (NOA-A2, NOA-A3, NOA-B1, NOA-B2, NOA-B3, NOA-B4, NOA-B5, NOA-B6, NOA-B7, NOA-B8, NOA-C2, NOA-C4, NOA-D4) n'ont apparemment pas encore été décrits à ce jour, et semblent donc restreints aux individus anciens de Pampa Grande, de la Quebrada de Humahuaca, de la vallée Calchaqui et de Doncellas. En ce qui concerne les haplotypes associés à la lignée B2, même en enlevant les positions hypervariables 16182 et 16183, un résultat identique a été trouvé. Ainsi, l'absence de correspondance dans notre base de données n'est pas due à la présence de ces positions hypervariables.

De manière attendue, les haplotypes NOA-A1 (16111T – 16223T – 16290T – 16319A – 16362C), NOA-C3 (16223T – 16298C – 16325C – 16327T) et NOA-D3 (16223T – 16325C – 16362C), que l'on peut qualifier de fondateurs des lignées A2, C1 et D1 respectivement, étant donné qu'ils présentent les mutations caractéristiques de ces lignées sans mutations supplémentaires, sont les plus fréquemment rencontrés, dans toute l'Amérique du Sud, aussi bien dans les populations anciennes que contemporaines. Les haplotypes NOA-B9 et NOA-B10 sont également très fréquemment retrouvés dans plusieurs régions d'Amérique du Sud.

Les 5 haplotypes restants sont retrouvés à des fréquences beaucoup plus faibles. L'haplotype NOA-A4 a été retrouvé dans trois populations : les Guaranis (8 individus), Moxos (2 individus) et le Chaco (1 individu). Trois individus de la population des Guahibos ainsi que deux individus de populations contemporaines andines (La Paz et Ancash), présentent le même haplotype NOA-C1, porté par un individu des Vallées. Les trois derniers haplotypes NOA-B11, NOA-B12 et NOA-D1 ne sont partagés qu'avec des populations andines anciennes et/ou contemporaines, et avec des individus du Chaco. L'haplotype NOA-B11 notamment, semble relativement fréquent dans ces régions.

Tableau 22 : Répartition et nombre des haplotypes mitochondriaux partagés (HVR-1 : 16024-16380), entre les populations anciennes de Pampa Grande, des Vallées, de Doncellas, et les populations anciennes et contemporaines sud-américaines.

	NOA- A1	NOA- A4	NOA- B9	NOA- B10	NOA- B11	NOA- B12	NOA- C1	NOA- C3	NOA- D1	NOA- D3	Total (%)
Vallées	11	1	2	2	1	.	1	1	.	4	/
Doncellas	1	/
Pampa Grande	2	.	.	1	2	3	/
Populations anciennes											
Palpa (92)	1	.	.	3	.	.	.	5	.	10	20,1
Laramate (65)	1	.	5	9	.	3	.	9	.	3	46,2
Ayacucho (28)	1	.	1	3	.	2	25
Arica (30)	1	3,3
Aire caribéenne											
Arawak (30)	7	.	.	23,3
Caracas (98)	17	3	.	1	21,4
Guahibo (59)	3	9	.	.	20,3
Amazonie											
Yanomami.1 (155)	.	.	.	75	.	.	.	49	.	.	80
Wayampi (53)	5	9
Emerillon (30)	9	.	.	19	93
Kalina (29)	.	.	2	1	.	.	.	3	.	.	21
Palikur (48)	.	.	3	2	10
Xavante (25)	4	13	68
Zoro (30)	13	43,3
Gaviao (27)	5	18,5
Apalai (102)	11	1	11,8
Matsiguenga (38)	.	.	.	3	2,6
Moxos (54)	5	2	1	7	.	3	33,3
Hauts-plateaux brésiliens											
Guarani (200)	79	8	13	50
Kaingang (77)	30	12	.	.	54,5
Tuacarembó (24)	1	.	.	4,2
Andes septentrionales											
Cayapa (32)	4	1	.	.	15,6
Andes centrales											
Piedmont (46)	1	2,2
Yungay (36)	1	2	8,3
San Martin (22)	1	3	18,2
Arequipa (22)	1	.	1	9,1
Tayacaja (61)	2	2	1	10	24,1
Ancash (33)	1	3	.	1	15,2
Titicaca (64)	3	.	1	4	9	1	.	2	.	4	37,5
Tupe (16)	1	.	.	6,25
Puno (44)	1	.	.	2,3
Aymara.2 (96)	2	.	7	9	9	.	.	3	.	3	34,3
Quechua.2 (93)	5	.	4	12	9	.	.	8	.	.	40,1
La Paz (110)	4	.	1	10	20	2	1	5	.	1	40
Andes méridionales											
Coyas (61)	2	.	2	3	3	1	.	4	.	1	26,2
Salta (18)	1	.	1	2	.	.	.	3	.	.	38,9
Catamarca(34)	8	.	1	2	.	.	32,4
Chaco											
Gran Chaco (204)	.	1	35	7	3	.	.	10	.	11	32,8
Patagonie											
Pehuenche (24)	7	.	.	29,2
Mapuche (34)	12	.	.	35,3

Tableau 23 : Haplotypes Y partagés avec les populations sud-américaines. Les comparaisons ont été effectuées à partir de sept loci (haplotype minimal moins le loci DYS385)

	Vallées	PG	Populations																			
			Colombie (305)	Pool Antioquia (1177)	Caracas (62)	Kichwa (102)	Metisses Colombie (102)	Afro-equatoriens (94)	Quito (120)	Pool Brésil (479)	Pool Sud Brésil (249)	Pérou (76)	Pool Beni (231)	Aymara (84)	Quechua (55)	Pool Colla (63)	Diaguita (24)	Pool Toba (47)	Cordoba (100)	Mapuche (76)	Salta (103)	Argentine (1575)
H2	1		1	.	.	.	1
H3	1		.	3	.	3	.	.	2	.	.	.	2	2	1	2	.
H4		1	1
H6	1		2	1	.	1	.	1	1	.	2	4	1	.	.	.	4	.
H7	1		1	2	1	3	.	1	2	.	7	4	.	6	1	1	1	8	.	.	1	4
H8		1	.	1	4	1	.	.	.	1
H10	2		1	1	1	.
H13	1		.	1	.	.	1	2
H14	1		1
H15		1	.	.	.	2	.	1	2	.	.	2
Total (%)			1,3	0,7	1,6	8,8	1	4,3	5,8	0,8	2,8	13,2	2,2	9,5	3,6	1,6	4,2	19,1	1	2,6	2,9	0,8

Les haplotypes Y n'ont été comparés que pour sept marqueurs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 et DYS393). L'ensemble des résultats est présenté dans le tableau 23 ci-dessus.

Sur les 16 haplotypes Y qui ont été comparés, 6 (H1, H5, H9, H11, H12 et H16) n'ont pas été retrouvés dans les populations de notre base de données et les 10 autres sont partagés par au moins 1 individu (H4 et H14) et au maximum 43 individus (H7). Les haplotypes H6 et H7 sont les plus fréquemment retrouvés dans les populations contemporaines de toutes les régions de l'Amérique du Sud. Seuls deux haplotypes semblent restreints aux populations andines (H4 et H14), alors que tous les autres sont partagés entre populations de toutes les aires géographiques d'Amérique du Sud.

Globalement, les pourcentages totaux d'haplotypes partagés sont relativement faibles, sauf pour quatre populations de l'aire andine : les Kichwa d'Equateur (8,8 %), une population du Pérou (13,2 %), les Aymaras de Bolivie (9,5 %) et les Tobas du nord-ouest de l'Argentine (19,1 %).

2.2.1. Analyses phylogénétiques.

La distribution des haplotypes au sein des différents haplogroupes mitochondriaux et Y détectés au sein de notre échantillon est illustrée grâce à des réseaux phylogénétiques réalisés à l'aide du logiciel Network.

Comme illustré sur la figure 25 suivante, les réseaux générés à partir des haplotypes des lignées maternelles A2, B2, C1 et D1 confirment les résultats déjà observés grâce à l'analyse des haplotypes partagés. Les haplotypes des groupes anciens étudiés se situent au sein de la diversité génétique contemporaine. La configuration de ces quatre réseaux est quasiment identique, avec au centre un nœud central que l'on peut considérer comme représentatif de l'haplotype fondateur et qui est partagé par de nombreuses populations de régions très différentes. Les radiations issues de ce nœud central correspondent le plus souvent à des haplotypes peu fréquents voire uniques, ou à des haplotypes partagés par des individus issus d'une même région ou de régions proches. Le schéma le plus simple pour expliquer la distribution des haplotypes observée sur ces réseaux est celui d'une expansion rapide puis d'une différenciation des haplotypes au sein des différentes régions. Dans l'ensemble, mis à part pour les haplotypes associés à la lignée C1, on retrouve assez peu d'haplotypes partagés entre populations andines et amazoniennes.

Pour le réseau illustrant la répartition des haplotypes de la lignée B2, qui est majoritaire au sein des populations que nous avons analysées ainsi que dans les populations andines, nous avons choisi d'exclure de l'analyse les positions 16182 et 16183 qui sont hypervariables et qui ne sont pas indiquées de manière cohérente par les différents auteurs (Malhi *et al.*, 2003; Pfeiffer *et al.*, 1999).

La majorité des haplotypes détectés au sein de notre échantillon sont uniques ou partagés avec des populations andines (essentiellement des Andes centrales et méridionales). Cependant, nous pouvons également remarquer que certains haplotypes partagent des affinités avec des individus d'autres régions géographiques, comme pour NOA-B8 qui présentent des liens avec le Gran Chaco, NOA-A4 qui possède des affinités avec les populations plus à l'est et notamment les hauts-plateaux brésiliens, et enfin NOA-C1 associé à des populations de l'Aire caribéenne.

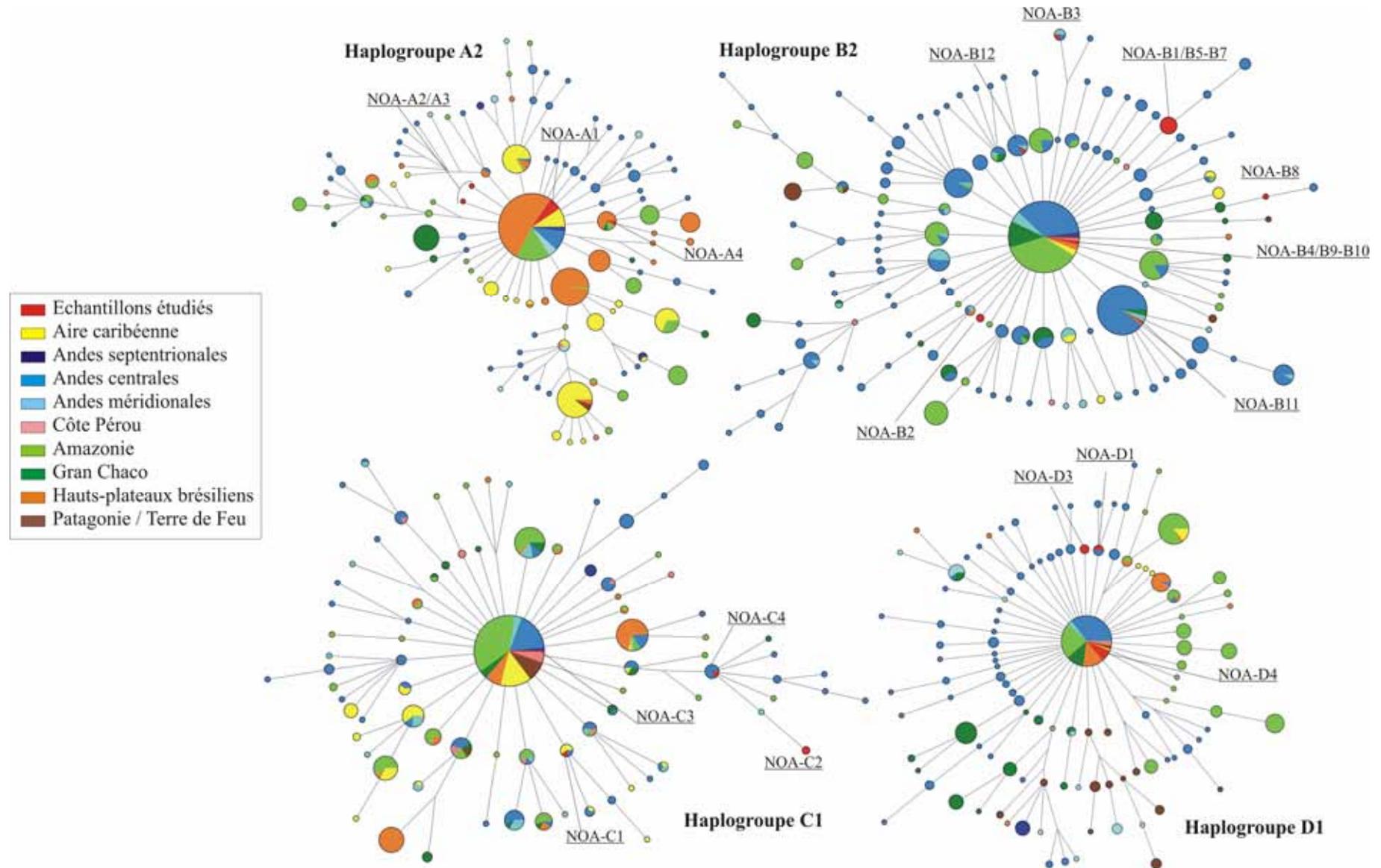


Figure 25 : Réseaux phylogénétiques illustrant la distribution des haplotypes au sein des quatre lignées maternelles amérindiennes fondatrices. La taille de chaque cercle est proportionnelle à la fréquence de l'haplotype au sein de l'échantillon. Pour l'haplogroupe B2, les positions hypervariables 16182 et 16183 n'ont pas été prises en compte dans l'analyse.

Le réseau phylogénétique représentant la distribution des haplotypes de la lignée Q présente une configuration très différente de celle des réseaux des lignées maternelles (Figure 26). Ce réseau présente en effet plusieurs branches clairement différenciées, dont une spécifique de la lignée Q1a3a4, typiquement andine (Jota *et al.*, 2011). Nous pouvons remarquer que nos échantillons ne présentent pas d'affinité particulière avec les individus partageant une lignée spécifique des Andes. Il est intéressant de noter que les haplotypes associés au paragroupe Q1a3* sont regroupés ensemble mais sont également associés à des haplotypes identifiés comme appartenant à la lignée Q1a3a*. Il faut cependant rappeler que nous disposons de relativement peu d'haplotypes associés à la lignée Q de manière fiable, ce qui a limité l'analyse réalisée. Il apparaît probable qu'une phylogénie plus précise de la lignée Q permettrait de révéler différents sous-haplogroupes au sein de cette lignée.

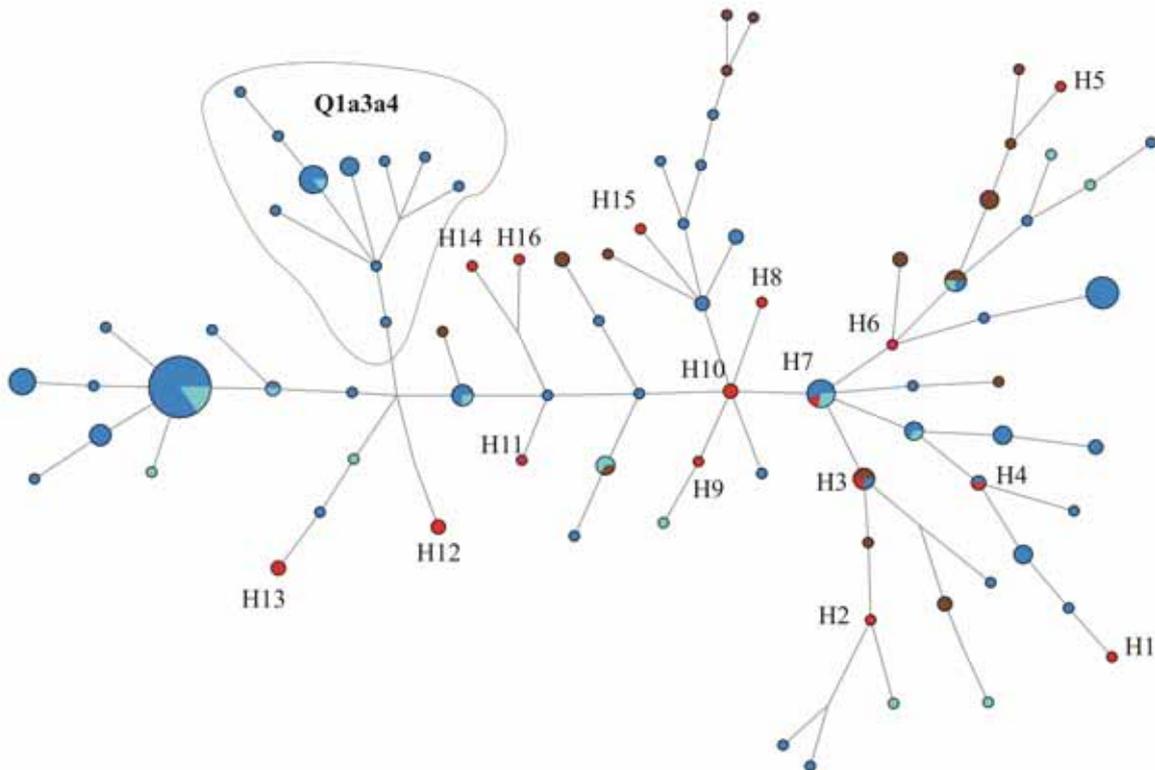


Figure 26 : Réseau phylogénétique présentant les haplotypes Y (haplotype minimal moins DYS385) amérindiens associés à la lignée paternelle fondatrice Q.

Les échantillons anciens étudiés au cours de ce travail sont colorés en rouge, les échantillons contemporains des Andes centrales en bleu foncé, les individus contemporains des Andes méridionales en bleu clair et les échantillons contemporains de Patagonie en marron. La taille des cercles est proportionnelle à la fréquence de l'haplotype dans l'échantillon.

2.3. Calculs de distances et analyses multidimensionnelles.

Deux matrices de distances ont été générées à partir des séquences HVR-1, mais aussi à partir des haplotypes Y. Les résultats complets de ces matrices de distances pour les deux populations étudiées sont présentés en annexes 8 et 9 (p.240-241). L'affinité des populations anciennes de Pampa Grande et des Vallées avec les populations contemporaines et anciennes de l'ensemble de l'Amérique du Sud a ainsi été évaluée.

La figure 27 montre que les populations anciennes analysées ne se différencient que très peu des populations andines contemporaines et anciennes, lorsque l'on considère les variations de séquences mitochondriales. En effet, même si elles apparaissent en périphérie, elles ne sont que faiblement voire modérément divergentes des populations andines. Les résultats obtenus soulignent la grande homogénéité des populations andines, qui a été évoquée par de nombreux auteurs. Ainsi, même si quelques populations, comme les Tupe, les Cayapa, les groupes du piedmont bolivien, ou la population de Catamarca se différencient quelque peu, au même titre que Pampa Grande et les Vallées, du pool global andin, une unité et homogénéité génétique importante se dégage de l'ensemble des régions andines (septentrionale, centrale et méridionale). Cette homogénéité contraste avec la divergence des populations des autres régions d'Amérique du Sud, et notamment les populations amazoniennes. Il est tout de même important de noter que lorsque l'on considère les variations de séquences, les populations andines sont un peu moins divergentes du reste des populations sud-américaines que lorsque l'on prend en compte les fréquences de lignées maternelles.

L'analyse réalisée à partir des haplotypes Y, présentée sur la figure 28, souligne également l'homogénéité andine. En effet, les populations de l'aire andine se détachent des populations des autres régions d'Amérique du Sud. Nous pouvons remarquer que les deux populations anciennes des Andes méridionales présentent des affinités claires avec les populations contemporaines andines, et notamment avec les individus masculins contemporains du nord-ouest argentin (Diaguïta et Colla), ce qui permet de supposer une certaine continuité génétique. Il est toutefois important de noter que les effectifs de chaque population étant relativement faibles, l'analyse n'est que faiblement représentative mais permet de dégager une tendance générale.

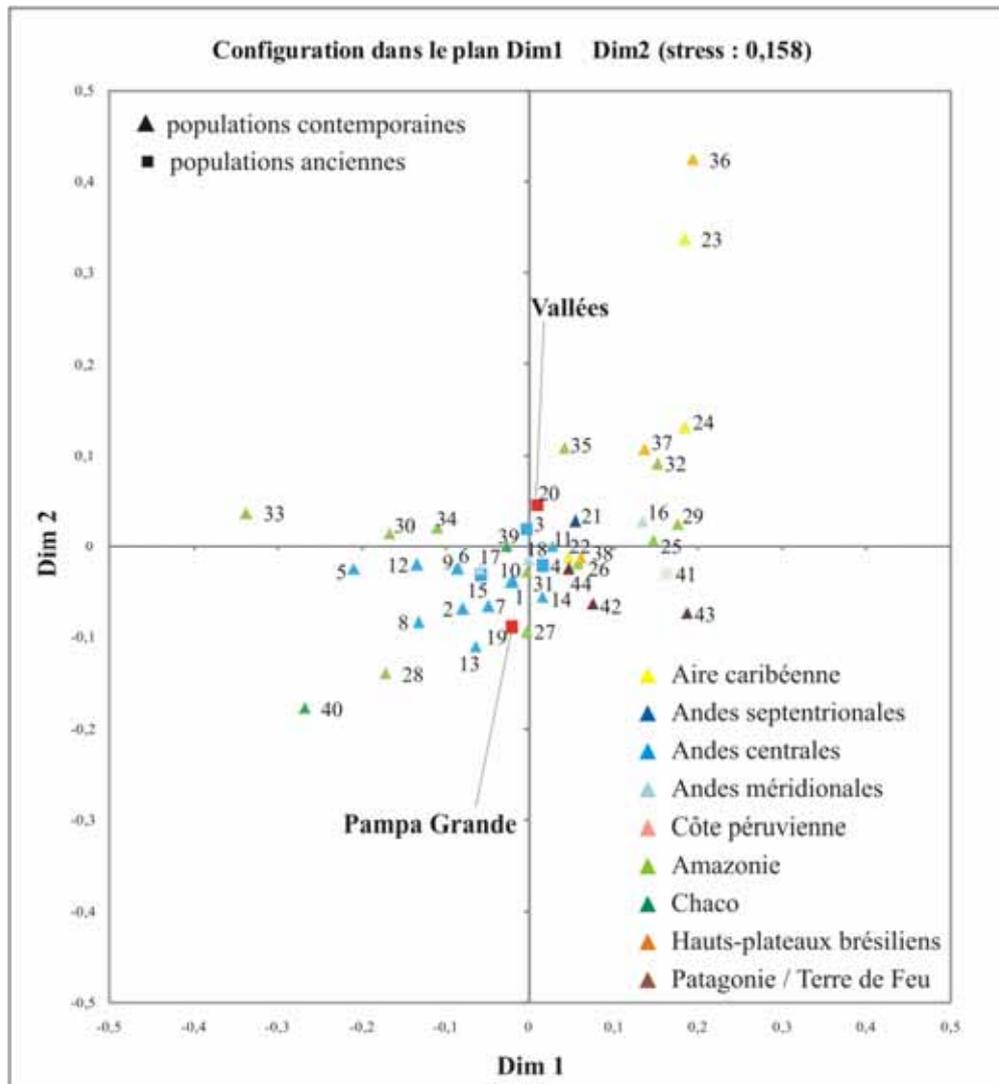


Figure 27 : MDS généré à partir des paires de distances F_{ST} de 44 populations andines contemporaines et anciennes, à partir des variations de séquences HVR-1 (16024-16380).

1 : Ancash ; 2 : Arequipa ; 3 : Arica ; 4 : Ayacucho ; 5 : Aymara (Bolivie) ; 6 : Quechua (Bolivie) ; 7 : Puno ; 8 : Piedmont ; 9 : La Paz ; 10 : San Martin ; 11 : Tayacaja ; 12 : Titicaca ; 13 : Tupe ; 14 : Yungay ; 15 : Laramate ; 16 : Catamarca/Tucuman ; 17 : Coyas ; 18 : Salta ; 19 : Pampa Grande ; 20 : Vallées ; 21 : Cayapa ; 22 : Arawak ; 23 : Chibchan ; 24 : Guahibo ; 25 : Gaviao ; 26 : Moxos ; 27 : Yanomami ; 28 : Xavante ; 29 : Zoro ; 30 : Emerillon ; 31 : Kali'na ; 32 : Apalai ; 33 : Matsiguenga ; 34 : Palikur ; 35 : Wayampi ; 36 : Guarani ; 37 : Kaingang ; 38 : Tuacarembo ; 39 : Gran Chaco ; 40 : Ache ; 41 : Palpa ; 42 : Huilliche/Mapuche ; 43 : Yaghan ; 44 : Pehuenche.

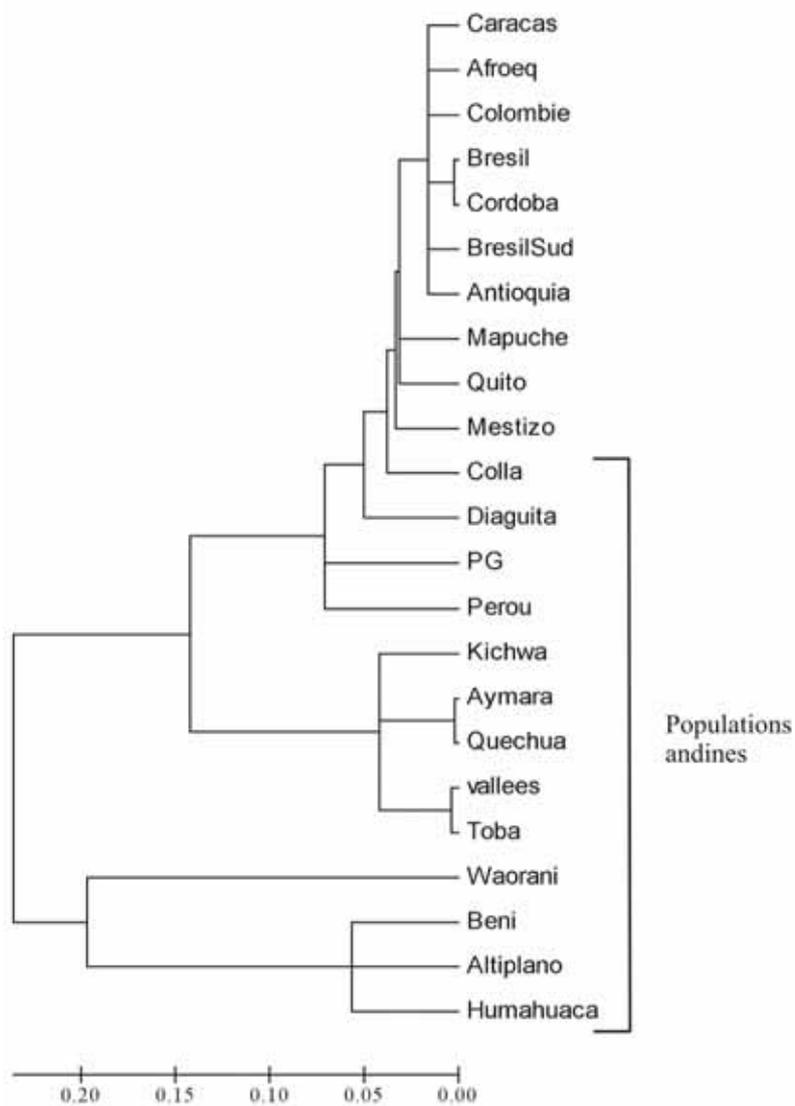


Figure 28 : Arbre phylogénétique (NJ) présentant les distances génétiques entre populations, calculées à partir des haplotypes Y.

Quatrième partie.

Discussion.

Même si ces dernières années des avancées importantes ont permis de mieux appréhender le peuplement de l'Amérique (Mazieres, 2011; O'Rourke and Raff, 2010; Perego *et al.*, 2009), les modalités de la diffusion des premières populations à toutes les régions d'Amérique du Sud et la mise en place de la diversité génétique observée actuellement demeurent assez mal compris. Une des raisons principales permettant d'expliquer les nombreuses problématiques encore irrésolues est le manque de données concernant les populations amérindiennes et plus particulièrement concernant les populations du passé.

Ce travail de thèse a donc été fondé sur la caractérisation de la diversité génétique, encore inconnue à ce jour, de populations anciennes des Andes méridionales, dans le but d'appréhender à la fois les pratiques de ces groupes mais aussi de mieux connaître leur histoire évolutive. Il apparaît que la question du peuplement de l'Amérique du Sud est tellement vaste qu'il est difficile de l'étudier dans sa globalité. Ainsi, l'apport de nouvelles données concernant des populations du passé permet de tester les hypothèses majeures et potentiellement d'en proposer de nouvelles.

Les populations anciennes des Andes méridionales ont fait l'objet de très peu d'études anthropologiques, menant à beaucoup d'incertitudes quant à leur origine, la nature des relations qu'elles ont pu nouer au cours du temps au sein même de la région ou avec des groupes voisins, ou encore leur évolution jusqu'à aujourd'hui. Ces différentes problématiques ont pu être abordées au cours de ce travail de thèse grâce à une approche originale consistant à étudier des marqueurs génétiques situés sur des régions différentes du génome humain et permettant d'obtenir des informations différentes mais complémentaires.

Au travers des différentes problématiques abordées, ce travail de thèse a également été l'occasion d'essayer de mettre en place de nouveaux protocoles permettant d'améliorer la méthodologie d'analyse de l'ADNa.

Avant de discuter des résultats obtenus concernant les populations en elles mêmes, nous discuterons d'abord de la stratégie d'analyse mise en place au cours de ce travail, en soulignant les bénéfices pour l'étude de l'ADNa mais aussi les limites auxquelles nous nous sommes heurtés.

I. Stratégie d'analyse de l'ADNa.

Dans le cadre de cette étude, nous avons essayé de mettre en place une stratégie analytique originale fondée sur l'analyse de différents marqueurs grâce à des techniques variées. La méthodologie employée au cours de ce travail nous permet de présenter des résultats authentiques et inédits concernant les populations anciennes des Andes méridionales, mais aussi d'envisager de nouvelles perspectives en ce qui concerne les techniques d'analyse de l'ADNa. Cependant, malgré les développements méthodologiques mis en place, nous nous sommes tout de même parfois heurtés à des limites qui ont pu freiner notre travail, comme par exemple l'accès aux échantillons ou l'état de conservation de l'ADN.

1. Etat de conservation de l'ADN.

L'étude présentée est la toute première fondée sur l'analyse d'échantillons anciens de la région des Andes méridionales. Si les résultats concernant les échantillons chiliens sont quelques peu décevants, les résultats concernant les échantillons du nord-ouest argentin sont assez exceptionnels tant par la quantité d'informations obtenues (marqueurs mitochondriaux mais aussi nucléaires) que par la qualité de ces informations. En effet, la stratégie employée au cours de ce travail, qui consistait à étudier plusieurs types de marqueurs sur des portions différentes du génome a conduit à l'obtention d'informations complémentaires nous permettant de discuter des pratiques de ces groupes amérindiens anciens mais aussi à une plus grande échelle de la dynamique d'évolution des populations des Andes méridionales avant l'arrivée des européens.

Concernant les échantillons du site de Solcor 3, il semble que l'absence de résultats exploitables soit la conséquence de la mauvaise qualité de l'ADN et/ou d'une contamination importante des échantillons. En effet, un premier test a été réalisé sur une petite partie de l'échantillonnage, révélant la présence d'ADN dans 9 des 11 échantillons testés. Cependant l'analyse des séquences obtenues a révélé de nombreuses hétéroplasmies ainsi que des incohérences importantes entre séquences obtenues d'amplifications différentes. Les contaminations observées correspondaient à des séquences amérindiennes, ainsi il ne

s'agissait pas d'une contamination par une personne ayant eu accès au laboratoire d'ADNa (présentant tous des séquences typiquement européennes), mais probablement des contaminations croisées entre échantillons, ou des contaminations par des fouilleurs ou chercheurs ayant travaillé sur cette collections auparavant. Cette collection de Solcor 3 est conservée au Musée Gustavo le Paige de San Pedro de Atacama depuis de nombreuses années, et a fait l'objet de nombreuses études anthropologiques (Bravo and Llagostera, 1986; Knudson and Price, 2007; Torres-Rouff, 2011), par différentes équipes de recherche. Il s'agit ainsi d'une collection qui a pu être manipulée par de nombreuses personnes. Il apparaît donc que malgré les précautions prises lors du choix des échantillons (seulement des dents intactes et en place ont été sélectionnées) et l'étape de décontamination réalisée, l'ADN de cette collection était trop profondément contaminé et dégradé pour obtenir des informations exploitables.

Au contraire, l'ADN au sein des échantillons du site de Pampa Grande était particulièrement bien conservé, ce qui est à mettre en relation avec l'environnement favorable des grottes dont est issu le matériel étudié. En plus des restes humains anciens découverts dans ces cavités, une quantité importante de divers matériaux comme du bois, des textiles, mais aussi des restes végétaux ont été mis au jour (Baffi *et al.*, 1996), attestant de l'excellente conservation des vestiges dans cette région.

Ces observations démontrent que les conditions de conservation des échantillons ainsi que les conditions taphonomiques sont déterminantes pour la préservation de l'ADNa.

2. Perspectives méthodologiques.

La méthodologie mise en place au cours de ce travail apparaît comme étant très intéressante pour les études d'ADNa. En effet, les divers marqueurs analysés nous permettent de discuter de plusieurs problématiques concernant les populations anciennes des Andes méridionales. Cependant, l'analyse de ces différents marqueurs ainsi que la nécessité de répliquer les manipulations impliquent qu'une quantité relativement importante de matériel soit disponible au départ. Un des objectifs de la paléogénétique est donc d'obtenir toujours plus d'informations fiables, à partir de la plus petite quantité de matériel possible. Dans ce cadre, la technique de génotypage de SNP par spectrométrie de masse, mise en place pour l'analyse des lignées maternelles et paternelles amérindiennes, nous est apparue comme étant très avantageuse pour l'analyse de l'ADNa.

2.1. Bénéfices du génotypage par spectrométrie de masse MALDI-TOF pour l'analyse de l'ADNa.

A la vue des récents développements des techniques utilisées en paléogénétique, qui permettent d'obtenir un grand nombre d'informations sur tous les génomes, il apparaît de plus en plus évident que le séquençage de petites régions ciblées du génome n'est plus suffisant. En effet, aujourd'hui, le séquençage nouvelle génération (NGS) s'impose d'avantage comme une technique prometteuse qui permettra de résoudre de nombreuses problématiques (Ho and Gilbert, 2010). Cependant, même si les avancées réalisées sont spectaculaires, il s'agit encore de techniques utilisées ponctuellement, réservées le plus souvent à des spécimens extrêmement précieux voire uniques (Rasmussen *et al.*, 2010; Shapiro and Hofreiter, 2010). Il semble donc encore utopique de pouvoir appliquer ces méthodes en routine pour des études de populations anciennes par exemple, d'autant plus que ces méthodes induisent un certain nombre de difficultés telles que le coût très important des analyses, la fiabilité des données obtenues, ou encore la quantité d'informations à analyser. Il est donc important de trouver en parallèle des techniques et des protocoles permettant d'obtenir un maximum d'informations des échantillons anciens analysés.

Depuis de nombreuses années l'intérêt pour les SNP est grandissant en anthropologie. Ces polymorphismes ponctuels présentent en effet de nombreuses caractéristiques intéressantes, présentées précédemment, et ont donc été utilisés aussi bien pour la caractérisation fine de phylogénies (Cerezo *et al.*, 2009; Endicott *et al.*, 2006; Geppert *et al.*, 2011; Grignani *et al.*, 2009; van Oven *et al.*, 2011), pour la détermination de traits phénotypiques (Bouakaze *et al.*, 2009), ou encore pour la détection de pathologies anciennes (Watson and Lockwood, 2009). Une des différences essentielles entre toutes ces études réside dans le choix de la méthode analytique utilisée (la plupart étant actuellement fondée sur une détection par électrophorèse capillaire).

Au cours de ce travail nous avons choisi d'utiliser une technique de génotypage par spectrométrie de masse MALDI-TOF. Plus particulièrement, la plateforme MassArray 96 de Sequenom, qui est la seule permettant de travailler sur les acides nucléiques a été utilisée. L'application iPLEX[®] Gold proposée sur cette plateforme s'est avérée extrêmement intéressante pour les applications de la paléogénétique.

- Premièrement, la **sensibilité** de cette technique est extrêmement intéressante pour l'analyse d'ADN dégradé et présent en faible quantité. En effet, le test de sensibilité réalisé a révélé que des résultats fiables et reproductibles (avec un taux de plus de 80%) étaient obtenus à partir d'ADNnu jusqu'à un seuil d'environ 150 pg d'ADN par réaction. En ce qui concerne l'ADNmt un seuil est difficile à détecter, d'une part car nous ne disposons pas d'une méthode permettant de quantifier avec précision l'ADNmt et d'autre part parce que mis à part les échantillons dont la quantité d'ADN s'est révélée trop faible pour pouvoir être quantifiée, nous avons obtenu des résultats pour quasiment tous les échantillons. Par la suite, l'ensemble des résultats obtenus lors des investigations menées ont permis de confirmer la très haute sensibilité de la technique. En effet, comme décrit dans l'article Mendisco et al. 2011, le taux de génotypage est élevé, même pour des échantillons ne présentant que quelques picogrammes d'ADN. Ainsi, le génotypage par spectrométrie de masse MALDI-TOF nous a permis d'obtenir des informations pour des échantillons pour lesquels aucuns résultats n'avaient été obtenus par une autre technique. De plus, cette technique nous a permis à la fois de contrôler les résultats obtenus à partir du séquençage de la région HV-1, et à partir de la détermination des haplotypes Y, mais aussi d'affiner de manière importante la détermination des lignées. Pour les lignées paternelles cette technique s'avère essentielle étant donnée que les outils permettant de prédire les haplogroupes à partir des haplotypes sont peu fiables et peu précis (surtout en ce qui concerne les lignées non-européennes). Grâce à l'analyse de quelques positions il est possible de couvrir l'ensemble des lignées fondatrices des populations amérindiennes, et réaliser des comparaisons fines par la suite. En ce qui concerne les lignées maternelles la situation est quelque peu différente puisque dans ce cas les lignées peuvent souvent être déduites à partir des polymorphismes détectés par le séquençage de la région HV-1. Cependant, il peut arriver que cette région ne soit pas assez discriminante, et/ou parfois, un même haplotype peut correspondre à différents haplogroupes. Dans ce cadre, l'analyse de positions complémentaires permet de confirmer la déduction faite par la détermination de l'haplotype mais aussi d'affiner cette détermination. Au cours de cette étude l'ensemble des haplogroupes déduits à partir du séquençage et du génotypage étaient concordants démontrant la **fiabilité** de la technique iPLEX[®] Gold. De plus, l'ensemble des génotypes obtenus à partir d'extraits ou d'amplifications différentes étaient également concordants. Cette fiabilité est en partie due au fait qu'il s'agit de la seule technique permettant l'analyse d'une propriété intrinsèque de la molécule, sans aucun intermédiaire. Ainsi, par rapport par exemple au séquençage, pour les échantillons dégradés, les résultats peuvent être obtenus sans trop de réplifications (ce qui permet d'économiser les extraits d'ADN

ancien).

- Deuxièmement, cette méthode iPLEX[®] Gold permet un taux de **multiplexage** intéressant pour les analyses d'ADN ancien. Un taux de multiplexage allant jusqu'à 40 SNP par réactions a été décrit avec cette technique (Gabriel *et al.*, 2009), et il apparaît assez facile d'analyser simultanément entre 15 et 20 positions, ce qui est très intéressant en comparaison avec les autres techniques de génotypage. En effet, la plateforme MassArray possède un logiciel intégré permettant de créer les amorces PCR et SBE simultanément ce qui est un atout par rapport aux autres techniques pour lesquelles cette étape peut s'avérer fastidieuse. En plus de faciliter la création des amorces, ce logiciel permet de réajuster les réactions créées, en enlevant ou rajoutant un SNP assez facilement, ce qui rend cet outil assez **flexible**. Ainsi, nous avons pu recréer une réaction multiplexe permettant l'analyse de 15 SNP à partir de deux réactions de 11 SNP générées dans un premier temps. Cette réaction finale permet de discriminer à la fois les lignées principales maternelles et toutes les lignées paternelles, à partir d'un volume d'extrait d'ADN de 3,4 µL, ce qui permet une **optimisation maximale** des échantillons anciens. En effet, par comparaison avec les méthodes les plus fréquemment utilisées (séquençage de HVR-1 ou STR autosomaux et du chromosome Y) qui nécessitent un volume initial entre 5 et 10 µL d'ADN (le séquençage nécessitant 10 µL pour chaque fragment HV-1a et HV-1b analysés), la technique iPLEX Gold est très discriminante et fine, à partir d'un faible volume de départ.

- Troisièmement, le protocole est extrêmement **simple** et **rapide**. Les résultats pour 96 réactions peuvent être obtenus après une seule journée de manipulations. Le protocole est simple et comme les différentes étapes sont réalisées dans un même tube du début à la fin, cela évite les problèmes de contamination, de perte de produits ou d'erreurs lors des manipulations.

L'ensemble de ces caractéristiques ainsi que le faible coût relatif de chaque réaction (qui dépend du nombre de SNP intégrés dans chaque réaction) rendent la méthode iPLEX Gold très attractive pour les analyses d'ADNa. En effet, cette méthode pourra être intégrée à différents protocoles, aussi bien lorsqu'il s'agit de discriminer des lignées génétiques, que d'étudier des positions particulières intéressantes (lactase, traits phénotypiques...) (Lacan *et al.*, 2011).

II. Populations et peuplement des Andes méridionales.

1. Pratiques funéraires : le culte des ancêtres.

L'analyse du recrutement funéraire de différents sites anciens de la région nous permet de montrer qu'il n'existait pas un schéma unique en ce qui concerne les pratiques funéraires dans les Andes méridionales à la période préhispanique, mais que ces pratiques pouvaient être fondées sur différents critères.

A la vue des résultats obtenus, il semble que pour la population de Pampa Grande, située dans les sierras du nord-ouest argentin, les critères de parenté n'étaient pas privilégiés, pendant la période céramique initiale au moins. En effet, aucun lien de proche parenté n'a été mis en évidence pour les 21 individus inhumés dans ces cavités. Cependant, étant donné le faible nombre d'individus étudiés pour chaque cavité, il est possible que ces liens n'aient pas été révélés à cause d'un biais d'échantillonnage. Quoi qu'il en soit, ces populations du piedmont andin utilisaient les cavités naturelles comme site funéraire, séparant ainsi les morts des lieux d'habitation. Chaque cavité ayant révélé des types d'inhumations différents, il est possible d'imaginer que différents groupes ou clans de ce secteur utilisaient ces cavités afin d'y inhumer les membres de leur clan sans forcément prendre en compte les liens de proche parenté.

En ce qui concerne les populations des vallées, la situation est tout autre. Pour la majorité des groupes des vallées tempérées (Humahuaca et Calchaqui), les morts étaient encore étroitement liées au monde des vivants et très souvent, inhumés au sein même des lieux d'habitation. Des liens de proche parenté ont clairement pu être mis en évidence entre les individus anciens inhumés simultanément et/ou dans la même structure funéraire, pour l'agglomération ancienne de Los Amarillos. En plus de certains liens de proche parenté mis en évidence, il apparaît que les individus inhumés au sein d'une même enceinte d'habitation (souvent plus de 10 individus) partagent une lignée commune, le plus souvent maternelle. Le recrutement funéraire au sein de ces sites était effectué en fonction des liens de proche parenté ainsi que de l'appartenance à un même groupe familial. Ces pratiques funéraires, mises au jour au sein de sites datés de la période céramique tardive, pourraient être en relation avec les

ayllus, qui semble t-il à cette époque se développaient et se renforçaient. Cette période tardive voit l'augmentation des conflits entre clans (*chefferies*) d'une même région ou de régions voisines. L'unité de base de ces clans était l'*ayllus*, chaque *ayllus* étant formé des familles se réclamant d'un ancêtre commun, et possédant un territoire transmis de génération en génération. L'inhumation des individus d'une même lignée familiale au sein même du lieu d'habitation pouvait être le moyen d'assurer et de pérenniser l'autorité et le pouvoir des *ayllus* sur leur territoire.

Enfin, lorsque l'on considère le site de Cortaderas Derecha, occupé pendant la phase Inca, on s'aperçoit que le recrutement a été fait en fonction des liens de proche parenté strictement. Il est intéressant de noter la présence d'un crâne isolé dans cette sépulture qui pourrait correspondre à un « crâne trophée ». La mort semble t-il au même moment de ces trois jeunes adultes apparentés ainsi que la présence de ce probable « crâne trophée », laisse imaginer qu'un évènement soudain et violent soit la cause du décès de ces trois individus anciens. L'arrivée des Incas dans le nord-ouest argentin a engendré de nombreux conflits ainsi que le soulèvement d'une grande partie des populations locales. Ainsi, cette sépulture, au sein d'un site occupé par une population locale mais dominée par les Incas (Acuto, 2004), pourrait être le reflet des tensions entre différents groupes de la région à cette époque.

Ces différences marquées au niveau des pratiques funéraires peuvent avoir plusieurs explications. Tout d'abord, elles pourraient être le résultat des différences d'environnement. En effet, dans les sierras, et donc à Pampa Grande les habitations étaient généralement construites en matériaux périssables et les villages n'étaient donc pas permanents. L'inhumation des individus au sein de cavités naturelles peut traduire la volonté des populations que les morts soient dans un endroit favorable à la conservation des restes, pour avoir un lieu de recueillement durable. Deuxièmement, l'absence d'un regroupement familial à Pampa Grande peut traduire une organisation sociale ou une culture différente de celle des populations des vallées. Le système des *ayllus* s'est mis en place à partir de période céramique initiale mais s'est surtout développé pendant la période tardive, ce qui pourrait expliquer l'absence de liens de parenté pour les individus de Pampa Grande. D'autre part, la région de Pampa Grande n'a pas toujours été dans la même sphère d'interactions que la Quebrada de Humahuaca, ce qui pourrait expliquer des pratiques culturelles différentes. Enfin, nous ne pouvons exclure que les différences observées soient dues à un biais d'échantillonnage. L'analyse d'un plus grand nombre d'ensembles funéraires permettra de mieux comprendre les pratiques en place dans la région aux périodes précolombiennes.

2. Populations de la période céramique initiale.

La diversité génétique de la période céramique initiale a pu être étudiée au travers de deux sites archéologiques issus de deux étages écologiques distincts : le site, ou plutôt la sépulture de Flores située dans la zone des vallées tempérées, et le site de Pampa Grande situé dans les sierras à l'est de ces vallées.

Flores (Tilcara, Quebrada de Humahuaca) :

Les analyses réalisées sur les restes anciens exhumés de l'habitation dite de Flores, au sein du site de Tilcara, ont permis de déterminer que les deux individus anciens (F-1 et F-2) partageaient le même haplotype (NOA-B2 : 16111G – 16183C – 16189C – 16217C) rattaché à la lignée maternelle B2. Aucune correspondance n'a été trouvée, que ce soit avec des individus anciens d'autres sites, ou avec des individus de notre base de données. Il s'agit donc d'un haplotype unique, qui n'a pas encore été décrit à ce jour, et qui dérive par le gain d'une transversion (16111G) d'un haplotype central de la lignée B2 (*cf.* figure 25, partie 3 chapitre III.2.2.1).

Aucun résultat exploitable n'a été obtenu à partir de l'analyse de marqueurs nucléaires de ces deux échantillons (la quantité d'ADN présente dans les extraits testés n'a pas pu être déterminée, annexe 2). Malgré la répétition des amplifications, seuls quelques STRa ont été déterminés de manière reproductible pour l'échantillon F-1, permettant d'affirmer qu'il s'agit d'un individu masculin. Ni l'analyse des STR ni l'analyse des SNP du chromosome Y n'ont permis de déterminer clairement la lignée paternelle de cet échantillon masculin. En effet, l'absence d'amplification pour le marqueur Q1a3a_M3, permet seulement de dire que cet individu appartient à une lignée Q1a3, sans pouvoir préciser s'il s'agit du paragroupe Q1a3* ou Q1a3a*.

Pampa Grande (Sierras) :

L'excellente conservation de l'ADN au sein des échantillons de ce site a rendu possible l'analyse des liens de parenté, ainsi que la détermination des lignées maternelles et paternelles pour l'ensemble des 21 individus étudiés.

Lignées maternelles :

Parmi les 21 échantillons, 3 individus (Lam-3, C.V-4 et Lam-1) ont été rattachés à l'haplogroupe A2, ce qui correspond à une fréquence dans la population de 14,3%. L'haplotype de deux de ces individus a été déterminé de manière reproductible, et s'avère identique : NOA-A1. Cet haplotype qui présente les cinq mutations caractéristiques de la lignée A2 (16111T – 16223T – 16290T – 16319A – 16362C), peut être considéré comme fondateur, et se retrouve donc dans de très nombreuses populations amérindiennes, aussi bien contemporaines qu'anciennes, dans toutes les régions d'Amérique du Sud. Cet haplotype n'est donc pas très informatif quant à l'origine ou les affinités de ces individus. Cependant, il est très intéressant de noter que même si Lam-3 et C.V-4 partagent le même haplotype (NOA-A1), ils appartiennent à un sous-haplogroupe différent. En effet, comme présenté dans le tableau 13 (partie 3, chapitre I.4), C.V-4, qui possède l'allèle dérivé (A) en position 12007, mais l'allèle ancestral (C) en position 64, pourrait appartenir à la lignée ancestrale A2, contrairement à Lam-3 qui appartient à un des sous-haplogroupes A2 c-r.

Le reste des individus de Pampa Grande sont associés aux lignées maternelles B2 (47.6 %) et D1 (38.1 %). Aucun haplogroupe C1 n'a été détecté au sein de cette population ancienne. Les 10 individus appartenant à la lignée B2 sont distribués entre 5 haplotypes différents : NOA-B4, NOA-B5, NOA-B6, NOA-B7 et NOA-B10. Seul l'haplotype NOA-B10, qui correspond à un haplotype central, est retrouvé au sein de très nombreuses populations distribuées sur tout le sous-continent. Les quatre autres haplotypes n'ont donc *a priori* jamais été décrits jusqu'à aujourd'hui. Il est très intéressant de noter que les individus appartenant aux trois haplotypes NOA-B5 (Li-1, Ap-3), NOA-B6 (C.II-3) et NOA-B7 (C.II-1, C.II-4, C.V-2, C.V-3) partagent un motif particulier de la région HVR-1 en commun : 16145A – 16156A – 16157C. Aucune correspondance n'a été trouvée pour ce motif parmi les 2269 séquences constituant notre base de données de comparaison. Cependant, deux individus des Vallées présentent un motif identique. Il s'agit de Sc-2 et Sc-5, issus du site de Tero, situé dans le nord de la vallée Calchaqui. Le partage de ce motif particulier entre populations de régions différentes et de périodes différentes peut être le signe soit de l'origine commune de l'ensemble des populations du nord-ouest argentin, soit de mouvements de populations et donc de flux de gènes ayant accompagné les interactions mises en place entre ces deux régions.

Enfin, les derniers 8 échantillons sont répartis au sein de 3 haplotypes NOA-D1, NOA-D3 et

NOA-D4, associés à la lignée D1. L'haplotype NOA-D3, comme illustré sur la figure 27 (partie 3 chapitre III.2.2.1) correspond à un haplotype central, encore une fois distribué dans toute l'Amérique du Sud, et donc pas très informatif de l'origine ou des affinités des trois individus présentant cette séquence. Aucune correspondance n'a été trouvée dans notre base de données pour l'haplotype NOA-D4. L'haplotype NOA-D1 est partagé avec deux individus contemporains issus de populations des Andes centrales : 1 Tayacaja et 1 Yungay du Pérou (Fuselli *et al.*, 2003; Lewis *et al.*, 2007b), laissant supposer une origine andine de cet haplotype.

Lignées paternelles :

Pour les 12 individus anciens identifiés comme masculins, nous avons pu déterminer 11 haplotypes, dont 5 présentent un haplotype minimal (moins le loci DYS385) complet (H1, H4, H8, H15 et H16). Parmi ces 5 haplotypes, 2 (H1 et H16) n'ont trouvé aucune correspondance dans notre base de données, ni sur les bases de données disponibles en ligne (Y-HRD). Comme illustré sur la figure 26 (partie 3 chapitre III.2.2.1) Li-1 partage son haplotype H4 avec un Quechua de Bolivie (Gaya-Vidal *et al.*, 2011). Il s'agit d'un haplotype assez fréquemment retrouvé dans différent pays d'Amérique du Sud (21 correspondances avec des individus de populations latines américaines dans la base Y-HRD). L'haplotype H8 a été retrouvé dans la population d'Antioquia en Colombie (Builes *et al.*, 2006; Gaviria *et al.*, 2005), 4 fois au Brésil (Pereira *et al.*, 2007), 1 fois dans la province de Cordoba en Argentine (Fondevila *et al.*, 2003) et enfin chez un individu argentin (base de donnée de référence d'Argentine en ligne). Ces deux derniers haplotypes font partie du même paragroupe Q1a3*. Pour finir, l'haplotype H15, associé au paragroupe Q1a3a*, est partagé avec 5 individus des Andes septentrionales [2 Kichwa, 1 afro-équatorien (Gonzalez-Andrade *et al.*, 2009; Gonzalez-Andrade *et al.*, 2008), 2 individus de la population contemporaine de Quito en Equateur (Baeza *et al.*, 2007)], et 2 individus du Pérou (Iannacone *et al.*, 2005).

Il apparaît que de nombreux haplotypes aussi bien mitochondriaux que du chromosome Y sont restreints à cette population de Pampa Grande, pouvant indiquer soit un certain isolement de ce groupe ancien avec peu de flux de gènes d'autres régions pendant la période céramique initiale, ou une discontinuité avec les populations contemporaines de la région, engendré par la perte de certains haplotypes. Il faut cependant garder en mémoire que le nombre de données de référence disponibles étant faible, un biais d'échantillonnage pourrait être

responsable du nombre important d'haplotypes appelés uniques.

3. Populations de la période céramique tardive.

La période céramique tardive a été étudiée au travers de deux régions écologiques différentes : la puna pour laquelle seuls trois échantillons étaient disponibles, et les vallées tempérées pour lesquelles 34 échantillons de la Quebrada de Humahuaca et 6 échantillons de la vallée Calchaqui ont été analysés.

Doncellas (Puna) :

L'ADN que nous avons pu extraire de ces échantillons anciens du site de Doncellas n'était pas présent en quantité suffisante (entre 10 et 50 pg environ pour Don-1 et Don-2 ; annexe 2). En conséquence, nous n'avons pu déterminer que les haplotypes mitochondriaux de deux des individus (Don-1 et Don-2) et un haplotype Y partiel de Don-1. Grâce au typage par spectrométrie de masse, la lignée maternelle des trois individus a pu être définie ainsi que la lignée paternelle de Don-1.

L'individu masculin Don-1 présente l'haplotype NOA-B12, faisant partie de la lignée B2, et qui n'a été retrouvé qu'une seule fois parmi les populations anciennes des Andes méridionales. Cet haplotype semble être restreint aux populations andines puisqu'il a été détecté chez trois individus anciens de la population de Laramate au Pérou (Fehren-Schmitz *et al.*, 2010a; Fehren-Schmitz *et al.*, 2010b), 3 individus contemporains de Pérou et Bolivie (1 de la région de Titicaca et 2 de La Paz) (Afonso Costa *et al.*, 2010; Barbieri *et al.*, 2011), et 1 individu contemporain de la population des Coyas du nord-ouest de l'Argentine (Alvarez-Iglesias *et al.*, 2007), révélant une origine plutôt andine de cet haplotype.

L'origine des Coyas (aussi appelés Colla ou Kolla), qui occupent aujourd'hui le nord du Chili, le nord-ouest argentin et le sud de la Bolivie, serait sur les rives du lac Titicaca, en Bolivie, où durant la période finale de l'Empire de Tiwanaku¹, une douzaine de chefferies de langue

¹ Tiwanaku était une culture dont le cœur se trouvait dans les hautes terres sur les rives du lac Titicaca, et dont l'influence s'est répandue dans toute la zone centrale andine et une partie de la région des Andes méridionales.

indépendante luttait pour des conquêtes de pouvoir et de territoire. Les Coyas se composaient de deux entités ethniques : les Coya eux mêmes, qui détenaient le pouvoir politique et les Puquina, qui constituent la population ancestrale de la région. Il semble qu'à cette période certains Coya aient migré vers la région d'Atacama et vers le nord-ouest argentin (Isbell, 2008). Par la suite, au cours du XV^{ème} siècle, le territoire Coya bolivien a été envahi par les Incas. Afin de réprimer le soulèvement Coya, l'Inca Tupac Yupanqui a intégré les rebelles au sein des troupes armées Inca pour la conquête d'autres territoires. Ainsi, une importante partie de la population Coya a été transférée en tant que *mitimaes* dans le nord-ouest argentin, notamment pour l'exploitation minière. Les Coyas contemporains de ces régions se revendiquent comme descendants de cette migration forcée. Ils sont aujourd'hui le résultat du métissage de plusieurs ethnies, comme les Omaguacas, les Diaguitas d'Argentine, les Atacamenos du Chili, et les populations de la région de Titicaca (*cf.* figure 6, p. 61).

La présence de cet haplotype particulier NOA-B12 au sein de la population contemporaine des Coyas d'Argentine peut être la preuve des mélanges ayant eu lieu entre les populations ancestrales du nord-ouest de l'Argentine et les populations venues des hauts-plateaux péruviens et boliviens. Ainsi, il apparaît qu'au travers de ce métissage une partie du pool génétique maternel ancestral des Andes méridionales a été conservé.

L'individu masculin Don-2 partage son haplotype NOA-C2 avec les trois individus du site de Cortaderas dans la vallée Calchaqui et avec 1 individu (LA-18) du site de Los Amarillos dans la Quebrada de Humahuaca. En revanche, aucune correspondance n'a été retrouvée pour cette séquence dans notre base de données. Le partage de cette lignée entre individus de différents étages écologiques de la région argue en faveur de l'origine commune des populations de la région ou de flux géniques qui auraient accompagnés la circulation des biens et des matières premières.

Vallées tempérées (Quebrada de Humahuaca, Vallée Calchaqui) :

Une fois enlevés de l'échantillon les individus pour lesquels des liens de proche parenté ont été démontrés, nous avons pu déterminer les haplotypes mitochondriaux de 32 individus et les haplotypes Y de 21 individus masculins anciens, dont 12 sont complets pour l'haplotype minimal.

Lignées maternelles :

Les 32 individus pour lesquels les séquences mitochondriales ont été déterminées de manière reproductible se répartissent entre 15 haplotypes différents, dont 7 (NOA-A2, NOA-A3, NOA-B1, NOA-B3, NOA-B8, NOA-C2 et NOA-C4) sont uniques. Parmi ces 7 haplotypes uniques, portés par 9 individus, l'haplotype NOA-C2 est partagé avec 1 individu ancien du site de Doncellas. Comme présenté précédemment, l'haplotype NOA-B1 porté par les individus Sc-2 et Sc-5 du site de Tero présente le même motif particulier (16145A – 16156A – 16157C) qui a été trouvé pour 7 individus du site de Pampa Grande (haplotypes NOA-B5, NOA-B6 et NOA-B7). A la vue de la littérature disponible actuellement, il semble que ce motif particulier soit typique de cette région et/ou soit apparu dans cette région. Il est également intéressant de noter que l'individu LA-7 caractérisé par l'haplotype mitochondrial NOA-A2, qui dérive de l'haplotype fondateur de la lignée A2 par le gain d'une mutation en position 16125, est le seul des échantillons présentant un haplotype associé à la lignée A2 à être rattaché à l'haplogroupe ancestral A2. L'ensemble des autres individus présentant un haplotype rattaché à la lignée A2, appartiennent à un des sous-haplogroupes A2 c-r.

Concernant les haplotypes restant, 3 sont partagés avec de très nombreuses populations de toutes les régions de l'Amérique du Sud, étant donné qu'il s'agit des haplotypes fondateurs (NOA-A1, NOA-C3 et NOA-D3). L'haplotype NOA-A4 porté par l'individu ancien du site de Huichairas a déjà été décrit pour 2 individus contemporains de la population des Moxo, dans la partie amazonienne de Bolivie (Bert *et al.*, 2004), et 8 Guarani des hauts-plateaux brésiliens (Marrero *et al.*, 2007). Cet haplotype est le seul parmi l'ensemble de notre échantillon qui pourrait présenter une origine plutôt amazonienne ou évoquer des affinités avec les populations de l'est. Les haplotypes NOA-B9 et NOA-B10, qui sont centraux au sein de la lignée B2, sont distribués essentiellement dans les Andes centrales, au sein de populations contemporaines (Afonso Costa *et al.*, 2010; Barbieri *et al.*, 2011; Gaya-Vidal *et al.*, 2011) et anciennes (Fehren-Schmitz *et al.*, 2010a; Fehren-Schmitz *et al.*, 2010b; Kemp *et al.*, 2009), mais aussi dans les Andes méridionales (Alvarez-Iglesias *et al.*, 2007; Tamm *et al.*, 2007), le Chaco (Cabana *et al.*, 2006; Demarchi *et al.*, 2001b) et dans quelques groupes amazoniens (Bert *et al.*, 2004; Mazieres *et al.*, 2008; Williams *et al.*, 2002). L'haplotype NOA-B11, obtenu pour un échantillon du site de Tilcara (Til-1) apparaît comme typiquement andin. En effet, il a été décrit dans la région de Titicaca pour 9 individus contemporains (Barbieri *et al.*, 2011), il a également été retrouvé chez 20 individus de la région de La Paz (Afonso Costa *et al.*, 2010), 18 Aymaras et Quechuas contemporains de Bolivie (Gaya-Vidal

et al., 2011), 3 individus du Gran Chaco (Cabana *et al.*, 2006; Demarchi *et al.*, 2001b) et chez 3 Coyas du nord-ouest argentin (Alvarez-Iglesias *et al.*, 2007). Enfin, l'individu Sc-2 du site de Tero dans la vallée Calchaqui présente l'haplotype NOA-C1, associé à la lignée C1b, qui a déjà été retrouvé dans trois populations contemporaines : au sein des Guahibo du Venezuela (Vona *et al.*, 2005), chez les Ancash du Pérou (Lewis *et al.*, 2005) et dans la région de La Paz en Bolivie (Afonso Costa *et al.*, 2010).

Mis à part, l'haplotype NOA-A2, l'ensemble des lignées décrites pour cette population des vallées reflètent les liens avec les populations andines, suggérant une origine andine de cette population du nord-ouest argentin. De plus, le nombre important d'haplotypes uniques permet d'imaginer que les populations ancestrales de cette région ont évolué localement sans flux de gènes excessifs d'autres régions, conduisant à l'accumulation de ces lignées uniques. Une autre explication pourrait être la disparition de certaines lignées à cause d'évènements démographiques récents.

Lignées paternelles :

Les 12 individus anciens masculins pour lesquels un haplotype minimal complet a été déterminé se distribuent entre 11 haplotype différents, dont 3 sont uniques (H5, H11 et H12). Si aucune correspondance n'a été retrouvée dans notre base de données pour l'haplotype H9 (Ju-5), la base de données en ligne Y-HRD a révélé une correspondance parfaite avec 1 homme issu d'une population native américaine de la région amazonienne de Bolivie et 1 homme issu d'une population métissée du Brésil. L'haplotype H2, correspondant aux deux individus masculins du site de Cortaderas a déjà été retrouvé 1 fois dans la région amazonienne du Béni en Bolivie (Lee *et al.*, 2007; Tirado *et al.*, 2009) et 1 fois chez un afro-équatorien (Gonzalez-Andrade *et al.*, 2008). Les haplotypes H10 (LA-4 et FA-1) et H14 (Til-1) sont peu fréquents dans les populations contemporaines (correspondance avec 3 et 1 individu respectivement) et localisés essentiellement en Argentine (Chaco) ou en Colombie pour H10 (Romero *et al.*, 2008). Les haplotypes H3, H6 et H13 sont plus fréquemment retrouvés et présentent une diffusion large à quasiment tous les pays d'Amérique du Sud (Baeza *et al.*, 2007; Blanco-Verea *et al.*, 2010; Builes *et al.*, 2006; Gaviria *et al.*, 2005; Gaya-Vidal *et al.*, 2011; Gonzalez-Andrade *et al.*, 2008; Iannacone *et al.*, 2005; Lee *et al.*, 2007; Romero *et al.*, 2008; Tirado *et al.*, 2009; Toscanini *et al.*, 2008). Pour terminer, l'haplotype H7, obtenu pour l'échantillon Ju-4 de l'agglomération préhispanique de Juella, est

extrêmement fréquent (47 correspondances dans notre base de données et 75 sur Y-HRD) dans toute l'Amérique du Sud. Il pourrait correspondre à un haplotype fondateur, comme illustré sur le réseau phylogénétique généré à partir des haplotypes Y (*cf.* figure 26, p. 160).

Le nombre relativement important d'haplotypes uniques (non décrits jusqu'à ce jour) mitochondriaux et du chromosome Y laisse imaginer que les flux de gènes de régions voisines ont dû être limités tout au long de l'histoire de ces vallées, conduisant à l'accumulation de lignées spécifiques de cette région. Il est également possible que les bouleversements causés par l'incorporation de la région à l'empire Inca, puis par la conquête espagnole, aient conduit à la disparition de certaines lignées (notamment maternelles) et donc à une certaine discontinuité génétique entre la période céramique tardive et aujourd'hui. Ces deux phénomènes ne s'excluent pas mutuellement et peuvent chacun avoir influé sur la diversité de ces populations. Si ces explications restent plausibles à la vue des résultats obtenus, il faut cependant garder en mémoire que l'état actuel des connaissances concernant la diversité génétique de populations amérindiennes contemporaines et anciennes est limité, et il est donc possible qu'aucune correspondance n'ait été trouvée par manque de données.

Outre ces lignées particulières, les haplotypes partagés avec d'autres populations sud-américaines traduisent une histoire quelque peu différente des hommes et des femmes. En effet, si les haplotypes mitochondriaux (mis à part les haplotypes fondateurs) sont essentiellement partagés avec des populations andines (anciennes et contemporaines), les haplotypes Y sont distribués dans toute l'Amérique du Sud. La faible différenciation des lignées paternelles en fonction de la géographie laisse imaginer une diffusion assez récente sans processus micro-évolutifs marqués pour ces lignées. Au contraire le schéma observé pour les lignées maternelles, avec une origine clairement andine, suggère une différenciation géographique plus marquée. Il est possible que l'ensemble de la diversité des lignées maternelles n'ait pas encore été totalement définie et que dans les prochaines années des sous-haplogroupes spécifiques de certaines régions soient découverts (comme le sous-haplogroupe Q1a3a4 spécifique des populations des hauts-plateaux péruviens et boliviens).

4. Les Andes méridionales : entre disparités culturelles et homogénéité génétique.

Les analyses menées à la fois sur les pratiques funéraires ainsi que sur les lignées maternelles et paternelles nous permettent de proposer un schéma général de la dynamique de peuplement des Andes méridionales ainsi que de l'évolution des populations amérindiennes locales.

Une certaine divergence a été mise au jour entre populations des Vallées et de Pampa Grande. Tout d'abord, les pratiques funéraires de ces deux groupes apparaissent très différentes. De plus, l'observation directe des fréquences des lignées maternelles et paternelles de ces deux groupes laissait supposer à première vue une divergence génétique relativement importante. En effet, la fréquence élevée de la lignée D1 (38,1 %) au sein du groupe de Pampa Grande, ainsi que l'absence totale de l'haplogroupe C1 tranche avec les groupes des vallées caractérisés par une fréquence élevée de la lignée A2 (42,5 %), la présence à 15 % de la lignée C1 et une faible fréquence de la lignée D1. En ce qui concerne les lignées paternelles, la différence semble également marquée, avec un paragroupe majoritaire différent pour chacune des deux populations. Cette divergence génétique (fondée sur les haplogroupes mitochondriaux) reste cependant modérée ($F_{ST} = 0,108$), et peut être expliquée par la distance géographique entre ces deux régions et/ou par le temps écoulé entre ces deux périodes (entre 300 et 700 ans de différences). Il est cependant difficile de déterminer pour l'instant lequel de ces deux facteurs a le plus influé sur les changements de fréquences de lignées maternelles observés.

La divergence modérée observée en étudiant les lignées est un peu moins importante lorsque l'on analyse les haplotypes mitochondriaux ($F_{ST} = 0,091$) et n'est pas significative pour les haplotypes Y. De plus, les deux groupes présentent des caractéristiques identiques, avec un grand nombre d'haplotypes uniques, des lignées maternelles essentiellement partagées avec des populations andines et des lignées paternelles partagées avec des populations de toute l'Amérique du Sud. Ces haplotypes uniques présents aussi bien à la période céramique initiale que tardive traduisent probablement une évolution *in situ* des groupes depuis leur arrivée dans la région, avec peu de flux de gènes d'autres régions. Cela peut également être expliqué par la disparition de lignées à cause d'événements démographiques récents, tels que l'annexion de la région par les Incas et/ou la colonisation européenne.

Le partage d'un motif mitochondrial particulier entre individus de deux étages écologiques différents et d'époques différentes impliquent deux hypothèses : (i) les différentes populations de la région des Andes méridionales sont issues d'un pool génétique ancestral commun, (ii) les nombreuses interactions nécessaires à la subsistance des groupes de la région ont pu être accompagnées d'échanges génétiques.

Enfin, nous pouvons constater que, mis à part les haplotypes fondateurs, seuls deux haplotypes mitochondriaux de la lignée B2 sont retrouvés dans une population contemporaine du nord-ouest de l'Argentine : les Coyas. Ceci traduit une certaine discontinuité des lignées maternelles entre les populations anciennes et contemporaines de la région, imputable très probablement au remplacement des populations de la région, engendré par la politique Inca, mais aussi par l'arrivée des européens et la dépopulation importante qui en a résulté. Les haplotypes Y sont plus nombreux (6/15) à être retrouvés actuellement dans les populations du nord-ouest argentin, ce qui implique une continuité un peu plus marquée du pool génétique masculin.

Ces différentes observations suggèrent une histoire évolutive assez complexe des populations des Andes méridionales. Même si ces populations semblent avoir une origine commune, il est possible que la dérive génétique ait conduit à une certaine différenciation des lignées maternelles entre les différents étages écologiques, aux périodes pré-céramique et céramique initiale. Par la suite, avec l'augmentation démographique de la période intermédiaire et l'intensification des interactions au travers de la région, le taux de migration a pu augmenter, conduisant à des flux géniques plus intenses. Pour finir, l'annexion de la région par les Incas et l'arrivée des Européens a conduit à des bouleversements démographiques profonds, conduisant à la disparition de certains haplotypes et à une baisse de la diversité des haplogroupes (lignée B2 prépondérante).

Ces premières hypothèses émises sont bien évidemment à tester avec des échantillons présentant des effectifs plus importants et issus d'aires géographiques et de périodes différentes.

III. Apports pour la compréhension de la dynamique de peuplement de l'Amérique du Sud.

L'objectif de ce type d'études à terme, outre la compréhension de processus micro-évolutif au sein de régions données, est de pouvoir mieux appréhender la dynamique de peuplement de l'Amérique du Sud. Actuellement, de nombreuses théories ou hypothèses, concernant notamment le peuplement de l'Amérique du Sud, ont été émises sur le fondement que le complexe andin (comprenant les populations andines depuis la Colombie jusqu'au nord du Chili et le nord-ouest de l'Argentine) formerait une entité génétiquement différenciée du reste de l'Amérique du Sud, et plus particulièrement de l'Amazonie.

Les données obtenues au cours de ce travail sur des populations anciennes faisant partie de ce complexe andin nous permettent de discuter des différentes hypothèses émises concernant la colonisation initiale de l'Amérique du Sud et dans un premier temps de l'histoire évolutive des populations de ce complexe andin.

1. Les Andes méridionales au sein du complexe andin.

Comme nous avons pu le vérifier au cours de ce travail, il apparaît que l'ensemble des populations andines contemporaines présentent des caractéristiques communes telles qu'une divergence inter-populationnelle faible, une diversité génétique intra-populationnelle élevée, des indices significatifs de populations en expansion et une forte fréquence de lignée maternelle B2. Comme cela a déjà été décrit par de nombreux auteurs, l'homogénéité génétique des populations andines contemporaines semble donc incontestable à ce jour (Fuselli *et al.*, 2003; Luiselli *et al.*, 2000; Tarazona-Santos *et al.*, 2001). Plusieurs hypothèses ont été émises pour expliquer l'origine de cette homogénéité génétique, parmi lesquelles, un pool génétique ancestral commun (qui aurait pu être différent de celui des populations amazoniennes) et/ou des flux géniques récurrents dans tout le complexe andin. En dépit de l'augmentation du nombre d'études concernant des populations amérindiennes ces dernières années (Afonso Costa *et al.*, 2010; Barbieri *et al.*, 2011; Bisso-Machado *et al.*, 2011; Fehren-Schmitz *et al.*, 2010b; Gaya-Vidal *et al.*, 2011; Toscanini *et al.*, 2011), la mise en place de la

diversité observée actuellement n'a toujours pas été élucidée.

L'arrivée des européens au cours du XVI^{ème} siècle a provoqué de profonds bouleversements aussi bien dans la répartition des populations sur le territoire, que dans la densité des populations natives et leur pool génétique (Hunley and Healy, 2011). Il apparaît donc indispensable de s'intéresser aux populations anciennes directement, afin de s'affranchir des biais causés par les populations contemporaines, et ainsi essayer de déterminer si les caractéristiques décrites pour ces populations étaient déjà établies aux périodes préhispaniques.

Les quelques études paléogénétiques réalisées jusqu'à aujourd'hui ont permis de démontrer qu'il existe une continuité génétique importante dans les Andes centrales depuis au moins la période céramique intermédiaire (600 – 900 ap. J.-C.). Les populations anciennes des Andes centrales présentent en effet des caractéristiques similaires (diversité génétique élevée, fréquence élevée d'haplogroupe B2), ainsi qu'une divergence faible avec les populations contemporaines de cette région, ce qui a conduit les auteurs à proposer que ces populations anciennes pouvaient correspondre aux ancêtres des populations andines actuelles (Fehren-Schmitz *et al.*, 2010a; Fehren-Schmitz *et al.*, 2010b; Kemp *et al.*, 2009; Lewis Jr *et al.*, 2007). La période intermédiaire voit le développement des premières sociétés étatiques complexes telles que Wari puis Tiwanaku, qui pratiquaient déjà la relocalisation de populations et ont étendu leur influence sur de larges territoires au travers de vastes réseaux d'interactions (Tung, 2008). Ces interactions, qui ont pu être accompagnées de flux de gènes, ainsi que les mouvements de populations pourraient expliquer l'homogénéisation des populations des hauts-plateaux andins, depuis des périodes anciennes. Cependant, comme cela est souligné par les auteurs, il est également possible que l'effectif des échantillons anciens analysés ne soit pas suffisamment important pour déceler des discontinuités du pool génétique.

La situation dans certaines aires en marge des hauts-plateaux des Andes centrales apparaît comme étant significativement différente. La côte Pacifique péruvienne a vu se développer et se succéder dès le VII^{ème} siècle avant notre ère, des civilisations telles les Paracas (800 – 200 av. J.-C.) puis les Nasca (200 av. J.-C. – 600 ap. J.-C.), qui possédaient toutes les deux des relations avec les populations des hauts-plateaux (Proulx, 2008). Malgré des affinités culturelles et idéologique évidentes, ces populations anciennes de la côte sont très divergentes génétiquement des populations andines anciennes des hauts-plateaux et contemporaines de la

côte. Ces résultats impliquent d'une part qu'entre ces deux aires géographiques les interactions n'ont pas été accompagnées de flux de gènes, et d'autre part qu'un évènement récent a conduit à modifier le pool ancestral des populations de la côte péruvienne (Fehren-Schmitz *et al.*, 2010a). De même, la population ancienne de la vallée d'Arica dans le nord du Chili a vu ses fréquences de lignées maternelles changer significativement entre la période archaïque et la période tardive (Moraga *et al.*, 2005).

Le complexe andin apparaît donc beaucoup moins homogène aux périodes préhispaniques qu'actuellement, ce qui implique que toutes les populations andines n'aient pas suivi le même processus évolutif. Dans ce cadre, il semblait intéressant d'étudier la région des Andes méridionales, pour laquelle aucunes données d'ADNa n'étaient disponibles, afin de mieux comprendre les relations entre populations anciennes des Andes, et la mise en place de la diversité génétique observée actuellement.

1.1. Interactions entre populations andines anciennes : l'impact des grandes civilisations sur les Andes méridionales.

Comme cela a déjà été présenté, les Andes méridionales sont caractérisées par une mobilité importante des populations, et des interactions nombreuses entre les différentes aires géographiques. Ces caractéristiques peuvent également s'appliquer aux populations des Andes Centrales, où la mobilité des groupes et le besoin d'interactions entre différentes régions, nécessaires à la gestion des ressources, ont été évoqués comme faisant partie des raisons qui ont conduit à la mise en place de sociétés aussi complexes¹. Certaines de ces sociétés ont étendue leur influence et/ou leur territoire jusqu'au nord des Andes méridionales. Il était donc intéressant d'essayer de comprendre le type d'influence en place entre ces deux grandes régions et si le développement des groupes des Andes méridionales a suivi un développement parallèle à celui de ces populations des Andes centrales.

La civilisation Tiwanaku (V^{ème} – XI^{ème} siècle) est la première à avoir étendu son influence de manière significative jusque dans l'extrême sud du complexe andin. Des interactions ayant

¹ Seules les populations de deux régions en Amérique ont atteint un degré de complexité aussi élevé : la Méso-Amérique avec le développement des Aztèques puis des Mayas, et les Andes centrales avec le développement de différentes civilisations telles que les Nasca, Paracas, Tiwanaku, Wari et les Incas.

mené à une forte influence du point de vue culturel et idéologique sont avérées, notamment dans la région d'Atacama (Stovel, 2008) et dans le nord de la puna (Albeck, 2001; Albeck and Ruiz, 2003). Cependant il n'est pas encore établi si ces interactions étaient accompagnées ou non de migrations d'individus et donc de flux de gènes.

Malheureusement, la conservation des échantillons de la région de San Pedro de Atacama, qui étant distribués sur des périodes pré-Tiwanaku et Tiwanaku auraient permis d'obtenir une première vision concernant cette problématique, était trop mauvaise pour pouvoir obtenir des résultats exploitables. En ce qui concerne les populations du nord-ouest argentin, comme cela a déjà été discuté précédemment, il semble que les flux de gènes depuis les Andes centrales aient été limités pendant les périodes céramiques initiale à tardive. Même si nous pouvons observer que la population des Vallées présente une affinité assez forte avec la population ancienne d'Arica (période céramique initiale et intermédiaire) dans le nord du Chili (Moraga *et al.*, 2005), les deux populations anciennes étudiées sont relativement divergentes des populations anciennes des Andes centrales, laissant supposer un certain isolement des populations des Andes méridionales par rapport aux Andes centrales. A partir des restes matériels archéologiques, et notamment de la céramique, il a été proposé que l'influence de Tiwanaku sur le nord-ouest argentin n'aurait été qu'indirecte, par le biais d'interactions avec la région de San Pedro de Atacama et le nord du Chili. Les résultats obtenus sont en faveur de cette hypothèse et montrent que l'influence des sociétés de l'altiplano sur les populations du nord-ouest argentin s'est probablement faite sans flux de gènes massifs. De plus, il s'avère que les interactions pendant les périodes céramiques dans l'aire méridionale se faisaient beaucoup plus dans le sens est/ouest, pour relier le désert aux plaines fertiles, que dans le sens nord/sud (Nielsen, 2008), ce qui confirme un peu plus cette hypothèse.

La seule grande civilisation andine qui semble avoir influé significativement sur le pool génétique des populations anciennes de la région méridionale est la civilisation Inca. La faible continuité génétique au niveau mitochondrial que nous avons observé entre les populations anciennes et contemporaines des Andes méridionales pourrait être expliquée par les déplacements de populations organisés par les Incas, avec entre autre les populations Coyas de la région de Titicaca installées de force dans le nord-ouest argentin. Comme cela a déjà été évoqué par de nombreux auteurs, il est possible que ces événements démographiques récents aient engendré ou au moins accentué la forte fréquence de lignée maternelle B2 observée actuellement (Barbieri *et al.*, 2011; Bert *et al.*, 2004; Corella *et al.*, 2007).

1.2. Histoire évolutive et mise en place du complexe andin.

Nous pouvons constater, à l'aide des différentes analyses réalisées, que la diversité andine observée actuellement n'est pas le reflet de la diversité ancestrale. En effet, mis à part dans les Andes centrales où une continuité génétique semble en place depuis la période intermédiaire, les autres principales aires du complexe andin présentent des caractéristiques qui suggèrent une certaine discontinuité au cours du temps.

De manière générale, il apparaît que la diversité globale des haplogroupes mitochondriaux semble avoir décliné entre les périodes anciennes et aujourd'hui (Raff *et al.*, 2011). En effet, les populations les plus anciennes présentent généralement des fréquences plus élevées d'haplogroupe D et moins élevées d'haplogroupe B en comparaison avec les périodes plus tardives (Demarchi *et al.*, 2001a; Fehren-Schmitz *et al.*, 2010a; Moraga *et al.*, 2005). De plus, les populations anciennes des différentes régions du complexe andin présentent une divergence génétique plus marquée en comparaison avec la faible divergence observée actuellement au sein du complexe.

En ce qui concerne plus particulièrement les deux populations anciennes étudiées au cours de ce travail, elles présentent une diversité génétique (fondée sur la diversité haplotypique mitochondriale) parmi les plus faibles de toutes les populations andines, ainsi qu'un nombre important d'haplotypes (mitochondriaux et Y) non retrouvés dans les populations actuelles et notamment dans les populations des Andes méridionales. Parmi ces haplotypes, il est possible de noter la présence d'un motif particulier retrouvé dans les différents étages de la région, mais dans aucune autre population d'Amérique du Sud. Ainsi, même si la divergence de cette région par rapport au reste du complexe andin est moins flagrante que pour la côte péruvienne par exemple, il semble tout de même qu'elle n'ait pas suivi la même histoire évolutive que les populations des hauts-plateaux des Andes centrales.

S'il apparaît probable que les caractéristiques des populations étudiées soient révélatrices d'une histoire évolutive en marge du reste du complexe andin, nous ne pouvons exclure que ces résultats soient dus à un biais d'échantillonnage. En effet, le faible effectif des populations analysées peut entraîner un biais important dans les calculs de fréquences ou de diversité. De plus, l'absence de correspondances dans notre base de données peut être due au faible nombre de séquences de populations amérindiennes disponibles dans la littérature. Seules des études

ultérieures, conduisant à l'augmentation du nombre de données de comparaisons, permettront de confirmer si les haplotypes présentés dans cette étude comme uniques sont effectivement spécifiques des populations du nord-ouest argentin ou pas.

En considérant que l'échantillonnage ne soit pas responsable entièrement des particularités génétiques observées, plusieurs hypothèses concernant l'évolution des populations andines peuvent être proposées.

- Premièrement, la divergence des populations des Andes méridionales par rapport aux populations anciennes des Andes centrales pourrait être expliquée, comme cela a été le cas pour les populations de la côte péruvienne (Fehren-Schmitz *et al.*, 2010a), par une colonisation initiale par des voies de migration différentes. Cependant, il apparaît que les haplotypes mitochondriaux partagés laissent supposer une origine commune des populations des Andes méridionales et centrales. Dans le cas d'une origine commune, cette divergence légère pourrait être expliquée par un certain isolement de la région aux périodes pré-céramique et céramique initiale, ayant pu favoriser une évolution par dérive génétique.

- Deuxièmement, les caractéristiques de ces populations anciennes peuvent être expliquées par une discontinuité génétique avec les populations contemporaines. Des événements démographiques récents ont pu causer la perte de certaines lignées (notamment maternelles). Deux événements prépondérants peuvent être responsables de la perte de diversité ancestrale : l'intégration de la région à l'empire Inca et la colonisation européenne. Ces deux événements ont provoqué de profonds bouleversements sur une période de temps très réduite (une centaine d'années). Cette époque est marquée à la fois par des conflits violents et quasiment ininterrompus depuis l'arrivée des incas jusqu'au XVII^{ème} siècle, ainsi que par de grands mouvements de populations natives venant essentiellement de l'altiplano à la période incaïque mais aussi des régions du Chaco ou de la plaine amazonienne à la période coloniale. Ainsi, non seulement l'effectif des populations natives de la région a baissé drastiquement mais en plus, les populations actuelles de cette région sont le reflet du métissage de groupes venant de régions très différentes. Il est cependant encore trop difficile d'évaluer de manière fiable, étant donnée le faible nombre de données disponibles, l'impact de chacun des événements sur le pool génique des populations ancestrales des Andes méridionales.

Pour résumer, le peuplement des Andes méridionales s'est fait apparemment à partir du même pool génique que celui des Andes centrales, probablement donc au même moment. Par la suite, il est possible que ces deux grandes régions aient évolué indépendamment (les premières grandes avancées techniques, et sociétés complexes sont apparues beaucoup plus tôt dans les Andes centrales que méridionales), conduisant à une légère divergence génétique. Enfin, à la période Inca, une partie du pool ancestral de la région méridionale a pu être remplacé conduisant à l'homogénéité observée actuellement dans tout le complexe andin.

2. Populations andines vs amazoniennes : une frontière de moins en moins évidente.

Même si les hypothèses émises concernant la dynamique de peuplement du complexe andin sont fondées sur un faible nombre d'échantillon et restent à tester par de nombreuses autres études, elles permettent d'imaginer qu'aux périodes préhistoriques le complexe andin ne présentait pas les caractéristiques qui le définissent actuellement.

Les résultats que nous avons obtenus sur les populations anciennes des Andes méridionales montrent que les hypothèses de peuplement concernant l'Amérique du Sud ne peuvent être fondées uniquement sur la diversité génétique des populations amérindiennes contemporaines. En effet, le pool génique des populations anciennes de diverses régions (côte péruvienne, nord Chili, nord-ouest argentin) a été remanié au cours du temps et les relations entre populations anciennes andines ne semblent pas aussi évidentes et simples que celles proposées jusqu'alors (différences de flux de gènes entre régions). Ainsi, il est encore très difficile de comprendre clairement la dynamique de peuplement du complexe andin.

De plus, les analyses réalisées montrent que certaines populations anciennes, dont la population des Vallées, présentent des affinités avec des populations contemporaines amazoniennes. Ainsi, il est possible et même probable que des interactions entre populations andines et amazoniennes aient eu lieu. Les populations du Chaco notamment sont très proches des populations anciennes des Vallées et de Pampa Grande, suggérant des liens dus à une origine commune ou à des flux géniques entre les deux régions. Ainsi, des interactions et des échanges entre populations de ces deux mondes, andin et amazonien, ont dû avoir lieu tout au long de l'histoire de l'Homme en Amérique du Sud, mais qui ont pu être masqués par des événements plus récents.

Certains auteurs ont même évoqué que les populations anciennes des hauts-plateaux andins aient pu être originaires de l'Amazonie (Moraga *et al.*, 2005; Rothhammer and Dillehay, 2009; Rothhammer and Silva, 1989) mais cela paraît assez peu probable à la vue de nos résultats. En effet même si quelques similitudes peuvent être observées du point de vue des fréquences d'haplogroupes, les haplotypes partagés sont clairement en faveur d'une origine typiquement andine.

Cependant, il faut noter que les données disponibles sur les populations amazoniennes, aussi bien anciennes que contemporaines, sont trop peu nombreuses et qu'elles ne permettent donc pas de discuter de manière fiable de ces hypothèses.

Si les données récoltées au cours de ce travail nous ont permis de discuter de l'évolution du pool génétique des populations anciennes des Andes méridionales depuis la période céramique initiale, elles conduisent à nous poser encore plus de questions sur la dynamique de peuplement de l'Amérique du Sud. De nombreuses études seront donc nécessaires pour comprendre de quelle manière s'est déroulée la diffusion des premiers migrants amérindiens à toutes les régions de l'Amérique du Sud.

Conclusion générale et perspectives.

La paléogénétique apparaît aujourd'hui comme une discipline essentielle pour appréhender les problématiques de peuplement, au même titre que des disciplines comme la génétique des populations, l'archéologie, l'anthropologie ou la linguistique, qui associées permettent des études multidisciplinaires indispensables à l'établissement d'hypothèses robustes.

Grâce à ce type d'études, ces dernières années, la compréhension de la dynamique de peuplement de l'Amérique s'améliore considérablement. Néanmoins, certaines régions restent très peu étudiées empêchant d'avoir une vision précise des modalités de diffusion de l'Homme à toutes les régions d'Amérique et plus particulièrement en Amérique du Sud.

Les difficultés engendrées par les particularités de l'ADN ancien, telles que la destruction du matériel ou l'incertitude quant aux résultats entre autres, peuvent freiner la mise en place d'analyses. Cependant, les avancées récentes des techniques et des méthodologies d'analyse permettent aujourd'hui, comme cela a été démontré dans ce travail, d'obtenir une quantité intéressante d'informations de qualité à partir de petits échantillons, comme une seule dent par exemple.

Parmi les avancées méthodologiques nous pouvons citer la spectrométrie de masse, que nous avons pu mettre en place au cours de ce travail au travers de la plateforme Mass Array (Sequenom). Cette plateforme de génotypage s'est révélée être un outil extrêmement intéressant pour l'étude d'ADN dégradé et présent en très faible quantité. En effet, cette technique se révèle être parmi les plus sensibles des techniques aujourd'hui utilisées en routine pour les études paléogénétiques, et extrêmement fiable et cela à partir de seulement quelques μL (entre 1,6 et 3,4 μL) d'extrait d'ADN. Dans le cadre de l'étude de populations amérindiennes cette technique se révèle particulièrement intéressante car, le nombre de lignées fondatrices étant réduit, une seule réaction permet de discriminer à la fois l'ensemble des lignées maternelles et paternelles.

A une période où les méthodes sont de plus en plus pointues et permettent d'obtenir une quantité d'informations très importante, la spectrométrie de masse apparaît comme une technique alternative très intéressante pour les études de paléogénétique.

Grâce à la mise en place d'une méthodologie rigoureuse et fondée sur l'analyse de plusieurs marqueurs génétiques complémentaires, ce travail de thèse avait plusieurs objectifs : premièrement, caractériser la variabilité génétique de populations anciennes de la région des Andes méridionales pour laquelle aucune données n'étaient disponibles à ce jour et deuxièmement, appréhender la dynamique de peuplement de cette région ainsi que l'histoire

évolutive des populations qui l'ont peuplée, dans le cadre d'une problématique plus générale qui est celle du peuplement de l'Amérique du Sud.

L'excellente conservation de l'ADN au sein de la majorité des échantillons du nord-ouest argentin a permis de discuter des différentes questions posées, grâce aux données génétiques authentiques obtenues pour 3 individus datant de la période tardive et issus de la puna de Jujuy, une population des vallées tempérées (41 individus) ayant occupé la région pendant la période tardive également et une population des sierras (21 individus) ayant peuplé cette zone pendant la fin de la période initiale.

1. Dans un premier temps, nous avons pu mettre en évidence des différences au niveau des pratiques funéraires entre les différents groupes étudiés. Certaines de ces pratiques funéraires, caractéristiques des populations tardives des vallées, pourraient être le reflet de l'importance des clans et des chefferies (*ayllus*) dans l'organisation sociale et politique des groupes de cette époque.

Un nombre très important de restes squelettiques ayant été mis au jour au sein des nombreux sites archéologiques de la région, des études complémentaires pourraient être menées afin d'essayer d'estimer l'origine de telles pratiques et les raisons des différences observées avec les régions voisines.

2. La variabilité génétique décrite pour ces groupes anciens correspond à la variabilité attendue pour des populations amérindiennes. En effet les quatre lignées maternelles fondatrices majoritaires ainsi que la lignée paternelle fondatrice principale ont été retrouvées au sein des populations étudiées. Quelques particularités intéressantes ont tout de même été décelées. Tout d'abord, le nombre d'haplotypes pour lesquels aucune correspondance n'a été observée dans nos bases de données de référence est très important. Cela souligne qu'une grande partie de la variabilité génétique des populations amérindiennes est probablement encore inconnue et que de nombreuses études seront encore nécessaires pour vraiment connaître la diversité présente et passée.

3. Nous avons pu détecter un motif particulier sur la région HVR-1 partagé entre populations d'aires géographiques et de périodes différentes, menant à proposer une origine commune pour l'ensemble des populations des Andes méridionales et/ou des flux géniques importants qui auraient pu accompagner les interactions intenses qui ont caractérisé les populations de cette région tout au long des périodes préhispaniques. Le pool génétique, en

particulier maternel, semble avoir été remplacé en partie, à partir de période incaïque. En effet, une certaine discontinuité avec les populations contemporaines de la région est observée, qui pourrait être le résultat de l'établissement dans la région de populations des hauts-plateaux des Andes Centrales (malgré un certain métissage avec les populations locales : Coyas), envoyée par les Incas, ainsi que de la colonisation européenne qui a engendré d'importants bouleversements démographiques.

Ces hypothèses devront être testées et approfondies par des études futures. Pour cela l'effectif de la cohorte d'individus devra être augmenté mais aussi d'autres aires écologiques devront être étudiées, tout comme d'autres périodes. En effet, nous nous sommes intéressé au cours de ce travail à des périodes relativement récentes (V^{ème} – XV^{ème} siècles), il serait donc intéressant dans le futur de pouvoir analyser des individus de la période archaïque (entre 10 000 et 2 000 ans av. J.-C.) afin de connaître l'origine des premiers groupes de cette région, mais aussi comprendre la mise en place de sociétés sédentaires et pratiquant l'agriculture. Des études paléogénétiques concernant ces périodes permettraient potentiellement de savoir si ces développements sont apparus après une évolution des individus sur place, ou par l'arrivée de groupes des Andes centrales (où ces phénomènes sont apparus un peu plus tôt que dans les Andes méridionales).

4. Les données obtenues sur cette région nous ont également permis de discuter des hypothèses émises actuellement pour le peuplement de l'Amérique du Sud. Associés aux récents résultats obtenus pour des populations anciennes de la côte péruvienne, nous pouvons proposer que l'homogénéité génétique des populations andines contemporaines est le résultat d'évènement récents et non pas le reflet de la diversité passée. Ainsi, les hypothèses fondées sur l'existence d'une telle homogénéité s'en retrouvent caduques. Malgré cela, il apparaît probable que l'origine des populations des Andes centrales et méridionales soit unique, mais que par la suite, ces deux régions aient évolué indépendamment, sans flux géniques massifs. Ainsi, la dynamique de peuplement du complexe andin apparaît complexe, d'autant plus que certains auteurs proposent que la côte Pacifique ait été peuplée par une voie différente de celle des hauts-plateaux andins. Un travail important devra donc être fait dans le futur afin de caractériser au mieux la diversité génétique de ces populations anciennes andines et d'affiner les hypothèses proposées à ce jour.

La colonisation de l'Amérique du Sud, par les premiers amérindiens, après le passage de l'isthme du Panama reste donc toujours autant problématique. Les relations entre populations

andines et amazoniennes sont encore floues, et les régions à la jonction entre ces deux mondes devront faire l'objet d'études approfondies dans le futur, afin de mieux appréhender ces problématiques.

Pour appréhender cette problématique, plusieurs perspectives sont envisageables :

- Premièrement, il serait intéressant de pouvoir évaluer de manière fiable l'effet des barrières géographiques, et notamment des Andes, sur la diffusion des gènes et sur la distribution des populations sur le territoire. En particulier, dans la région des Andes méridionales, il serait intéressant de pouvoir évaluer les relations entre les populations de la côte chilienne et les populations que nous avons étudiées au cours de ce travail. L'hypothèse de deux migrations séparées ayant peuplé d'une part la côte Pacifique et d'autre part les hauts-plateaux andins a été évoquée par plusieurs auteurs et pourrait pas ce biais être testée.

- Deuxièmement, l'hypothèse d'une origine des populations des hauts-plateaux brésiliens depuis le nord-ouest argentin pourrait être testée grâce à l'étude des lignées maternelles et paternelles de populations anciennes des périodes archaïques de ces régions. De même, il serait intéressant de pouvoir caractériser la diversité génétique ancienne des populations natives du Gran Chaco. Dans les populations contemporaines, des affinités génétiques entre populations du Chaco et des Andes méridionales sont évidentes, mais il est encore difficile de dire si elles sont seulement dues à des événements récents (et notamment l'arrivée des jésuites dans le Chaco aux XVI^{ème} et XVII^{ème} siècles) ou si elles pré-datent l'arrivée des européens.

- Troisièmement, il semble qu'un travail important sur les marqueurs du chromosome Y pourrait apporter de nouvelles perspectives intéressantes. En effet, il apparaît que la phylogénie évolue rapidement et il est probable que des lignées spécifiques de régions particulières et pouvant révéler des mouvements de migration soient découverts dans les prochains temps. Ainsi, l'étude de la lignée Y sera probablement un enjeu important dans les prochaines années.

Bibliographie

A.

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, and Bandelt HJ (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764.
- Acreche N, Albeza MV, Caruso GB, Broglia VG, and Acosta R (2004) Diversidad biológica humana en la provincia de Salta. *Cuadernos FHyCS-UNJu* 22:171-194.
- Acuto F (2001) Domestic life under Inka domination : the case of the Calchaqui valley (Argentina). 66th annual meeting of the Society for American Archaeology.
- Acuto F, Gigliotti V, Jacob C, Podesta M, Pupkin P, and Sprovieri M (2004) Bajo el pulgar imperial : primeros estudios en el sitio Inka de Cortaderas (Valle Calchaqui, Salta). *Mosaico. Trabajos en Antropología Social y Arqueología*, pp. 1-24.
- Acuto FA (2007) Fragmentacion vs. integracion comunal : Repensando el Periodo Tardio des Noroeste Argentino. *Estudios Atacamenos : Arqueología y Antropología Surandinas* 34:71-95.
- Adcock GJ, Dennis ES, Easton S, Huttley GA, Jeremiin LS, Peacock WJ, and Thorne A (2001) Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proc Natl Acad Sci U S A* 98:537-42.
- Afonso Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S, Vieira DN, and Corte-Real F (2010) Mitochondrial DNA sequence analysis of a native Bolivian population. *J Forensic Leg Med* 17:247-253.
- Albeck ME (2001) La puna argentina en los periodos medio y tardio. In Berberian E. E. and AE Nielsen (eds.): *Historia Argentina Prehispanica*. Cordoba: Editorial Brujas, pp. 347-388.
- Albeck ME, Conti VE, and Ruiz M (2003) Entre les oasis d'Atacama et la Puna de Jujuy : voies de communication et mobilité dans la longue durée. . *Histoire des Alpes*.:79-98.
- Albeck ME, and Ruiz M (2003) El tardio en la puna de Jujuy : poblados etnias y territorios. *Cuadernos FHyCS-UNJu* 20:199-221.
- Altuna ME, Modesti NM, and Demarchi DA (2006) Y-chromosomal evidence for a founder effect in Mbya-guarani Amerindians from northeast Argentina. *Hum Biol* 78:635-9.
- Alvarez-Iglesias V, Jaime JC, Carracedo A, and Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet* 1:44-55.
- Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SD, and

- Prado VF (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444-61.
- Amorim A, and Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int* 150:17-21.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, and Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-65.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, and Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Angiorama CI (2005a) Minereros quebradenos o altiplanicos? La circulacion de metales y minerales en el extremo noroccidental de Argentina (1280-1535 AD) *Intersecciones en Antropologia* 7:147-161.
- Angiorama CI (2005b) Nuevas evidencias de actividades metalurgicas pre-incaicas en la Quebrada de Humahuaca (Jujuy, Argentina). *Anales del Museo de America* 13:173-198.
- Aranda C, and Luna L (2005) Evidencias del impacto inkaico en el sitio Cortaderas Derecho (Sector Septentrional del Valle Calchaquí, Provincia de Salta). *Actas del XIII Congreso Nacional de Arqueología Argentina*, pp. 309-314.
- Avila F (2009) Interactuando desde el estilo. Variaciones en la circulacion espacial y temporal del estilo alfarero Yavi. *Esrudios Atacamenos : Arqueologia y Antropologia Surandinas* 37:29-50.

B.

- Baeza C, Guzman R, Tirado M, Lopez-Parra AM, Rodriguez T, Mesa MS, Fernandez E, and Arroyo-Pardo E (2007) Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Sci Int* 173:214-9.
- Baffi EI, Torres MF, and Cocilovo JA (1996) La poblacion antigua de las Pirguas (Salta Argentina), un enfoque integral. *Revista Argentina de Antropologia Biologica* 1:204-218.
- Bailliet G, Ramallo V, Muzzio M, Garcia A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, and Demarchi DA (2009) Brief communication:

- Restricted geographic distribution for Y-Q* paragroup in South America. *Am J Phys Anthropol* 140:578-82.
- Baldini L (2003) Proyecto arqueologia del valle Calchaqui central (Salta, Argentina). Sintesis y perspectivas. *ANALES Nueva Epoca*. 6:219-239.
- Baldini L, and Baffi EI (2007) Aportacion al estudio de practicas mortuorias durante el Periodo de Desarrollos Regionales. Entierros en vasijas utilitarias des sector central del valle Calchaqui (SAlta, Argentina). *Revista Espanola de Antropologia Americana*. 37:7-26.
- Bandelt HJ, Forster P, and Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, and Zhang YP (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 67:512-24.
- Barbieri C, Heggarty P, Castri L, Luiselli D, and Pettener D (2011) Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol* 23:89-99.
- Berardi G, Toscanini U, and Raimondi E (2003) STR data for PowerPlex 16 System from Buenos Aires population, Argentina. *Forensic Sci Int* 134:222-4.
- Bert F, Corella A, Gene M, Perez-Perez A, and Turbon D (2001) Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum Biol* 73:1-16.
- Bert F, Corella A, Gene M, Perez-Perez A, and Turbon D (2004) Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. *Ann Hum Biol* 31:9-28.
- Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martinez-Marignac VL, and Pena SD (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. *Am J Phys Anthropol* 102:79-89.
- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, and Lopez-Camelo JS (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* 63:1862-71.
- Bisso-Machado R, Jota MS, Ramallo V, Paixao-Cortes VR, Lacerda DR, Salzano FM, Bonatto SL, Santos FR, and Bortolini MC (2011) Distribution of Y-chromosome q

- lineages in native americans. *Am J Hum Biol* 23:563-6.
- Blanco-Verea A, Jaime JC, Brion M, and Carracedo A (2010) Y-chromosome lineages in native South American population. *Forensic Sci Int Genet* 4:187-93.
- Bobillo MC, Zimmermann B, Sala A, Huber G, Rock A, Bandelt HJ, Corach D, and Parson W (2009) Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med* 124:263-8.
- Bobillo MC, Zimmermann B, Sala A, Huber G, Rock A, Bandelt HJ, Corach D, and Parson W (2010) Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med* 124:263-8.
- Bolnick DA, Bolnick DI, and Smith DG (2006) Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol Biol Evol* 23:2161-74.
- Bonatto SL, and Salzano FM (1997a) Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* 61:1413-23.
- Bonatto SL, and Salzano FM (1997b) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci U S A* 94:1866-71.
- Bonilla C, Bertoni B, Gonzalez S, Cardoso H, Brum-Zorrilla N, and Sans M (2004) Substantial Native American female contribution to the population of Tacuarembó, Uruguay, reveals past episodes of sex-biased gene flow. *Am J Hum Biol* 16:289-97.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, and Ruiz-Linares A (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524-39.
- Bouakaze C, Keyser C, Crubezy E, Montagnon D, and Ludes B (2009) Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *Int J Legal Med* 123:315-25.
- Bravo L, and Llagostera A (1986) Solcor 3 : un aporte al conocimiento de la cultura San Pedro. Periodo 500 al 900 d.C. *Chungara* 16-17:323-332.
- Bray MS, Boerwinkle E, and Doris PA (2001) High-throughput multiplex SNP genotyping

- with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat* 17:296-304.
- Brenner CH (1997) Symbolic kinship program. *Genetics* 145:535-42.
- Brevnov MG, Pawar HS, Mundt J, Calandro LM, Furtado MR, and Shewale JG (2009) Developmental validation of the PrepFiler Forensic DNA Extraction Kit for extraction of genomic DNA from biological samples. *J Forensic Sci* 54:599-607.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev VB, Golovanova LV, de la Rasilla M, Fortea J, Rosas A, and Paabo S (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318-21.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, and Paabo S (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104:14616-21.
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, and Cooper A (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717-28.
- Bruzek J (2002) A method for visual determination of sex, using the human hip bone. *Am J Phys Anthropol* 117:157-68.
- Builes JJ, Bravo ML, Gomez C, Espinal C, Aguirre D, Gomez A, Rodriguez J, Castaneda P, Montoya A, Moreno M, Amorim A, and Gusmao L (2006) Y-chromosome STRs in an Antioquian (Colombia) population sample. *Forensic Sci Int* 164:79-86.

C.

- Cabana GS, Merriwether DA, Hunley K, and Demarchi DA (2006) Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol* 131:108-19.
- Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E, and Gilbert MT (2011) DNA in ancient bone - Where is it located and how should we extract it? *Ann Anat*.
- Cano RJ, Poinar HN, Pieniazek NJ, Acra A, and Poinar GO, Jr. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* 363:536-8.
- Cardoso S, Merino M, Gómez L, Alfonso-Sánchez MA, Dipierri JE, Alfaro E, Valverde L, Odriozola A, Bravi CM, Peña JA, and de Pancorbo MM (2009) Mitochondrial DNA

- control region in native population from the province of Jujuy (northwestern Argentina). *Forensic Science International: Genetics Supplement Series* 2:340-341.
- Cerezo M, Cerny V, Carracedo A, and Salas A (2009) Applications of MALDI-TOF MS to large-scale human mtDNA population-based studies. *Electrophoresis* 30:3665-73.
- Cooper A, and Poinar HN (2000) Ancient DNA: do it right or not at all. *Science* 289:1139.
- Corach D, Lao O, Bobillo C, van Der Gaag K, Zuniga S, Vermeulen M, van Duijn K, Goedbloed M, Vallone PM, Parson W, de Knijff P, and Kayser M (2010) Inferring continental ancestry of argentineans from Autosomal, Y-chromosomal and mitochondrial DNA. *Ann Hum Genet* 74:65-76.
- Corach D, Sala A, and Marino M (2006) Relevant genetic contribution of Amerindian to the extant population of Argentina. *International Congress Series ; Progress in Forensic Genetics* 11 1288:397-399.
- Corella A, Bert F, Perez-Perez A, Gene M, and Turbon D (2007) Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Mosen, Aymara and Quechua. *Ann Hum Biol* 34:34-55.

D.

- da Costa Francez PA, Rodrigues EM, de Velasco AM, and Dos Santos SE (2011) Insertion-deletion polymorphisms-utilization on forensic analysis. *Int J Legal Med*.
- Dalton R (2003) Archaeology: The coast road. *Nature* 422:10-2.
- de Azevedo S, Nocera A, Paschetta C, Castillo L, Gonzalez M, and Gonzalez-Jose R (2011) Evaluating microevolutionary models for the early settlement of the New World: The importance of recurrent gene flow with Asia. *Am J Phys Anthropol*.
- De Marrais E (2001) La arqueologia del norte del valle Calchaqui. In Berberian E. E. and AE Nielsen (eds.): *Historia Argentina Prehispanica*. Cordoba: Editorial Brujas, pp. 289–346.
- de Micou CP (2001a) Cesteria y cordeleria para los muertos. *Chungara (Arica)* 33:1-7.
- de Micou CP (2001b) Tecnologia cestera en la colleccion Doncellas (Jujuy, Argentina). *Museo de Arqueologia y Antropologia* 1:1-8.
- Demarchi DA, and Mitchell RJ (2004) Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers. *Hum Biol* 76:413-29.
- Demarchi DA, Panzetta-Dutari GM, Colantonio SE, and Marcellino AJ (2001a) Absence of

- the 9-bp deletion of mitochondrial DNA in pre-Hispanic inhabitants of Argentina. *Hum Biol* 73:575-82.
- Demarchi DA, Panzetta-Dutari GM, Motran CC, Lopez de Basualdo MA, and Marcellino AJ (2001b) Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. *Am J Phys Anthropol* 115:199-203.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanecek T, Villems R, and Zakharov I (2007) Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81:1025-41.
- Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, and Zakharov IA (2001) The presence of mitochondrial haplogroup x in Altaians from South Siberia. *Am J Hum Genet* 69:237-41.
- Dillehay TD (2003) Palaeoanthropology: tracking the first Americans. *Nature* 425:23-4.
- Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, and Pino-Navarro JD (2008) Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320:784-6.
- Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S, and Bianchi NO (1998) Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001-10.
- Dixon EJ (2001) Human colonization of the Americas : Timing, technology and process. *Quaternary Science Reviews* 21:1-7.
- Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, and Salzano FM (2004) Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479-88.
- Dornelles CL, Bonatto SL, De Freitas LB, and Salzano FM (2005) Is haplogroup X present in extant South American Indians? *Am J Phys Anthropol* 127:439-48.

E.

- Easton RD, Merriwether DA, Crews DE, and Ferrell RE (1996) mtDNA variation in the Yanomami: evidence for additional New World founding lineages. *Am J Hum Genet* 59:213-25.
- Egeland T, Mostad PF, Mevag B, and Stenersen M (2000) Beyond traditional paternity and identification cases. Selecting the most probable pedigree. *Forensic Sci Int* 110:47-59.

- Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, and Sanchez JJ (2006) Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. *PLoS One* 1:e81.
- Excoffier L, Laval G, and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.

F.

- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK, Santos SE, Petzl-Erler ML, and Bonatto SL (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82:583-92.
- Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S, and Herrmann B (2010a) Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. *Am J Phys Anthropol* 141:208-21.
- Fehren-Schmitz L, Warnberg O, Reindel M, Seidenberg V, Tomasto-Cagigao E, Isla-Cuadrado J, Hummel S, and Herrmann B (2010b) Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. *Ann Hum Genet* 75:266-83.
- Fix AG (2005) Rapid deployment of the five founding Amerind mtDNA haplogroups via coastal and riverine colonization. *Am J Phys Anthropol* 128:430-6.
- Fondevila M, Jaime JC, Salas A, Lareu MV, and Carracedo A (2003) Y-chromosome STR haplotypes in Cordoba (Argentina). *Forensic Sci Int* 137:217-20.
- Forster P, Harding R, Torroni A, and Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935-45.
- Forster P, Torroni A, Renfrew C, and Rohl A (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18:1864-81.
- Francalacci P, Bertranpetit J, Calafell F, and Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* 100:443-60.
- Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, and Pettener D (2003) Mitochondrial DNA diversity in South America and the genetic history of Andean

highlanders. *Mol Biol Evol* 20:1682-91.

G.

- Gabriel MN, Huffine EF, Ryan JH, Holland MM, and Parsons TJ (2001) Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J Forensic Sci* 46:247-53.
- Gabriel S, Ziaugra L, and Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet Chapter 2:Unit 2* 12.
- Garcia-Bour J, Perez-Perez A, Alvarez S, Fernandez E, Lopez-Parra AM, Arroyo-Pardo E, and Turbon D (2004) Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization. *Am J Phys Anthropol* 123:361-70.
- Garcia A, and Demarchi DA (2009) Incidence and distribution of Native American mtDNA haplogroups in central Argentina. *Hum Biol* 81:59-69.
- Garcia F, Moraga M, Vera S, Henriquez H, Llop E, Aspillaga E, and Rothhammer F (2006) mtDNA microevolution in Southern Chile's archipelagos. *Am J Phys Anthropol* 129:473-81.
- Gascue C, Mimbacas A, Sans M, Gallino JP, Bertoni B, Hidalgo P, and Cardoso H (2005) Frequencies of the four major Amerindian mtDNA haplogroups in the population of Montevideo, Uruguay. *Hum Biol* 77:873-8.
- Gaviria AA, Ibarra AA, Palacio OD, Posada YC, Triana O, Ochoa LM, Acosta MA, Brion M, Lareu MV, and Carracedo A (2005) Y-chromosome haplotype analysis in Antioquia (Colombia). *Forensic Sci Int* 151:85-91.
- Gaya-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, and Dugoujon JM (2011) mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *Am J Phys Anthropol* 145:215-30.
- Geppert M, Baeta M, Nunez C, Martinez-Jarreta B, Zweynert S, Cruz OW, Gonzalez-Andrade F, Gonzalez-Solorzano J, Nagy M, and Roewer L (2011) Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Sci Int Genet* 5:100-4.
- Gibbons A (1996) The peopling of the Americas. *Science* 274:31-3.
- Gilbert MT, Bandelt HJ, Hofreiter M, and Barnes I (2005) Assessing ancient DNA studies.

- Trends Ecol Evol 20:541-4.
- Gilbert MT, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, Carlson JE, Leebens-Mack JH, and Schuster SC (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35:1-10.
- Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, and Cooper A (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet* 72:48-61.
- Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Gotherstrom A, Campos PF, Rasmussen M, Metspalu M, Higham TF, Schwenninger JL, Nathan R, De Hoog CJ, Koch A, Moller LN, Andreassen C, Meldgaard M, Villems R, Bendixen C, and Willerslev E (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320:1787-9.
- Goebel T (2007) Anthropology. The missing years for modern humans. *Science* 315:194-6.
- Goebel T, Waters MR, and O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science* 319:1497-502.
- Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, and Zurawski G (1990) Chloroplast DNA sequence from a miocene Magnolia species. *Nature* 344:656-8.
- Gonzalez-Andrade F, Roewer L, Willuweit S, Sanchez D, and Martinez-Jarreta B (2009) Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Sci Int Genet* 3:e83-91.
- Gonzalez-Andrade F, Sanchez D, Martinez-Jarreta B, and Budowle B (2008) Y-chromosome STR haplotypes in three different population groups from Ecuador (South America). *J Forensic Sci* 53:512-4.
- Greenberg JH, Turner CGI, and Zegura SL (1986) the settlement of the Americas : a comparison of linguistic, dental, and genetic evidence. *Curr Anthropol* 27:477-497.
- Grignani P, Turchi C, Achilli A, Peloso G, Alu M, Ricci U, Robino C, Pelotti S, Carnevali E, Boschi I, Tagliabracci A, and Previdere C (2009) Multiplex mtDNA coding region SNP assays for molecular dissection of haplogroups U/K and J/T. *Forensic Sci Int Genet* 4:21-5.
- Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 17:475-92.

H.

- Hagelberg E (1997) Ancient and modern mitochondrial DNA sequences and the colonization of the Pacific. *Electrophoresis* 18:1529-33.
- Hagelberg E, and Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proc Biol Sci* 252:163-70.
- Hamilton MJ, and Buchanan B (2010) Archaeological support for the three-stage expansion of modern humans across northeastern Eurasia and into the Americas. *PLoS One* 5:e12472.
- Hartmann A, Thieme M, Nanduri LK, Stempfl T, Moehle C, Kivisild T, and Oefner PJ (2009) Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat* 30:115-22.
- Hasegawa M, Di Rienzo A, Kocher TD, and Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37:347-54.
- Haynes CV, Jr. (1964) Fluted Projectile Points: Their Age and Dispersion: Stratigraphically controlled radiocarbon dating provides new evidence on peopling of the New World. *Science* 145:1408-13.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, and Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-4.
- Ho SY, and Gilbert MT (2010) Ancient mitogenomics. *Mitochondrion* 10:1-11.
- Hofreiter M, Serre D, Poinar HN, Kuch M, and Paabo S (2001) Ancient DNA. *Nat Rev Genet* 2:353-9.
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, and Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10:23-47.
- Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, and Paabo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res* 24:1304-7.
- Hunley K, and Healy M (2011) The impact of founder effects, gene flow, and European admixture on native American genetic diversity. *Am J Phys Anthropol*.

I.

- Iannacone GC, Tito RY, Lopez PW, Medina ME, and Lizarraga B (2005) Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample. *J Forensic Sci* 50:239-42.

- Ingman M, Kaessmann H, Paabo S, and Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-13.
- Isbell WH (2008) Wari and Tiwanaku : International identities in the central Andean middle horizon. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 731-759.
- Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, and Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12:417-20.

J.

- Jobling MA, Hurles ME, and Tyler-Smith C (2004) *Human evolutionary genetics : origins, peoples and disease*. London/New York: Garland Science Publishing.
- Jota MS, Lacerda DR, Sandoval JR, Vieira PP, Santos-Lopes SS, Bisso-Machado R, Paixao-Cortes VR, Revollo S, Paz YMC, Fujita R, Salzano FM, Bonatto SL, Bortolini MC, and Santos FR (2011) A new subhaplogroup of native American Y-Chromosomes from the Andes. *Am J Phys Anthropol*.

K.

- Kalinowski ST, Wagner AP, and Taper ML (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* 6:576-579.
- Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, and Hammer MF (1997) Y chromosome markers and Trans-Bering Strait dispersals. *Am J Phys Anthropol* 102:301-14.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830-8.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, and Hammer MF (2002) High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74:761-89.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W,

- Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, and Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817-31.
- Kemp BM, Tung TA, and Summar ML (2009) Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. *Am J Phys Anthropol* 140:80-91.
- Keyeux G, Rodas C, Gelvez N, and Carter D (2002) Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Hum Biol* 74:211-33.
- Kitchen A, Miyamoto MM, and Mulligan CJ (2008) A three-stage colonization model for the peopling of the Americas. *PLoS One* 3:e1596.
- Knudson KJ, and Price TD (2007) Utility of multiple chemical techniques in archaeological residential mobility studies: case studies from Tiwanaku- and Chiribaya-affiliated sites in the Andes. *Am J Phys Anthropol* 132:25-39.
- Kolman CJ, Sambuughin N, and Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321-34.
- Korstanje MA, Cuenya P, and Williams VI (2010) Taming the control of chronology in ancient agricultural structures in the Calchaqui Valley, Argentina. Non-traditional data sets. *Journal of Archaeological Science* 37:343-349.

L.

- Lacan M, Keyser C, Ricaut FX, Brucato N, Duranthon F, Guilaine J, Crubezy E, and Ludes B (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 108:9788-91.
- Lalueza C, Perez-Perez A, Prats E, Cornudella L, and Turbon D (1997) Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. *Hum Mol Genet* 6:41-6.
- Lander N, Rojas MG, Chiurillo MA, and Ramirez JL (2008) Haplotype diversity in human mitochondrial DNA hypervariable regions I-III in the city of Caracas (Venezuela). *Forensic Sci Int Genet* 2:e61-4.
- Lavallée D (2006) Secuencias y consecuencias de algunos procesos de neolitización en los Andes Centrales. *Estudios Atacameños : Arqueología y Antropología Surandinas*

- 32:35-41.
- Lee J, Ewis AA, Hurles ME, Kashiwazaki H, Shinka T, and Nakahori Y (2007) Y chromosomal STRs haplotypes in two populations from Bolivia. *Leg Med (Tokyo)* 9:43-7.
- Leite FP, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CH, Raimann PE, Schwengber SP, Sortica VA, Tsuneto LT, Petzl-Erler ML, Salzano FM, and Hutz MH (2008) Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol* 20:359-63.
- Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, and Wallace DC (1997) Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. *Hum Genet* 100:536-43.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, and Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192-206.
- Leoni JB, and Acuto FA (2008) Social landscapes in Pre-Inca northwestern argentina. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 587-603.
- Lewis CM, Jr., Buikstra JE, and Stone AC (2007a) Ancient DNA and genetic continuity in the south central andes. *Latin American Antiquity* 18:145-160.
- Lewis CM, Jr., Lizarraga B, Tito RY, Lopez PW, Iannacone GC, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM, and Stone AC (2007b) Mitochondrial DNA and the peopling of South America. *Hum Biol* 79:159-78.
- Lewis CM, Jr., Tito RY, Lizarraga B, and Stone AC (2005) Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol* 127:351-60.
- Lewis Jr CM, Buikstra JE, and Stone AC (2007) Ancient DNA and genetic continuity in the south central andes. *Latin American Antiquity* 18:145-160.
- Li H, and Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493-6.
- Lindahl T (1993) Recovery of antediluvian DNA. *Nature* 365:700.
- Lindahl T (1997) Facts and artifacts of ancient DNA. *Cell* 90:1-3.
- Lorenz JG, and Smith DG (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. *Am J Phys Anthropol* 101:307-23.

Lorenz JG, and Smith DG (1997) Distribution of sequence variation in the mtDNA control region of Native North Americans. *Hum Biol* 69:749-76.

Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, and Pettener D (2000) Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* 113:5-17.

M.

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, and Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034-6.

Malhi RS, Cybulski JS, Tito RY, Johnson J, Harry H, and Dan C (2010) Brief communication: mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am J Phys Anthropol* 141:494-7.

Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gorodezky C, and Smith DG (2003) Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* 120:108-24.

Mannucci A, Sullivan KM, Ivanov PL, and Gill P (1994) Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. *Int J Legal Med* 106:190-3.

Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, and Bortolini MC (2007) Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am J Phys Anthropol* 132:301-10.

Martin PS (1973) The Discovery of America: The first Americans may have swept the Western Hemisphere and decimated its fauna within 1000 years. *Science* 179:969-74.

Mazieres S (2011) Towards a reconciling model about the initial peopling of America. *C R Biol* 334:497-504.

Mazieres S, Guitard E, Crubezy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, and Salzano FM (2008) Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations--implications for the region's colonization. *Ann Hum Genet* 72:145-56.

Mellars P (2006) Going east: new genetic and archaeological perspectives on the modern

- human colonization of Eurasia. *Science* 313:796-800.
- Melton PE, Briceno I, Gomez A, Devor EJ, Bernal JE, and Crawford MH (2007) Biological relationship between Central and South American Chibchan speaking populations: evidence from mtDNA. *Am J Phys Anthropol* 133:753-70.
- Mendisco F, Keyser C, Hollard C, Seldes V, Nielsen AE, Crubezy E, and Ludes B (2011) Application of the iPLEX Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: benefits for ancient population studies. *Electrophoresis* 32:386-93.
- Merriwether DA, Rothhammer F, and Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98:411-30.
- Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, and Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am J Hum Genet* 67:1277-86.
- Monsalve MV (1997) Mitochondrial DNA in ancient Amerindians. *Am J Phys Anthropol* 103:423-5.
- Monsalve MV, Cardenas F, Guhl F, Delaney AD, and Devine DV (1996) Phylogenetic analysis of mtDNA lineages in South American mummies. *Ann Hum Genet* 60:293-303.
- Monsalve MV, Groot de Restrepo H, Espinel A, Correal G, and Devine DV (1994) Evidence of mitochondrial DNA diversity in South American aboriginals. *Ann Hum Genet* 58:265-73.
- Moraga M, Santoro CM, Standen VG, Carvallo P, and Rothhammer F (2005) Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile. *Am J Phys Anthropol* 127:170-81.
- Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, and Carvallo P (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol* 113:19-29.
- Motti JMB, Rodenak B, Muzzio M, Ramallo V, Santos MR, Castro C, Alfaro EL, Dipierri JE, Scheible M, Saunier JL, Irwin JA, Coble MD, Bailliet G, and Bravi CM (2009) The genetic composition of Argentina prior to the massive immigration era: Insights from matrilineages of extant criollos in central-western Argentina. *Forensic Science*

International: Genetics Supplement Series 2:342-343.

Mulligan CJ, Kitchen A, and Miyamoto MM (2008) Updated three-stage model for the peopling of the Americas. *PLoS One* 3:e3199.

Mullis KB, and Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-50.

N.

Neel JV, Gershowitz H, Mohrenweiser HW, Amos B, Kostyu DD, Salzano FM, Mestriner MA, Lawrence D, Simoes AL, Smouse PE, Oliver WJ, Spielman RS, and Neel JV, Jr. (1980) Genetic studies on the Ticuna, an enigmatic tribe of Central Amazonas. *Ann Hum Genet* 44:37-54.

Nemecek S (2011) Les premiers américains: Dossier pour la science, pp. 8-13.

Neves W (2000) Luzia is not alone. *Science* 287:974-5.

Neves WA, Hubbe M, Okumura MM, Gonzalez-Jose R, Figuti L, Eggers S, and De Blasis PA (2005) A new early Holocene human skeleton from Brazil: implications for the settlement of the New World. *J Hum Evol* 48:403-14.

Nielsen AE (2001) Evolucion social en Quebrada de Humahuaca (AD 700-1536). In Berberian E. E. and AE Nielsen (eds.): *Historia Argentina Prehispanica*. Cordoba: Editorial Brujas, pp. 171-264.

Nielsen AE (2008) Una arqueologia de los "espacios vacios" de los Andes Circumpunenos. *Novedades de Antropologia* 59:3-6.

Nielsen AE, and Rivolta MC (1997) Asentamientos residenciales de ocupacion breve en la quebrada de Humahuaca (Jujuy, Argentina). *Chungara* 1:19-33.

Nunez Regueiro VA, and Tartusi MRA (2002) Aguada y el proceso de integracion regional. *Estudios Atacamenos*. 24:9-19.

Nunez Regueiro VA, and Tartusi MRA (2003) Los mecanismos de control y la organizacion del espacio durante los periodos formativos y de integracion regional. *Cuadernos FHyCS-UNJu* 20:37-50.

O.

O'Rourke DH, and Raff JA (2010) The human genetic history of the Americas: the final frontier. *Curr Biol* 20:R202-7.

Olivera DE (2001) Sociedades agropastoriles tempranas: el formativo inferior del noroeste

argentino. In EE Berberian and AE Nielsen (eds.): *Historia argentina indigena*. Cordoba: Editorial Brujas, pp. 83-125.

P.

- Paabo S (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314:644-5.
- Paabo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A* 86:1939-43.
- Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, and Hofreiter M (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-79.
- Pakendorf B, and Stoneking M (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165-83.
- Palma JR, do Rio SF, Runcio MA, and Capizzi L (2006) Museo arqueologico e historico de Huacalera (Quebrada de Huamahuaca, provincia de Jujuy) : un trabajo junto a la comunidad. *Intersecciones en Antropologia* 8:163-171.
- Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, and Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. *Nat Genet* 11:15-6.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, and Torroni A (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19:1-8.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballa A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward SR, Achilli A, and Torroni A (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20:1174-9.
- Pereira RW, Monteiro EH, Hirschfeld GC, Wang AY, and Grattapaglia D (2007) Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *Forensic Sci Int* 171:226-36.
- Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, and Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112:291-8.

- Phillips C, Rodriguez A, Mosquera-Miguel A, Fondevila M, Porras-Hurtado L, Rondon F, Salas A, Carracedo A, and Lareu MV (2008) D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. *Forensic Sci Int Genet* 3:7-13.
- Pitulko VV, Nikolsky PA, Girya EY, Basilyan AE, Tumskey VE, Koulakov SA, Astakhov SN, Pavlova EY, and Anisimov MA (2004) The Yana RHS site: humans in the Arctic before the last glacial maximum. *Science* 303:52-6.
- Polzin T, and Daneschmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* 31.
- Proulx DA (2008) Paracas and Nasca: regional cultures on the south coast of Peru. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 563-585.

R.

- Raff JA, Bolnick DA, Tackney J, and O'Rourke DH (2011) Ancient DNA perspectives on American colonization and population history. *Am J Phys Anthropol*.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Gronnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Ponten T, Villems R, Nielsen R, Krogh A, Wang J, and Willerslev E (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757-62.
- Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, and Excoffier L (2010) A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol* 27:337-45.
- Ribeiro-dos-Santos AK, Santos SE, Machado AL, Guapindaia V, and Zago MA (1996) Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian Natives of the Amazon region. *Am J Phys Anthropol* 101:29-37.
- Ricaut FX, Keyser-Tracqui C, Bourgeois J, Crubezy E, and Ludes B (2004) Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian

- migrations. *Hum Biol* 76:109-25.
- Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, and Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* 65:519-30.
- Rivera MA (2008) The Archaeology of Northern Chile. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 963-977.
- Rivolta MC (2007) Abandono y reutilización de sitios. La problemática de los contextos habitacionales en quebrada de Humahuaca. *Estudios Atacamenos : Arqueología y Antropología Surandinas* 34:31-49.
- Rocco P, Morales C, Moraga M, Miquel JF, Nervi F, Llop E, Carvallo P, and Rothhammer F (2002) [Genetic composition of the Chilean population. Analysis of mitochondrial DNA polymorphism]. *Rev Med Chil* 130:125-31.
- Rodriguez-Delfin LA, Rubin-de-Celis VE, and Zago MA (2001) Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. *Hum Hered* 51:97-106.
- Rohland N, Siedel H, and Hofreiter M (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour* 10:677-83.
- Romero RE, Briceno I, Lizarazo Rdel P, Willuweit S, Roewer L, and Gomez A (2008) A Colombian Caribbean population study of 16 Y-chromosome STR loci. *Forensic Sci Int Genet* 2:e5-8.
- Rothhammer F, and Dillehay TD (2009) The late Pleistocene colonization of South America: an interdisciplinary perspective. *Ann Hum Genet* 73:540-9.
- Rothhammer F, Moraga M, Rivera M, Santoro CM, Standen VG, Garcia F, and Carvallo P (2003) Analisis de ADNmt de restos esquelerales del sitio arqueologico de Tiwanaku y su relacion con el origen de sus constructores. *Chungara* 35:269-274.
- Rothhammer F, and Silva C (1989) Peopling of Andean South America. *Am J Phys Anthropol* 78:403-10.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-7.
- Ruiz-Linares A, Ortiz-Barrientos D, Figuerola M, Mesa N, Munera JG, Bedoya G, Velez ID, Garcia LF, Perez-Lezaun A, Bertranpetit J, Feldman MW, and Goldstein DB (1999) Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc Natl Acad Sci U S A* 96:6312-7.

S.

- Sala A, Penacino G, Carnese R, and Corach D (1999) Reference database of hypervariable genetic markers of Argentina: application for molecular anthropology and forensic casework. *Electrophoresis* 20:1733-9.
- Salaberry MT, Dejean CB, Goicoechea AS, Baffi EI, and Carnese FR (1999) Grupos sanguíneos en restos momificados de una población prehispánica : "las Pirguas" (Salta, Argentina). *Revista Argentina de Antropología Biológica* 2:267-280.
- Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, and Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713-24.
- Santos FR, Rodriguez-Delfin L, Pena SD, Moore J, and Weiss KM (1996) North and South Amerindians may have the same major founder Y chromosome haplotype. *Am J Hum Genet* 58:1369-70.
- Schanfield MS, Crawford MH, Dossetor JB, and Gershowitz H (1990) Immunoglobulin allotypes in several North American Eskimo populations. *Hum Biol* 62:773-89.
- Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, and Salzano FM (2004) Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. *Ann Hum Biol* 31:87-94.
- Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, and Smith DG (2007) A private allele ubiquitous in the Americas. *Biol Lett* 3:218-23.
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613-23.
- Schurr TG, and Sherry ST (2004) Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* 16:420-39.
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, and Wells RS (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* 73:700-5.
- Seldes V (2006) Bioarqueología de poblaciones prehistóricas de Humahuaca (Jujuy,

- Argentina). *Estudios Atacamenos* 31:47-61.
- Seldes V (2007) Aportes de la bioarqueología al estudio de la complejidad y la desigualdad social en la Quebrada de Humahuaca (Jujuy, Argentina), Facultad de Filosofía y Letras., Buenos Aires.
- Shapiro B, and Hofreiter M (2010) Analysis of ancient human genomes: using next generation sequencing, 20-fold coverage of the genome of a 4,000-year-old human from Greenland has been obtained. *Bioessays* 32:388-91.
- Shimada I, Shinoda K, Farnum J, Corruccini R, and Watanabe H (2004) An integrated analysis of pre-hispanic mortuary practices. *Current Anthropology* 45:369-392.
- Shinoda K, Adachi N, Guillen S, and Shimada I (2006) Mitochondrial DNA analysis of ancient Peruvian highlanders. *Am J Phys Anthropol* 131:98-107.
- Silva WA, Jr., Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixao BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paco-Larson ML, Petzl-Erler ML, Valente V, Santos SE, and Zago MA (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am J Hum Genet* 71:187-92.
- Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, and Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674-9.
- Soltis PS, Soltis DE, and Smiley CJ (1992) An *rbcL* sequence from a Miocene *Taxodium* (bald cypress). *Proc Natl Acad Sci U S A* 89:449-51.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, and Wallace DC (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69:67-89.
- Stone AC, and Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet* 62:1153-70.
- Stovel E (2008) Interaction and Social Fields in San Pedro de Atacama, South Central Andes. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 979-1002.
- Sullivan KM, Mannucci A, Kimpton CP, and Gill P (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* 15:636-8, 640-1.
- Surovell TA (2000) Early Paleoindian women, children, mobility, and fertility. *Am Antiq*

65:493-508.

Szathmary EJ (1984) Peopling of northern North America: clues from genetic studies. *Acta Anthropogenet* 8:79-109.

Szibor R, and Michael M (1999) Correct mitochondrial L-strand sequencing after C-stretches.

W. Parson et al. *Int J Legal Med* (1998) 111: 124-132. *Int J Legal Med* 112:348-9.

T.

Taboada C, and Angiorama CI (2003) Buscando los indicadores arqueologicos de la unidad domestica. *Cuadernos FHyCS-UNJu* 20:393-407.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, and Malhi RS (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2:e829.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28:2731-9.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, and Santos FR (2001) Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68:1485-96.

Tarrago MN (1999) Las sociedades del Sudeste andino. In TR Rabiela and JV Murra (eds.): *Las sociedades originarias (Historia general de america latina)*. Madrid: Trotta.

Tartusi MRA, and Nunez Regueiro VA (2001) Fenomenos culticos tempranos en la subregion valliserrana. In EE Berberian and AE Nielsen (eds.): *Historia argentina indigena*. Cordoba: Editorial Brujas, pp. 83-125.

Tartusi MRA, and Nunez Regueiro VA (2003) Procesos de interaccion entre poblaciones de los valles intermontanos del noroeste argentino y las del piedemonte. *ANALES Nueva Epoca*. 6:43-62.

Tirado M, Lopez-Parra AM, Baeza C, Bert F, Corella A, Perez-Perez A, Turbon D, and Arroyo-Pardo E (2009) Y-chromosome haplotypes defined by 17 STRs included in AmpFlSTR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia). *Leg Med (Tokyo)* 11:101-3.

- Torres-Rouff C (2011) Hiding inequality beneath prosperity: Patterns of cranial injury in middle period San Pedro de Atacama, Northern Chile. *Am J Phys Anthropol*.
- Torres MM, Bravi CM, Bortolini MC, Duque C, Callegari-Jacques S, Ortiz D, Bedoya G, Groot de Restrepo H, and Ruiz-Linares A (2006) A revertant of the major founder Native American haplogroup C common in populations from northern South America. *Am J Hum Biol* 18:59-65.
- Torrioni A, Achilli A, Macaulay V, Richards M, and Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339-45.
- Torrioni A, Chen YS, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, and Wallace DC (1994) mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *Am J Hum Genet* 54:303-18.
- Torrioni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC (1993a) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563-90.
- Torrioni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, and et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153-62.
- Torrioni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, and Wallace DC (1993b) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* 53:591-608.
- Toscanini U, Gusmao L, Berardi G, Amorim A, Carracedo A, Salas A, and Raimondi E (2008) Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data. *Forensic Sci Int Genet* 2:274-80.
- Toscanini U, Gusmao L, Berardi G, Gomes V, Amorim A, Salas A, and Raimondi E (2011) Male lineages in South American native groups: Evidence of M19 traveling south. *Am J Phys Anthropol*.
- Tovar F, Chiurilloa MA, Borjas L, Lander N, and Ramírez JL (2006) Chromosome Y haplotypes database in a Venezuelan population. *International Congress Series* 1288:246-248.
- Tung TA (2008) Life on the move : bioarchaeological contributions to the study of migration and diaspora communities in the Andes. In H Silverman and W Isbell (eds.): *Handbook of South American Archaeology*. New York: Springer pp. 671-680.

U.

Underhill PA, Jin L, Zemans R, Oefner PJ, and Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci U S A* 93:196-200.

V.

van Oven M, and Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386-94.

van Oven M, Ralf A, and Kayser M (2011) An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int J Legal Med.*

Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, and Sukernik RI (2008) Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am J Hum Genet* 82:1084-100.

Vona G, Falchi A, Moral P, Calo CM, and Varesi L (2005) Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. *Am J Phys Anthropol* 127:361-9.

W.

Wallace DC (1995) Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet* 57:201-23.

Wallace DC, Garrison K, and Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. *Am J Phys Anthropol* 68:149-55.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, and Ruiz-Linares A (2007) Genetic variation and population structure in native Americans. *PLoS Genet* 3:e185.

Ward RH, Frazier BL, Dew-Jager K, and Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci U S A* 88:8720-4.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra Jr CEA, and Santos RV (1996)

- Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am J Hum Biol* 8:317-372.
- Waters MR, and Stafford TW, Jr. (2007) Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 315:1122-6.
- Watson CL, and Lockwood DN (2009) Single nucleotide polymorphism analysis of European archaeological *M. leprae* DNA. *PLoS One* 4:e7547.
- Weber JL, David D, Heil J, Fan Y, Zhao C, and Marth G (2002) Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet* 71:854-62.
- Willerslev E, and Cooper A (2005) Ancient DNA. *Proc Biol Sci* 272:3-16.
- Willerslev E, Gilbert MT, Binladen J, Ho SY, Campos PF, Ratan A, Tomsho LP, da Fonseca RR, Sher A, Kuznetsova TV, Nowak-Kemp M, Roth TL, Miller W, and Schuster SC (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evol Biol* 9:95.
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, and Cooper A (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300:791-5.
- Williams SR, Chagnon NA, and Spielman RS (2002) Nuclear and mitochondrial genetic variation in the Yanomamo: a test case for ancient DNA studies of prehistoric populations. *Am J Phys Anthropol* 117:246-59.
- Woodward SR, Weyand NJ, and Bunnell M (1994) DNA sequence from Cretaceous period bone fragments. *Science* 266:1229-32.

Y.

- YCC (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339-48.

Z.

- Zaburlin MA (2008) Historia de ocupacion del Pucara de Tilcara (Jujuy, Argentina). *Intersecciones en Antropologia* 10:89-103.
- Zazula GD, Froese DG, Schweger CE, Mathewes RW, Beaudoin AB, Telka AM, Harington CR, and Westgate JA (2003) Palaeobotany: Ice-age steppe vegetation in east Beringia. *Nature* 423:603.
- Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF (2004) High-resolution SNPs and

microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21:164-75.

Zhang W, Reading C, and Deisseroth AB (1992) Improved PCR sequencing with formamide. *Trends Genet* 8:332.

Zuckerkandl E (1987) On the molecular evolutionary clock. *J Mol Evol* 26:34-46.

Zuckerkandl E, and Pauling L (1965) Molecules as documents of evolutionary history. *J Theor Biol* 8:357-66.

Annexes

Annexe 1 : Tableau récapitulatif présentant la séquence des amorces PCR et SBE utilisées pour le génotypage par spectrométrie de masse MALDI-TOF, et la taille des fragments amplifiés.

Marqueur	Amorce PCR-2	Amorce PCR-1	Longueur de l'amplicon (pb)	Amorce SBE	Orientation de l'amorce SBE*
A2_12007	ACGTTGGATGAGCCCTATACTCCCTCTACA	ACGTTGGATGTGTTGTTAATGTGGTGGGTG	88	ACATATTTACCACAACACAATG	F
A2_64	ACGTTGGATGACCCTATTAACCACTCACGG	ACGTTGGATGTCTCGCAATGCTATCGCGTG	100	AGCTCTCCATGCATTTGGTATTTTCGT	F
B2_3547	ACGTTGGATGGGGTTGGGTATGGGGAGGG	ACGTTGGATGATCACCTCTACATCACCGC	100	GTTCATAGTAGAAGAGCGA	R
C1b_493	ACGTTGGATGACATTATTTCCCCTCCCAC	ACGTTGGATGTGTGTGTGCTGGGTAGGATG	96	ACTACTAATCTCATCAACACA	F
C1c_15930	ACGTTGGATGAAATGGGCCTGTCCTTGTAG	ACGTTGGATGCTGATTTGTCCTTGAAAAAG	102	TAATACACCAGTCTTGTAACCGGA	F
C1d_7697	ACGTTGGATGGTGAGTGTTAGGAAAAGGGC	ACGTTGGATGTCACCTTTCATGATCACGCC	99	TTAGGAAAAGGGCATAACAGGA	R
D1_2092	ACGTTGGATGTTGCCACAGAACCCTCTAA	ACGTTGGATGTCCTAGTGTCCAAAGAGCTG	97	CTCTAAATCCCCTTGTAATTTAA	F
Q_M242	ACGTTGGATGCAAAAAACACGTTAAGACC	ACGTTGGATGGTAAGCTTTCTACGGCATAG	99	AAACACGTTAAGACCAATGCCAA	R
Q1a3_M346	ACGTTGGATGAGATGGGAAAGGCAGCCAAG	ACGTTGGATGTCACTCTGCCTACCTGATAG	98	GCCAAGAGGACAGTAAGA	F
Q1a3a_M3	ACGTTGGATGTAGGTTGACCTGACAATGGG	ACGTTGGATGCTTTCATTTTAGGTACCAGC	88	GTCACCTCTGGGACTGA	F
Q1a3a1_M19	ACGTTGGATGACTGATGTAGAGACATCTG	ACGTTGGATGTTTGGGGTTTGTAAAGTGGC	95	AGAGACATCTGAAACCCAC	R
Q1a3a2_M194	ACGTTGGATGCCAAAAGGATTTTGAATAC	ACGTTGGATGATACAGTCGTTGCCTTCTCG	92	AACATACAGGGAGTGTTTTT	F
Q1a3a3_M199	ACGTTGGATGGCTCTGTAAAGCTGCATTG	ACGTTGGATGTGTTAGTCTTAAGATTGTC	92	GAAATGTTAAAATGGCTTACACTTG	F
Q1a3a4_SA01	ACGTTGGATGAAATATGTATTTTGGCATGT	ACGTTGGATGTGCAGCTAAATTAGGCAACC	97	TGTATTTTGGCATGTAAATGACC	R
C_M216	ACGTTGGATGCTGCTAGTTATGTATACCTG	ACGTTGGATGCTCAACCAGTTTTTATGAAGC	100	ATGTATACCTGTTGAAT	R

* F = forward, R = reverse.

Annexe 2 : Résultats de la quantification obtenue pour au moins un extrait de chaque échantillon ancien (la valeur la plus haute est présentée).

Echantillon	Quantité (pg/ μ L)	Inhibiteurs	Echantillon	Quantité (pg/ μ L)	Inhibiteurs
Quebrada de Humahuaca			Quebrada de Humahuaca (suite)		
BP-1	79,3	non	Ju-1	10	non
BP-2	n.d.	non	Ju-2	n.d.	non
Hu	30	non	Ju-3	n.d.	non
LA1	4,48	non	Ju-4	11,6	non
LA-2	n.d.	non	Ju-5	13,1	non
LA-3	151	non	Vallée Calchaqui		
LA-4	56,3	non	FA-1	89	non
LA-5	n.d.	non	FA-2	n.d.	non
LA-6	3,4	non	Sc-1	n.d.	non
LA-7	17	non	Sc-2	10,2	non
LA-8	55,7	non	Sc-3	1,87	non
LA-9	80,9	non	Sc-4	n.d.	non
LA-10	13,3	non	Sc-5	1,31	non
LA-11	n.d.	non	Sc-6	n.d.	non
LA-12	31,7	non	Sc-7	n.d.	non
LA-13	1430	non	Sc-8	42,6	non
LA-14	43,1	non	Cor-1	48,6	non
LA-15	n.d.	non	Cor-2	7,2	non
LA-17	24,4	non	Cor-3	189	non
LA-18	57,3	non	Doncellas		
LA-19	n.d.	non	Don-1	10	non
LA-20	69,8	non	Don-2	43,8	non
LA-21	28,5	non	Don-3	n.d.	non
LA-22	n.d.	non	Pampa Grande		
LA-23	140	non	Li-1	39,8	non
LA-24	10	non	Li-2	4,95	non
Mu-1	n.d.	non	Li-3	n.d.	non
Mu-2	n.d.	non	Li-4	138	non
Mu-3	n.d.	non	C.II-1	301	non
Mu-4	8	non	C.II-2	n.d.	non
Mu-5	370	non	C.II-3	1	non
Mu-6	n.d.	non	C.II-4	3,66	non
Mu-8	11,2	non	C.V-1	20,8	non
PB	n.d.	non	C.V-2	1680	non
Sa-1	38,7	non	C.V-3	42,5	non
Sa-2	n.d.	non	C.V-4	n.d.	non
SJ	10	non	Lam-1	n.d.	non
Til-1	20,8	non	Lam-2	7,56	non
Til-2	n.d.	non	Lam-3	366	non
Til-3	26,4	non	Lam-4	165	non
Til-4	38,9	non	Lam-5	718	non
Til-5	104	non	Ap-1	123	non
Til-6	n.d.	non	Ap-2	105	non
F-1	n.d.	non	Ap-3	58	non
F-2	n.d.	non	Ap-4	261	non
Hui	277	non			

* Les résultats de l'extrait présentant la quantité la plus élevées sont présentés.

Annexe 3 : (A.) Profils génétiques des fouilleurs (F) et personnels ayant eu accès au laboratoire d'ADNa (M), obtenus avec le kit Identifier + ;
 (B.) Séquences HVR-1 obtenues pour le personnel ayant eu accès au laboratoire d'ADNa (M) et les fouilleurs ayant eu accès au matériel (F).

A.	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA	AMEL
M1	14/16	30/32,2	10/11	10/12	16/17	9,3/9,3	11/13	9/12	16/25	13,2/15	18/19	8/11	12/15	11/12	21/21	XX
M2	12/14	29/29	9/11	12/12	14/18	9/9,3	12/12	11/11	17/19	13/13	15/15	11/11	14/17	11/11	21/23	XX
M3	10/11	30/32,2	10/10	11/12	16/17	9/9,3	8/11	11/12	20/23	14/14	16/18	11/12	12/13	11/13	19/25	XX
M4	8/13	30/31,2	10/11	12/12	16/17	6/8	12/12	9/12	25/25	13/14	17/17	11/11	11/18	11/13	22,2/23	XX
F1	9/11	28/31	10/11	10/11	15/15	6/9,3	12/12	12/12	18/20	14/16,2	13/16	8/10	14/21	12/13	19/21	XY
F2	10/14	32,2/33,2	10/11	11/12	15/15	6/9	8/9	13/14	16/18	14/15,2	16/19	8/10	15/16	11/12	20/22	XX
F3	11/14	29/30,2	10/10	10/11	15/16	6/9,3	9/11	12/12	22/23	13,2/14	15/16	8/8	13/19	12/12	22/23	XX
F4	13/13	28/31,2	8/10	10/13	15/17	7/9,3	12/13	11/11	19/20	13/14	17/18	11/12	12/14	11/13	21/25	XY
F5	13/14	28/30	10/11	10/12	14/15	9/13	11/11	13/13	17/17	14/15	18/19	8/11	13/16	10/11	22,2/28	XY
F6	10/13	31,2/33,2	8/10	11/11	15/16	9/13	8/12	12/12	17/20	14/16	17/18	8/11	14/20	9/13	22/25	XX
F7	13/14	27/32,2	10/11	10/13	15/16	6/9	8/9	12/13	17/25	14/14	17/19	11/11	12/15	7/12	19/24	XX
F8	10/14	32,2/33,2	10/11	11/12	15/15	6/9	8/9	13/14	16/18	14/15,2	16/19	8/10	15/16	11/12	20/22	XX
F9	12/13	29/30	11/12	10/11	15/17	6/9	11/13	10/11	16/24	13/15	16/18	8/8	13/14	11/13	19/20	XY

B. Polymorphismes de la région HV-1 (16024-16383 pb)

M1	16192T - 16223T - 16325C
M2	16270T
M3	16093C - 16189C - 16270T - 16274G
M4	CRS
F1	16154C - 16193T - 16223T - 16319A - 16325C - 13362C
F2	16069T - 16126T
F3	16224C - 16311C
F4	16183C - 16189C - 16217C

Annexe 4 : Ensemble des génotypes consensus obtenus par spectrométrie de masse MALDI-TOF pour les 70 échantillons anciens amérindiens analysés. Seuls les génotypes des 15 SNP conservés dans la dernière réaction sont présentés.

Echantillon	Sexe	SNP mt							HG_Mt	SNP Y							HG_Y
		A2_12007	A2_64	B2_3547	C1b_493	C1c_15930	C1d_7697	D1_2092		Q_M242	Q1a3_M346	Q1a3a_M3	Q1a3a1_M19	Q1a3a2_M194	Q1a3a3_M199	Q1a3a4_SA01	
Vallée Calchaqui																	
FA-1	XY	A	T	A	A	G	G	C	A2 c-r	T	G	T	T	T	DEL	C	Q1a3a*
FA-2	XX	A	T	A	A	G	G	C	A2 c-r								
Sc-1	XX	G	C	A	G	G	G	C	C1b								
Sc-2	XY	G	C	G	A	G	G	C	B2	T	G	T	T	T	DEL	C	Q1a3a*
Sc-3	XX	A	T	A	A	G	G	C	A2 c-r								
Sc-4	Ind.		C		A		G										
Sc-5	XY	G	C	G	A	G	G	C	B2	T	G	C	T	T	DEL	C	Q1a3*
Sc-6	Ind.		T		A		G										
Sc-7	Ind.		T	A		G	G										
Cor-1	XY	G	C	A	G	G	G	C	C1b	T	G	C	T	T	DEL		Q1a3*
Cor-2	XY									T							
Cor-3	XX	G	C	A	G	G	G	C	C1b								
Doncellas																	
Don-1	XY	G	C	G	A	G	G	C	B2	T	G	T	T	T	DEL	C	Q1a3a*
Don-2	XX	G	C	A	G	G	G	C	C1b								
Don-3	XX	G	C	G	A	G	G	C	B2								
Pampa Grande																	
Li-1	XY	G	C	G	A	G	G	C	B2	T	G	C	T	T	DEL	C	Q1a3*
Li-2	XX	G	C	A	A	G	G	T	D1								
Li-3	XY	G	C	A	A	G	G	T	D1	T	G	T	T	T	DEL	C	Q1a3a*
Li-4	XY	G	C	A	A	G	G	T	D1	T	G	T	T	T	DEL	C	Q1a3a*
C.II-1	XY	G	C	G	A	G	G	C	B2	T	G	T	T	T	DEL	C	Q1a3a*
C.II-2	XX	G	C	G	A	G	G	C	B2								
C.II-3	XX	G	C	G	A	G	G	C	B2								
C.II-4	XX	G	C	G	A	G	G	C	B2								
C.V-1	XY	G	C	A	A	G	G	T	D1	T	G	C	T	T			Q1a3
C.V-2	XY	G	C	G	A	G	G	C	B2	T	G	T	T	T	DEL	C	Q1a3a*
C.V-3	XY	G	C	G	A	G	G	C	B2	T	G	C	T	T	DEL	C	Q1a3*
C.V-4	XX	A	C	A	A	G	G	C	A2								
Lam-1	XX	A	T	A	A	G	G	C	A2 c-r								
Lam-2	XY	G	C	A	A	G	G	T	D1	T	G	C	T	T	DEL	C	Q1a3*
Lam-3	XY	A	T	A	A	G	G	C	A2 c-r	T	G	C	T	T	DEL	C	Q1a3*
Lam-4	XY	G	C	A	A	G	G	T	D1	T		C	T	T	DEL		Q1a3*
Lam-5	XX	G	C	G	A	G	G	C	B2								
Ap-1	XX	G	C	A	A	G	G	T	D1								
Ap-2	XY	G	C	G	A	G	G	C	B2	T	G	C	T	T	DEL	C	Q1a3*
Ap-3	XX	G	C	G	A	G	G	C	B2								
Ap-4	XY	G	C	A	A	G	G	T	D1	T	G	C	T	T	DEL	C	Q1a3*

Annexe 5 : Présentation des profils génétiques consensus obtenus à partir de l'utilisation du kit Mentype DIPplex (Qiagen), pour 14 échantillons anciens pour lesquels des liens de parenté étaient supposés.

	AM	HLD77	HLD45	HLD131	HLD70	HLD6	HLD111	HLD58	HLD56	HLD118	HLD92	HLD93	HLD99	HLD88	HLD101	HLD67	HLD83	HLD114	HLD48	HLD124	HLD122	HLD125	HLD64	HLD81	HLD136	HLD133	HLD97	HLD40	HLD128	HLD39	HLD84		
C+	XY	+/+	-/-	+/-	-/-	+/+	-/-	+/-	-/-	+/-	+/+	+/-	+/-	+/-	+/-	+/+	+/-	-/-	+/-	+/-	-/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	-/-		
LA-9	XY	+/-	+/+	+/-	+/-	+/+	+/-	-/-	-/-	+/-	+/-	+/-	-/-	-/-	+/-	+/+	+/-	+/-	+/+	+/-	-/-	+/-	+/+	+/+	+/-	-/-	+/-	+/-	+/-	-/-	+/-		
LA-20	XY	+/+	+/+	+/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	-/-	+/+	-/-	-/-	+/+	+/+	+/-	-/-	+/+	+/+	+/+	-/-	+/-	+/-	+/-	+/-	+/-		
LA-8	XX	+/-	+/+	-/-	+/-	+/+	+/-	+/-	+/-	-/-	-/-	+/-	-/-	-/-	-/-	+/+	-/-	+/+	+/-	-/-	-/-	+/-	+/+	+/-	-/-	+/+	+/+	+/+	+/-	+/+	+/+		
LA-4	XY					+/+	+/-	+/-	-/-		-/-				+/-	+/+						-/-	+/+	+/-		+/-	+/-	+/+	+/+	+/-	+/-		
LA-23	XX	+/-	+/-	-/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/+	-/-	+/+	+/-	-/-	-/-	-/-	+/+	+/-	-/-	-/-	+/+	+/-	+/+	-/-	+/-	+/-	+/-	+/-	+/-	+/+	
LA-14	XX	+/+	+/-	-/-	+/+	+/+	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/+	+/+	-/-	-/-	-/-	+/-	-/-	-/-	-/-	+/+	+/-	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/+	
LA-24	XY	+/-	+/+	-/-	+/+	+/+	+/-		-/-	-/-	+/-	+/-	+/+	+/+	+/-	+/+	-/-	-/-	+/+	+/+	+/+	-/-	+/+	+/+	+/-	-/-	+/-	+/-	+/+	+/+	-/-		
LA-3	XY	+/-	+/+	+/-	+/-	+/-	-/-	-/-	+/-	-/-	-/-	+/+	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-	-/-	-/-	+/+	+/+	+/-	+/+	+/+	+/-	+/-	+/-	+/-	+/-	
LA-10	XY	+/-	+/+	-/-	+/+	+/+	+/-	+/+	+/+	-/-	+/+	+/-	+/-	+/+	+/+	+/-	-/-	-/-	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	-/-	
LA-12	XY	+/-	+/+	+/-	+/+	+/-	+/-	-/-	+/+	+/-	+/-	+/-	+/+	+/-	+/-	-/-	+/-	-/-	+/+	+/-	-/-	-/-	+/-	-/-	+/-	+/-	-/-	-/-	+/+	+/-	+/-	-/-	
LA-13	XY	-/-	+/-	-/-	+/-	+/-	+/-	-/-	+/+	-/-	+/-	+/+	+/+	+/-	+/-	+/-	+/-	-/-	+/+	+/-	-/-	-/-	+/+	-/-	+/-	+/-	+/+	+/+	+/+	+/-	-/-	-/-	
Cor-1	XY	+/-	+/+	-/-	+/+	+/-	+/-	-/-	+/+	+/-	-/-	+/+	+/+	-/-	-/-	-/-	-/-	+/-	+/+	-/-	-/-	+/-	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	-/-	-/-	+/+
Cor-2	XY	-/-	+/+	+/-	+/+	+/-	-/-	-/-	+/+	+/-	-/-	+/+	+/+	-/-	-/-	-/-	-/-	-/-	+/+	-/-	-/-	+/-	+/+	+/-	+/+	+/+	+/+	+/+	+/+	+/-	-/-	-/-	+/+
Cor-3	XX	-/-	+/+	+/-	+/+	+/+	+/-	-/-	+/+	+/-	-/-	+/-	+/+	-/-	-/-	+/-	-/-	-/-	+/-	+/-	-/-	+/-	+/+	+/-	+/-	-/-	+/-	-/-	-/-	-/-	+/-	-/-	-/-

Annexe 6 : Fréquences des haplogroupes mitochondriaux de 69 populations d'Amérique du Sud compilées à partir des données disponibles dans la littérature.

Populations	N	A	B	C	D	Autre	Populations	N	A	B	C	D	Autre
Aire caribéenne							Andes septentrionales						
Arawak	50	37	26	35	0	2	Zenu - Embera	59	39	33,9	18,7	3,4	5
Chibchan	138	73,2	0,7	26,1	0	0	Pool Andes Colombie	241	24,9	28,3	40,2	3,3	3,3
Pool Nord Colombie	379	31,4	38,5	24	5,3	0,8	Ingano	27	15	44	37	0	4
Caracas	100	34	5	20	8	33	Cayapa	62	29	22,6	21	27,4	0
Guahibo	59	47,5	3,39	49,2	0	0	Côte péruvienne						
Amazonie							<i>Palpa</i>	130	2,9	11,5	26,9	58,7	0
Pool Est Colombie	433	28,4	18	31,9	12,5	9,2	<i>Sican</i>	28	21,4	25	3,6	25	25
Ticuna	82	10	10	36	39	5	Andes centrales						
Yanomami.1	238	0	38,2	46,2	11,4	4,2	Yungay	36	3	47	36	14	0
Palikur	48	0	56,2	0	43,8	0	Ancash	33	9	52	18	21	0
Emerillon	30	30	70	0	0	0	Lima / Pasco	52	3,9	53,8	17,3	19,2	5,8
Kali'na	29	6,9	41,4	37,9	6,9	6,9	San Martin	22	9	55	5	27	4
Wayampi	53	47,2	20,7	0	32,1	0	Tupe	16	0	69	31	0	0
Apalai	102	39,2	1	29,4	30,4	0	Tayacaja	61	21	33	30	13	3
Gaviao	27	14,8	14,8	0	70,4	0	QuechuaP	19	26,3	36,8	5,3	31,6	0
Zoro	30	20	3,3	13,4	60	3,3	Puno	44	4,5	63,6	20,5	11,4	0
Xavante	28	16	84	0	0	0	Arequipa	22	9	68	14	9	0
Matsiguenga	38	5,3	92,1	0	2,6	0	Titicaca	64	11	69	9	11	0
Plaines Moxos	54	18,5	24,1	50	5,6	1,8	Aymara	270	6	74,7	9,3	10	0
Llanos pool	168	27,3	39,3	26,2	3,6	3,6	La Paz	110	10	57,3	19,1	6,4	7,3
Chaco							Piedmont andin	46	19,6	69,5	2,2	6,5	2,2
AyoreoB	57	0	0	86	14	0	QuechuaB	125	15,2	64,8	16,8	3,2	0
AyoreoP	34	0	0	79,4	20,6	0	Aymara.C	120	7	57	18	16	2
Ache	63	10	90	0	0	0	<i>Ayacucho</i>	32	21,9	34,4	37,5	6,2	0
Gran Chaco	345	14,2	43,4	14,8	27,6	0	<i>Laramate</i>	72	5,6	55,6	33,3	5,6	0
Hauts-plateaux brésiliens							<i>Ollantaytambo</i>	35	8,6	65,7	22,9	2,9	0
Tuacarembo	100	21	33,9	32,2	12,9	0	<i>ChenChen</i>	23	39	39	17	4	0
Montevideo	108	36	32	18	14	0	<i>Tiwanaku</i>	13	8	15	23	23	31
Guarani	200	84	0	9,5	6,5	0	<i>Arica</i>	83	33,3	43,7	18,7	4,2	0
Kaingang	78	47	4	49	0	0	Andes méridionales						
Pampa							Atacama	73	11	68,5	15	5,5	0
Centre Argentine	310	18,5	12	42,6	26,9	0	Jujuy	100	16	59	16	9	0
Patagonie Terre de Feu							Puna	65	12	65	8	15	0
Pehuenche	205	2,4	9,8	39	48,8	0	Coyas	61	13,1	57,4	23	4,9	1,6
Huilliche - Mapuche	191	1,6	16,2	33,5	48,7	0	Salta	18	22,2	38,9	27,8	11,1	0
Chiloe	158	3,2	20,9	36,7	34,8	4,4	Nord Ouest Argentine	98	21,2	11,1	21,2	12,1	34,4
Yaghan	21	0	0	47,6	52,4	0	Catamarca/Tucuman	34	26,5	5,8	32,4	35,3	0
Sud Argentine	47	10,9	10,9	13	34,8	30,4	Centre Ouest Argentine	358	12,6	13,4	18,7	21,5	33,8
<i> Fuegian</i>	24	0	0	58,2	41,8	0							

Annexe 7 : Matrices de distances F_{ST} calculés à partir des fréquences d'haplogroupes mitochondriaux de 15 populations anciennes amérindiennes.

	Pampa Grande			Vallées	
	F_{ST}	P value		F_{ST}	P value
Arica PCT	0,04337	0,17578+-0,0113	Chen Chen	0	0,80762+-0,0138
Sican	0,05263	0,06836+-0,0073	Arica PCIni	0	0,86035+-0,0114
Vallées	0,10871	0,00977+-0,0029	Arica PCInt	0,00031	0,39941+-0,0128
Laramate PCT	0,11135	0,01270+-0,0039	Amazonie	0,01235	0,26855+-0,0159
Chen Chen	0,11437	0,01660+-0,0039	Ayacucho	0,04156	0,08398+-0,0086
Arica PCIni	0,12814	0,03613+-0,0050	Sican	0,06249	0,02637+-0,0050
Tiwanaku	0,13366	0,02832+-0,0044	Pampa Grande	0,10871	0,00977+-0,0029
Arica PCInt	0,13647	0,01758+-0,0039	Tiwanaku	0,12168	0,00586+-0,0030
Ollantaytambo	0,14975	0,00586+-0,0022	Laramate PCT	0,1465	0,00000+-0,0000
Ayacucho	0,15486	0,00293+-0,0016	Arica PCT	0,15128	0,00488+-0,0025
Palpa	0,18247	0,00195+-0,0014	Laramate PCInt	0,17966	0,00000+-0,0000
Laramate PCInt	0,21517	0,00195+-0,0014	Ollantaytambo	0,18065	0,00098+-0,0010
Amazonie	0,3218	0,00000+-0,0000	Palpa	0,31452	0,00000+-0,0000
Fuegian	0,55972	0,00000+-0,0000	Fuegian	0,50765	0,00000+-0,0000

Annexe 8 : Matrice de distances F_{ST} calculés à partir des haplotypes mitochondriaux de 44 populations sud-américaines, avec la méthode Kimura 2P, corrigé avec Slatkin.

	Pampa Grande			Vallées	
	F_{ST}	P value		F_{ST}	P value
Gran Chaco	0,07086	0,00684+-0,0027	Salta	0,01029	0,27539+-0,0144
San Martin	0,07522	0,06348+-0,0077	Wayampi	0,0395	0,03320+-0,0059
Salta	0,07719	0,05664+-0,0068	Tayacaja	0,0442	0,01562+-0,0037
Coyas	0,08237	0,00781+-0,0028	Ayacucho	0,04957	0,03613+-0,0062
La Paz	0,09036	0,00488+-0,0025	Gran Chaco	0,05014	0,00684+-0,0023
Vallées	0,09165	0,02441+-0,0050	Cayapa	0,05958	0,01074+-0,0033
Ancash	0,09432	0,00977+-0,0029	Coyas	0,06629	0,00195+-0,0014
Quechua	0,09584	0,00488+-0,0020	Kalina	0,07472	0,01367+-0,0042
Tayacaja	0,10273	0,00391+-0,0019	Tuacarembó	0,0768	0,00977+-0,0036
Ayacucho	0,10548	0,01367+-0,0031	Pehuenche	0,07916	0,01270+-0,0034
Kalina	0,11032	0,00684+-0,0023	Arawak	0,07946	0,00684+-0,0027
Laramate	0,11767	0,00391+-0,0019	Quechua	0,08237	0,00195+-0,0014
Palikur	0,11874	0,01758+-0,0037	San Martin	0,08399	0,01758+-0,0039
Titicaca	0,12799	0,00195+-0,0014	Arica	0,08708	0,00195+-0,0014
Pehuenche	0,13996	0,00586+-0,0022	Ancash	0,08817	0,00391+-0,0019
Arequipa	0,14374	0,00977+-0,0029	Moxos	0,08875	0,00098+-0,0010
Arica	0,14625	0,00000+-0,0000	La Paz	0,09049	0,00098+-0,0010
Yungay	0,1468	0,00293+-0,0016	PG	0,09165	0,02441+-0,0050
Puno	0,15195	0,00098+-0,0010	Laramate	0,10395	0,00293+-0,0016
Tuacarembó	0,15835	0,00391+-0,0019	Catamarca	0,10872	0,00000+-0,0000
Cayapa	0,16153	0,00098+-0,0010	Emerillon	0,11157	0,02051+-0,0043
Wayampi	0,16292	0,00098+-0,0010	Titicaca	0,13549	0,00000+-0,0000
Emerillon	0,1682	0,00879+-0,0029	Yungay	0,13846	0,00000+-0,0000
Yanomami	0,17956	0,00098+-0,0010	Apalai	0,13985	0,00000+-0,0000
Moxos	0,19006	0,00000+-0,0000	Arequipa	0,14313	0,00488+-0,0020
Arawak	0,19732	0,00098+-0,0010	Kaingang	0,14952	0,00000+-0,0000
Gaviao	0,19894	0,00098+-0,0010	Gaviao	0,15013	0,00000+-0,0000
Mapuche	0,20102	0,00098+-0,0010	Puno	0,15157	0,00000+-0,0000
Piedmont	0,20797	0,00000+-0,0000	Zoro	0,16046	0,00000+-0,0000
Aymara	0,20926	0,00000+-0,0000	Mapuche	0,16455	0,00000+-0,0000
Tupe	0,20952	0,01172+-0,0033	Guahibo	0,16482	0,00000+-0,0000
Catamarca	0,24491	0,00000+-0,0000	Piedmont	0,1791	0,00000+-0,0000
Palpa	0,26954	0,00000+-0,0000	Palikur	0,17924	0,00000+-0,0000
Zoro	0,27161	0,00000+-0,0000	Yanomami	0,18162	0,00000+-0,0000
Xavante	0,31686	0,00098+-0,0010	Palpa	0,21568	0,00000+-0,0000
Yaghan	0,31756	0,00000+-0,0000	Tupe	0,21989	0,00098+-0,0010
Apalai	0,32132	0,00000+-0,0000	Aymara	0,25445	0,00000+-0,0000
Kaingang	0,36426	0,00000+-0,0000	Yaghan	0,26065	0,00000+-0,0000
Guahibo	0,41036	0,00000+-0,0000	Xavante	0,29588	0,00000+-0,0000
Matsiguenga	0,43615	0,00000+-0,0000	Chibchan	0,31107	0,00000+-0,0000
Ache	0,61553	0,00000+-0,0000	Guarani	0,47614	0,00000+-0,0000
Chibchan	0,86466	0,00000+-0,0000	Matsiguenga	0,5126	0,00000+-0,0000
Guarani	1,25455	0,00000+-0,0000	Ache	0,57249	0,00000+-0,0000

Annexe 9 : Matrice des distances F_{ST} (somme des différences au carré : R_{ST}) établies à partir des données des STR du chromosome Y (sept marqueurs) de 23 populations sud-américaines de notre base de données.

	Pampa Grande			Vallées	
	F_{ST}	P value		F_{ST}	P value
Kichwa	0,01296	0,33105+-0,0170	Toba	-0,00741	0,51074+-0,0138
Mestizo	0,03483	0,23535+-0,0137	Kichwa	0,03532	0,10254+-0,0075
Diaguaita	0,03716	0,26562+-0,0154	Perou	0,04458	0,11133+-0,0126
Perou	0,04677	0,22070+-0,0123	Pampa Grande	0,06758	0,20703+-0,0092
Vallées	0,06758	0,20703+-0,0092	Quechua	0,07008	0,07031+-0,0072
Colla	0,09286	0,10840+-0,0117	Diaguaita	0,1263	0,01270+-0,0039
Quito	0,09327	0,09082+-0,0094	Aymara	0,1429	0,00488+-0,0020
Mapuche	0,1315	0,03809+-0,0034	Colla	0,1606	0,00293+-0,0016
Quechua	0,1729	0,04590+-0,0059	Mestizo	0,1961	0,00000+-0,0000
Toba	0,1758	0,03027+-0,0044	Quito	0,2358	0,00000+-0,0000
Antioquia	0,184	0,00781+-0,0028	Beni	0,2923	0,00000+-0,0000
Colombie	0,1964	0,01465+-0,0037	Mapuche	0,2975	0,00000+-0,0000
BresilSud	0,2041	0,01172+-0,0033	Antioquia	0,3368	0,00000+-0,0000
Caracas	0,2494	0,00391+-0,0019	Colombie	0,3421	0,00000+-0,0000
Bresil	0,2571	0,00391+-0,0019	BresilSud	0,3635	0,00000+-0,0000
Beni	0,2592	0,00684+-0,0027	Afro-equatoriens	0,3822	0,00000+-0,0000
Cordoba	0,2628	0,00098+-0,0010	Waorani	0,3854	0,00000+-0,0000
Afro-equatoriens	0,3081	0,00000+-0,0000	Caracas	0,3877	0,00000+-0,0000
Aymara	0,3438	0,00195+-0,0014	Bresil	0,4064	0,00000+-0,0000
Waorani	0,496	0,00000+-0,0000	Cordoba	0,438	0,00000+-0,0000
Altiplano	0,5988	0,00000+-0,0000	Altiplano	0,5896	0,00000+-0,0000
Humahuaca	0,6413	0,00391+-0,0019	Humahuaca	0,6941	0,00000+-0,0000

Publications

Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina).

Carnese FR, Mendisco F, Keyser C, Dejean CB, Dugoujon JM, Bravi CM, Ludes B, Crubézy E.

Am J Phys Anthropol;141(3):452-62. 2010.

Paleogenetical Study of Pre-Columbian Samples From Pampa Grande (Salta, Argentina)

Francisco R. Carnese,¹ Fanny Mendisco,^{2,3*} Christine Keyser,³ Cristina B. Dejean,¹ Jean-Michel Dugoujon,² Claudio M. Bravi,⁴ Bertrand Ludes,³ and Eric Crubézy²

¹Universidad de Buenos Aires, Facultad de Filosofía y Letras, Instituto de Ciencias Antropológicas, Sección Antropología, Biológica, Buenos Aires 1406, Argentina

²Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), CNRS FRE 2960, Toulouse 31000, France

³Laboratoire d'Anthropologie moléculaire EA: 4438, Institut de Médecine Légale, Université de Strasbourg, Strasbourg Cedex 67085, France

⁴Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), La Plata 1900, Argentina

KEY WORDS ancient DNA; Amerindians; mtDNA; STRs; Y-STR

ABSTRACT Ancient DNA recovered from 21 individuals excavated from burial sites in the Pampa Grande (PG) region (Salta province) of North-Western Argentina (NWA) was analyzed using various genetic markers (mitochondrial DNA, autosomal STRs, and Y chromosomal STRs). The results were compared to ancient and modern DNA from various populations in the Andean and North Argentinean regions, with the aim of establishing their relationships with PG. The mitochondrial haplogroup frequencies described (11% A, 47% B, and 42% D) presented values comparable to those found for the ancient Andean populations from Peru and San Pedro de Atacama. On the other hand, mitochondrial and Y chro-

mosomal haplotypes were specific to PG, as they did not match any other of the South American populations studied. The described genetic diversity indicates homogeneity in the genetic structure of the ancient Andean populations, which was probably facilitated by the intense exchange network in the Andean zone, in particular among Tiwanaku, San Pedro de Atacama, and NWA. The discovery of haplotypes unique to PG could be due to a loss of genetic diversity caused by recent events affecting the autochthonous populations (establishment of the Inca Empire in the region, colonization by the Europeans). *Am J Phys Anthropol* 141:452–462, 2010.

© 2009 Wiley-Liss, Inc.

Interest in the prehistory of South America has increased in recent years, in particular due to the latest archaeological discoveries which have modified our understanding of how the continent was populated (Dixon, 2001; Dillehay et al., 2008). The development of molecular biology techniques has now become an essential tool for the study of populations and their history. However, despite the greater volume of data available (Bailliet et al., 1994; Bianchi et al., 1998; Mesa et al., 2000; Moraga et al., 2000; Rodríguez-Delfin et al., 2001; Tarazona-Santos et al., 2001; Salzano, 2002), numerous questions still remain unanswered, in particular those regarding the ancient populations of South America. Indeed, studies concerning the genetic diversity of extinct populations are still rare, notably for regions such as North-Western Argentina (NWA) (Dejean et al., 2004; Goicoechea et al., 2001). Thus, questions related to the evolution of populations in this region, the relationships among them, or the processes of admixture, remain unexplored.

Argentina can be subdivided into four major regions: Central, South, North-Eastern, and North-Western, each one having a different history and evolution. The region of NWA, including the provinces of Salta, Jujuy, Catamarca, Santiago del Estero, and Tucuman, is composed of diverse environments: the Andean highlands, the inter-mountainous valleys, and the plains of the Chaco, with marked topographic and climatic variations. These physical characteristics influenced the distribution of

human populations and the relationships established among them. NWA has been characterized by a complex cultural development beginning 11,000 years ago when hunter-gatherer groups first colonized the region (Tarrago, 2000). With the development of agriculture, four phases are described in the cultural development of NWA, each one marked by an important influence of Andean culture, notably from the Aymara and

The first two authors contributed equally to this work.

Grant sponsors: ECOS/Sud/ (France—Ministère des Affaires Étrangères and Ministère de la Recherche et de l'Enseignement Supérieur), ECOS-SETCIP (Argentina—Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas CONICET), Laboratoire AMIS (Toulouse; France), Institut de Médecine Légale (Strasbourg; France), CNRS (Centre National de la Recherche Scientifique).

*Correspondence to: Fanny Mendisco, Laboratoire d'Anthropologie moléculaire, Institut de Médecine Légale, 11 rue Humann, Strasbourg Cedex 67085, France. E-mail: fanny.mendisco@neuf.fs

Received 19 April 2009; accepted 21 July 2009

DOI 10.1002/ajpa.21165

Published online 16 November 2009 in Wiley InterScience (www.interscience.wiley.com).

Quechua populations. The Early Period (500 BC–650 AD) (sometimes called the Formative) marked the appearance of agriculture and saw the development of sedentary village societies. Several local cultures (Cienaga, Candelaria) coexisted during this period. The end of this period saw the development of more complex and organized cultures, such as La Aguada, which extended its influence throughout NWA. The next phase, named the Middle Period (650–900 AD) was characterized by an important increase in population size and the appearance of extensive agriculture. The social hierarchy became more linear and important influences from the Andean highlands were observed, notably with the emergence of the Tiwanaku civilization around 600 AD (Hastorf, 2008, Leoni and Acuto, 2008). Traces of the Tiwanaku influence, mainly through material culture, were observed in the north of Bolivia to the north of Chile and Argentina. The Late Period (900–1470 AD) represented a time of regional development and increased sociopolitical complexity. Finally, the last phase before European colonization is the Imperial Period, with the arrival and the annexation of NWA by the Incas (Leoni and Acuto, 2008). This period was characterized by important population movements, in particular by the "mitimaes"; groups which moved to hostile regions to assure the ascendancy of the Incas. Bolivian Inca groups were sent by the Empire to NWA to work, but also to try and gain control of the local populations (Leoni and Acuto, 2008).

The physical characteristics of the NWA region, including the current conditions of its relative isolation, and the fluctuations and movements of its populations, many of which are small in size, make this region interesting for anthropological studies. Since the discovery of this continent by the Europeans, numerous historians and archaeologists have been interested in the Andean populations, in particular by the incredible civilizations which reigned there. However, despite a relatively good understanding of the history of the region from a cultural point of view, knowledge of the biological diversity of NWA including both its current and ancient populations is lacking. This information is essential if the history of these populations and the population settlement dynamics of the region are to be fully understood and appreciated.

Our team had the opportunity to work with remains from Pampa Grande (PG) in NWA, a region that has been influenced by various cultures throughout its history. The aim of this article is to attempt to clarify the history of the population, and at the same time, to clarify its evolution and biological structure. To this end, we decided to analyze the genetic diversity of this population with different genetic markers, and to establish biological distances with other extinct and modern South American populations.

Little information exists concerning the structure of this population before contact with Europeans at the end of the 15th century. Moreover, the lack of historical data over the postcontact period means that the exact origin and/or the degree of admixture of the inhabitants of this area are still unknown. Thus, as well as the characterization of the genetic diversity of the PG population, which is important to provide new data for the knowledge of South American settlement, we can analyze the evolution of this diversity to gauge the impact of the European colonization on the population's genetic diversity and how the population of this region has evolved. The comparison of the diversity observed at PG with modern regional populations will permit us to determine

whether PG contributed to the genetic structure of today's populations in the region.

MATERIALS AND METHODS

Site and samples

PG (Salta province, NWA) (see Fig. 1) is a region located near the Tucuman province, between the mountains of "Las Pirguas" and "Alto el Rodeo" at an altitude of 2500–3000 m (25°46' south and 65°24' west). A large survey was conducted between 1969 and 1971 in this region by Gonzalez and his team from the Museum of La Plata (National University of La Plata), and during this time the archaeological site of Las Pirguas was discovered (Gonzalez, 1972). Las Pirguas consists of several burial caves, seven of which were excavated: four caverns El Litro, II, III, and V located in Cuevitas Ravine, two others caverns I and IV located in Lampazar Ravine, and finally a great rocky shelter named Los Aparejos, located on the hillside of Pirgua Chica. In these burial caves, around 120 primary graves and secondary deposits, including 15 naturally mummified individuals, were excavated. Some of these remains were discovered in large burial urns (individual or multiple), with a varying diameter of 50 cm for immature individuals to 140 cm for adults. All the skeletal remains have been studied in the Museum of La Plata (Baffi et al., 1996). No evidence of occupation of these sites as living spaces was discovered, except in cavern IV and V situated on the margin of the ravine (Baldini et al., 1998). Thus, Las Pirguas was probably only ever used as a burial ground and represents the cemetery of a population over two to six generations (according to the cultural phase and radiocarbon dating).

The archaeological remains and particularly the funerals urns permit the site to be dated between 400 AD and 650 AD. Indeed, these remains are associated with the Candelaria cultural tradition. This is confirmed by uncalibrated radiocarbon dates obtained from skeletal remains of approximately 1310 ± 40 years BP (Beta analytic analysis number 200032). When calibrated, this date corresponds to the Candelaria period. This is a little known culture, which was described for the first time by Heredia (1969). Candelaria is an Argentinean local culture which appeared during the "Early period" (BC 300–650 AD), and remained confined to the south zone of the Salta province. It is divided into five phases, I to V, with Las Pirguas belonging to the III phase, ranging from 400 to 650 AD.

To undertake the present biological study, 21 femora were sampled. These remains are kept in the Anthropology Division of the School of Natural Sciences, and in the Museum of the University National of La Plata.

Precautions taken to avoid contamination

Given the age of the excavations, the analyzed samples had not been collected nor preserved under optimal conditions to avoid contamination. Consequently, drastic measures were taken thereafter to avoid contamination with modern DNA and to allow authentication of the results obtained. Manipulations were performed in a completely separate laboratory dedicated to ancient DNA, with UV irradiated surfaces, laminar flux and instruments, and all personnel wore gloves, lab coats, and face masks. Pipettes, plastic ware, and aerosol resistant tips were used exclusively for ancient DNA,



Fig. 1. Map of South America showing the locations of PG and other sites mentioned in the text. 1: PG (this study); 2: San Pedro de Atacama (Moraga et al., 2005); 3: Tiwanaku (Rothhammer et al., 2003); 4: Ancash (Lewis et al., 2004); 5: Arequipa, 6: Tayacaja, 7: San Martin (Fuselli et al., 2003); 8: Aymara, 9: Quechua (Merriwether et al., 1995; Bert et al., 2001); 10: Gran Chaco (Demarchi et al., 2001; Cabana et al., 2006x); 11: Movima, 12: Ignaciano, 13: Trinitario, 14: Yuracare (Bert et al., 2001, 2004); 15: Atacameños, 16: Huilliches (Bailliet et al., 1994); 17: Kichwas (Gonzalez-Andrade et al., 2008); 18: Cerro Largo (Sans et al., 2006); 19: ancient Peruvians (Shinoda et al., 2006); 20: Puenenos (Dipierrri et al., 1998).

and blank reactions were always included in PCR protocols. The experimental areas of pre- and post-PCR were strictly separated. Moreover, for each sample, multiple independent extractions and PCR amplifications were realized. All personnel involved in the sample processing were tested using the same techniques to compare the DNA profiles obtained to those of the PG population (Keyser-Tracqui et al., 2003).

DNA extraction and purification

For each of the 21 samples, DNA was extracted following the protocol established by Keyser-Tracqui et al. (2003). The outer bone surface (1–2 mm) was removed with a sanding machine (Dremel) and then powdered bone was obtained by grinding bone fragments under liquid nitrogen in a 6800 Freezer Mill (Fischer Bioblock).

Two grams of the pulverized material was incubated at 50°C overnight in 5 ml of a solution containing 5 mmol EDTA, 2% SDS, 10 mmol Tris HCL (pH 8.0), 0.3 mol sodium acetate, and 1 ml proteinase K/ml. A phenol/chloroform/isoamyl alcohol extraction on the supernatant was performed. The aqueous phase was purified with the Cleanmix kit (Talent). After the elution, DNA was concentrated by passing it through a Microcon YM30 filter (Millipore). For each sample, three separate extractions were done.

Real-time PCR

To determine the quantity and presence of PCR inhibitors, a real time PCR was performed. The reaction was done employing Quantifiler Human DNA Quantification kit (Applied Biosystems), following the protocol provided by the manufacturers.

Autosomal STR analysis

Nine autosomal short tandem repeats (STR) and the amelogenin sex marker were amplified using the AmpFISTR Profiler Plus Kit (Applied Biosystems), which analyzed the STR: D3S1358, D8S1179, D5S818, VWA, D21S11, D13S137, FGA, D7S820, D18S51. Multiplexes were performed according to the manufacturer's procedure (Applied Biosystems). Each sample was amplified at least twice for each extraction.

Y chromosome STR analysis

The DNA of the male ancient specimens were analyzed at 11 STR loci (DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) from the nonrecombining portion of the Y chromosome, employing a Promega Kit, Power Plex Y. Analyses were performed according to the manufacturer's recommendations (Promega). Determinations were done at least twice for each extraction.

Mt DNA analysis

Analyses of the HVI control region of the mitochondrial DNA were performed in the laboratory of the Institut de Médecine Légale, Université de Strasbourg in France. Amplifications were done in two overlapping portions, using primers F15989 (Gabriel et al., 2001), R16239 (Ivanov et al., 1996), and F16190/R16410 (Gabriel et al., 2001). When B haplogroup was detected, an alternative reverse primer was employed for the first segment: 16167 (5'-GGGTTTGATGTGGATGGG-3') (Ricaud et al., 2004). PCR was performed with Hot start Taq-polymerase (Eurogentec) as follows: predenaturation at 94°C for 10 min, followed by 38 cycles of 94°C for 30 s, 48°C or 51°C for 30 s, and 72°C for 45 s, with a final extension at 72°C for 10 min. Amplification products were visualized on a 1% of agarose gel and purified with Microcon-PCR filter (Millipore). The sequence reaction was carried out with the same primers for each strand with the ABI Prism Big Dye Terminator Kit (Applied Biosystems), employing forward and reverse primers in two different reactions. The products were purified by ethanol precipitation and the DNA sequence obtained was analyzed in an ABI Prism 3100 Genetic Analyzer. Consensus sequences were obtained comparing the amplified results for each one of the three extractions.

Restriction enzymes, including -5176 *AluI* (to detect haplogroup D), and the mtDNA 9-bp deletion (in the COII/tRNA^{lys} region) were used to determine and confirm the haplogroup determination (this step was carried out in the Sección Antropología Biológica, Instituto de Ciencias Antropológicas, Centro de Genética, Facultades de Filosofía y Letras y Ciencias Veterinarias, Universidad de Buenos Aires).

Data analysis

To carry out comparative analyses (haplotype and haplogroup frequencies), data available in the literature on modern and ancient South American populations were incorporated into the study.

Biological distances between the populations were estimated from mtDNA haplogroup frequencies using pairwise F_{ST} in the Arlequin program package (Shneider et al., 2000), which also calculated the genetic diversity and performed an analysis of molecular variance (AMOVA). To better visualize the F_{ST} results, a principal component analysis (PCA) based on mtDNA haplogroup frequencies was constructed as a harmonic image. Moreover, the haplotypes detected were compared to a database of about 5000 South American mtDNA sequences (Bravi, personal database), with the aim of visualizing the haplotype distribution. Median networks for each mtDNA haplogroup described were generated with the median joining algorithm (Bandelt et al., 1995, 1999) using the NETWORK 4.5 software program (Fluxus Technology Limited). The weighting scheme for the nucleotide positions used in this analysis (nps 16,101–16,376) followed Richards et al. (1998).

For the Y chromosome, a personal database of about 8000 Y-STR haplotypes covering all of America was used to interpret the haplotype distribution. In addition, Haplogroup Predictor (<http://home.comcast.net/~hapest5/index.html>) was used for inferring haplogroup status of Y-STR profiles using flat a priori probabilities. Then, we searched and compared in the YHRD (<http://www.yhrd.org/index.html>) for each of the profiles detected, assuming that the number of occurrences of particular profiles in a worldwide database could indicate their most natural geographical origin.

RESULTS

Evaluation of authenticity

We used a series of criteria to evaluate the authenticity of the ancient DNA obtained in this analysis. Indeed, the decision to analyze autosomal STR was essentially based on their ability to indicate the authenticity of the profiles (Keyser-Tracqui et al., 2003). The comparison of the PG's amplified products to the profiles of the persons involved in this study underlines the absence of contamination (data can be obtained upon request). Furthermore, the length of the obtained fragments (rather short, near 200 bp) is coherent with the characteristics of ancient DNA. The coherence between the morphological and genetic sex determination (in our case, approximately 80%) is another criterion in favor of the authenticity of the obtained results (Mooder et al., 2006). Finally, the mtDNA and Y-STR haplotype analysis revealed that all the samples' polymorphisms were in accordance with the geographic location under study (haplotypes found in Amerindian populations). Median joining networks were built for the different mitochondrial haplogroups to evaluate the

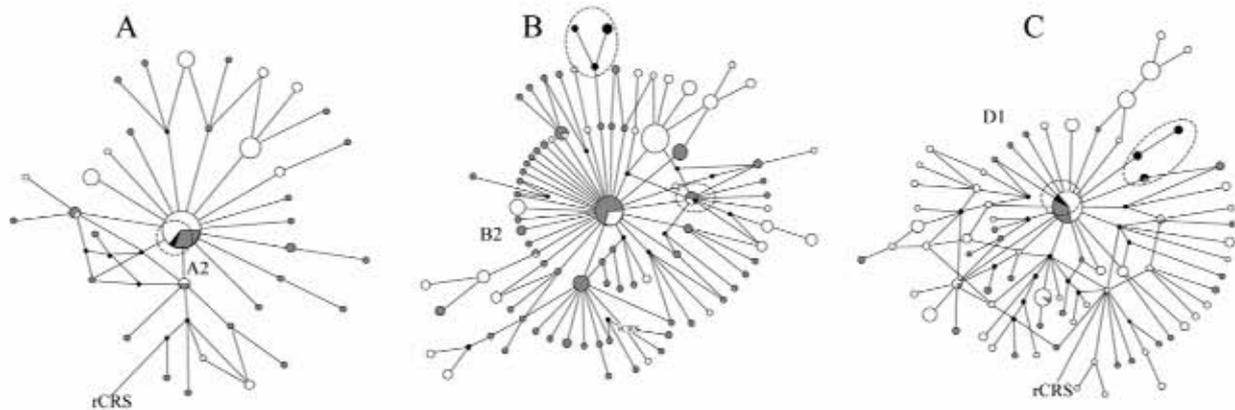


Fig. 2. Median joining networks representing South American mitochondrial sequences of the haplogroup A (A), the haplogroup B (B), and the haplogroup D (C). The sequences detected for the individuals of PG are surrounded in black. The white circles represent non-Andean populations and gray circles represent Andean populations.

TABLE 1. Autosomal allelic profiles of Pampa Grande individuals

Site	Sample	Amg. ^a	MS ^b	D3S135	D5S818	D7S820	D8S1179	D13S31	D18S51	D21S11	FGA	VWA
El Litro	17838	XY	XY	15/15	11/11	11/13	9/14	9/13	—	29/32.2	21/23	16/17
	17863	XY	XX	15/15	11/11	—	13/13	9/12	—	—	—	17/17
	17895	XY	XX	17/17	7/9	—	10/13	11/13	(12/12)	31.2/32.2	23/25	16/18
	18432	XY	XY	15/15	9/11	11/12	13/15	12/13	15/18	30/31	21/22	16/17
Cavern II	17825	XY	XY	15/15	10/11	11/14	13/13	9/13	(14/14)	31.2/33.2	20/25	16/17
	17843	XY?	XX	—	—	—	—	—	—	30/38	(23)/24	—
	17864	XX	XX	15/15	10/11	—	10/14	9/13	13/13	32.2/33.2	21/25	16/18
	17894	XX	XX	14/16	11/11	11/11	12/13	10/13	42339	30/30	21/22	15/16
Cavern V	17809	XY	Ind.	14/15	—	—	—	—	—	24.2/35	—	19/19
	17890	XY	Ind.	15/17	9/9	—	13/13	9/13	(13/13)	30/31.2	21/21	15/17
	17891	XY	XY	16/17	8/9	11/11	12/13	9/13	—	30/30	20/24	16/16
	18414	XX	XX	15/15	9/11	—	11/15	9/12	14/15?	—	22/25	16/17
Cavern IV	17824	Ind.	XY	—	—	—	—	—	—	—	—	—
	17885	XY	XY	15/15	11/11	—	12/14	12/13	15/15	28/30	21/25	17/19
	18365	XY	XY	15/15	11/11	—	11/12	9/12	(12/14)	31.2/31.2	21/24	15/17
	18417	XY	XY	15/15	11/11	—	14/14	12/12	(18/18)	—	22/26	17/19
	18430	XX	XX	15/17	9/12	—	14/14	12/14	(12/16)	31.2/31.2	20/24	16/17
Los Aparejos	17842	XX	XX	16/16	9/12	—	13/15	9/10	(16/16)	30/32.2	20/26	16/20
	17886	XY	XY	15/18	9/11	(11/12)	11/15	12/13	13/15	31/31.2	21/25	16/16
	18416	XX	XY	15/15	—	—	—	—	—	—	—	17/17
	18426	XY	XY	15/16	9/12	(11/11)	13/14	11/12	14/15	28/31	25/26	16/17

() not included in the analysis (ambiguity could not be eliminated even after reiteration of the experimentation).

^a Amg., amelogenin.

^b MS, morphological sex.

molecular affinity of PG's sequences with South American populations (see Fig. 2).

To remove doubts concerning the validity of the results, data were kept for the analyses only if reproducible PCR results were obtained from multiple extractions and amplifications of the same samples made at different times. The samples presenting discordant RFLP and HVI markers were excluded from the analysis. As regards the mitochondrial lineages, the haplotypes of two samples were interpreted as a mixture between endogenous sequences of haplogroup A and B respectively, and modern rCRS-type lineage. These two sequences were excluded from the analysis.

Autosomal STRs

DNA quantification showed that concentrations ranged from 0 ng (17824) to 3.66×10^{-1} ng (18365). No PCR

inhibitor was detected in DNA extracts. For all determinations we obtained reproducible results in about 80–85% of the samples.

Of the 21 samples analyzed, four were too degraded: for sample 17824 no amplifiable product could be observed and for the three other samples, only two (17843, 18416) and three (17809) alleles could be typed. The remaining 17 samples present more or less complete allelic profiles (Table 1). To ensure that all the heterozygote alleles were detected, several amplifications were realized (4–5 times) for samples presenting genotypes apparently homozygous. Morphological and genetic sex determination was not in accordance for four samples (17863, 17895, 17843, and 18416). We indicated earlier that samples 18416 and 17843 were too degraded, which can explain this difference of results. For samples 17863 and 17895, we can suppose an error in the morphological sex determination.

ANCIENT DNA FROM PAMPA GRANDE

457

TABLE 2. Y-STR haplotypes detected in the PG samples

Hp	Sample	DYS19	DYS385	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	Hg	Prob. (%)
I	17838	13	14/18	13	29	24	11	14	13	14	11	11	Q	99.9
II	17895	14	15/16	14	32	24	10	-	13	14	11	12	Q	62.4
III	18432	13	13.2/17	12	28	25	11	14	13	14	11	11	Q	99.9
IV	17825	14	15/16	13	31	25	10	14	13	14	11	12	Q	99.9
V	17890	14	15/16	14	32	25	10	14	13	14	11	12	Q	99.5
VI	17885	14	15/18	13	29	24	11	-	13	14	11	13	Q	60.8
VII	18365	-	14/18	13	29	24	11	-	13	14	11	11	Q	85.4
VIII	18417	-	15/18	13	29	24	10	14	13	14	11	-	Q	100
IX	17886	14	15/17	13	29	24	11	-	13	14	11	12	Q	88
X	17891	14	15/17	13	-	-	11	-	13	14	11	12	Q	61.7

- means that amplification cannot be done.

In this population, the number of alleles by locus varied from 4 to 10, with an average of 6.2 ± 1.7 , which is lower than what is reported in literature for South American populations (7.9 ± 2.7 in the Calchaqui valley (Acreche, 2003), 12 ± 3.9 in Amazonian-Orinoquian groups (Paredes et al., 2003). The rather poor amplification for some loci (D7S820 or D18S51 for example) can explain such low values.

Our study is one of the first on autosomal STRs from a pre-columbian South American population; therefore, a comparative analysis was performed with allelic frequencies of extant populations of South America (Gonzalez-Andrade et al., 2003; Barrot-Feixat et al., 2004; Gonzalez Martin et al., 2004; Jaime et al., 2004; Goulart Lanes et al., 2008). We compiled data from populations with the highest possible indigenous composition, such as samples from the Puna and Calchaqui valley (Albeza et al., 2002). The alleles D3S135*15, D5S818*11, D8S1179*13, D13S31*12:13, D18S51*15, D21S11*30, FGA*21, and VWA*16:17 present the highest frequencies. These alleles are also the most frequent in most of the populations of South America (Kichwas, Puna, or Colombian Andean, Albeza et al., 2002; Paredes et al., 2003; Gonzalez-Andrade et al., 2006), except for loci D18S51 and FGA. It shows that the allelic frequencies compared between the ancient population of PG and the South American current populations are similar. Even if we noted a weaker allelic diversity, which could be explained by the incomplete profiles obtained, the genetic diversity from PG does not seem so different from the modern genetic diversity.

The comparison of all the pairs of complete and incomplete profiles reveals no concordance between PG's individuals.

Y chromosome STR analysis

Y chromosome STR analysis was carried out on 14 samples (including 17843 which presented a doubtful result (XY?) for sex determination). The combination of the 11 STRs analyzed allowed us to construct the haplotype of each individual (Table 2). It is noted that no locus was amplified for sample 17843, which may correspond to a XX individual or to a highly degraded DNA sample. As for the autosomal STRs, sample 17809 did not give good results because no loci were amplified. For sample 17863 only two loci were amplified making the designation of one haplotype impossible. In the remaining 10 samples the number of loci amplified varied from 8 to 11, and 10 different haplotypes were identified, each one corresponding to only one individual (Table 2). In these

samples, we can see that the number of alleles by loci varied from 1 to 2 with an average of 1.8 ± 0.4 , which is lower than for the South American populations found in the literature (3.8 ± 0.8 in Chaco groups and 4.4 ± 1.8 in Mexico Amerindians (Paez-Riberos et al., 2006; for example). The alleles DYS19*14 (67%), DYS389I*13 (73%), DYS389II*29 (50%), DYS390*24 (60%), DYS391*10/11 (50%), DYS392*14 (83%), DYS393*13 (100%), DYS437*14 (100%), DYS438*11 (92%), and DYS439*12 (50%) present the highest frequencies.

Comparisons of each of the haplotypes with the various databases revealed no strict matches. For sample 17825, four profiles with one mismatch (at the locus DYS385b which is a very variable site) were detected, which were described in a modern Equatorial population (Quito, Kichwas) (Baeza et al., 2007; Gonzalez-Andrade et al., 2008). For the remaining samples, only corresponding profiles with two or three mismatches were observed. The minimal haplotype (DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS437) was also compared with the YHRD database (approximately 23,075 haplotypes belonging to 200 worldwide populations, in <http://www.ystr.org/index.html>). Still, no perfect correspondence was observed for the PG individuals.

Moreover, with the databases available on the Internet (Haplogroup Predictor: <http://www.hprg.com/hapest5/and> World Haplogroup & Haplo-I Subclade Predictor: <http://members.bex.net/jtcullen515/haplotest.htm>), we could estimate from the allelic frequencies the correspondence with a particular haplogroup. With the different databases we found, for each haplotype, between 60 and 100% to belong to haplogroup Q (Table 2), which is the major lineage among the Native Americans. The distribution of this haplogroup outside the Americas is limited: it is found at high frequencies in some Siberian groups, and at low frequencies in Europe, East Asia, and the Middle East (Karafet et al., 2008).

HVS-1 lineages and RFLP analysis

As we can see in Table 3, nine different mitochondrial haplotypes were observed for the 19 samples (as indicated earlier, 2 of 21 samples were excluded from the study). The examination of the different haplotypes showed that the haplotype number 1, 2, and 7 (Table 3) carry the HVS-1 founder mutations for haplogroups A2, B2, and D1 (Forster et al., 1996; Bandelt et al., 2003). These haplotypes are dispersed in the American continent and thus, are not very good indicators for biological and phylogeographical relations. The remaining haplotypes are derived from these founder haplotypes, by one

TABLE 3. mtDNA haplotypes of PG samples

Hp #	Sample code	Control region: HVS-I	RFLP coding region		Hg.	Total
			9-bp	AluI		
1	18414, 18365	111 223 290 319 362			A2	2
2	17886	183C 189 217	+		B2	1
3	18430	142 182C 183C 189 217	+		B2	1
4	17838, 18416	145 156 157 183C 189 217	+		B2	2
5	17864	145 156 157 182C 183C 189 217	+		B2	1
6	17825, 17894, 17890, 17891	145 156 157 183C 189 217 278	+		B2	4
7	18432, 17863, 17809	223 325 362		-	D1	3
8	18417, 17885	129 223 325 362		-	D1	2
9	17895, 17842, 18426	223 287 325 362		-	D1	3

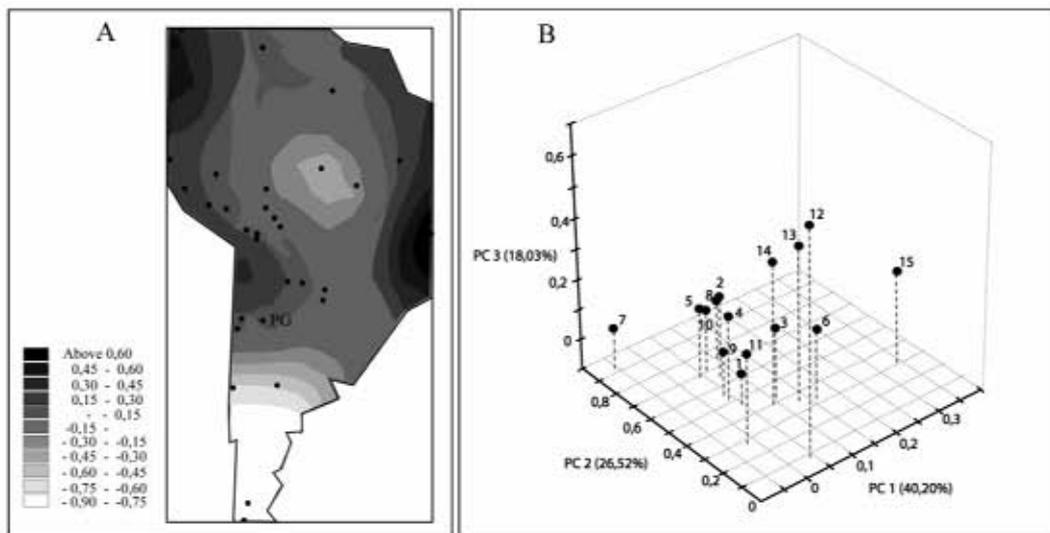


Fig. 3. (A) Harmonic image and (B) three-dimensional PCA of 15 populations based on the mitochondrial haplogroup frequencies. The populations included in this analysis are as follows: 1: PG, 2: Punenos, 3: Gran Chaco, 4: Ancash, 5: Arequipa, 6: Tayacaja, 7: Aymara, 8: Quechua, 9: SanMartin, 10: Atacamenos, 11: Huilliche, 12: Movima, 13: Ignaciano, 14: Trinitario, 15: Yuracare. (For references and location see Fig. 1).

to four mutations. The comparison analysis shows that there is an exact match between haplotype 3 and one Uruguyan individual from Cerro Largo (sample 21 in Sans et al., 2006). An exact match was also discovered between haplotype 8 and one Taino of Dominican Republic (Lalueza-Fox et al., 2001) and with one Tayacaja of Peru (Fuselli et al., 2003). However, a monophyletic origin for these three lineages cannot be asserted based on the occurrence of a single position known to be hyper-variable (16129). Moreover, near matches (one difference) were found between the haplotypes 3, 5, and 6 and individuals of the Peruvian Andes described by Fuselli et al. (2003) and Shinoda et al. (2006), for haplotype 6.

The RFLP analysis confirmed the previous haplogroup assignment by HVS-1 lineages. The 9-bp deletion was observed in all samples defined as B haplogroup by sequencing. For the assignment of D haplogroup, the absence of *AluI* site at position 5176 was confirmed for 7/8 of the samples D (for one of the samples there was not amplification) (Table 3). Thus, the haplogroup distribution in the PG population is the following: 11% of A2 haplogroup, 47% of B2, and 42% of D1.

Statistical analysis

To visualize the biological affinities between PG and modern Amerindian populations, a three-dimensional PCA and a harmonic image were performed from the mitochondrial haplogroups frequencies (see Fig. 3). The harmonic image (showing the first principal component: about 40% of the variability) underlines the different genetic structure of the Patagonian populations (because of the high frequencies of the D haplogroup). Concerning PG, this image illustrates the genetic proximity with the current Andean and sub-Andean populations. The PCA allows to centre the study on the Andean region, and thus, to place PG in this presumably homogeneous region. This analysis accounts for approximately 85% of the variability observed. The first and the third components (40% and 18% of the variance) show a certain separation between the sub-Andean populations and the others. The other Andean populations, including PG, show biological proximity. We performed AMOVA on the same populations, geographically and culturally close to the population of PG.

We can observe that most of the genetic variation (~90%) appears within the populations, and the variation between populations is negligible (~4%). Moreover, the genetic differentiation measured by F_{ST} distances between these populations shows a relatively high level of genetic homogeneity (0.09).

DISCUSSION

Despite the age of the excavation and the nonoptimal storage conditions of samples, the conservation of the DNA was good, due to the climatic conditions (low temperature and humidity) of this mountainous region, and ancient genetic profiles could be obtained. We realize that the analysis of autosomal STRs is an interesting criterion for the validation of the obtained results. Indeed, this marker allowed the sequences obtained to be compared with those of the researchers' who had contact with the bone remains during analyses. The comparison of the morphological and molecular sex data also contributed to this authentication.

Links between PG and Andean populations

Regarding the genetic diversity observed among PG individuals, the analyzed markers allowed us to evaluate this group's affinities with extinct and extant populations of the region.

The analysis of maternal lineages showed that very few haplotypes described for these ancient individuals had been described previously. Indeed, close or total concordance was found with one present day Uruguayan individual (sample 3), three present day Quechuas (sample 3, 5, and 8), and two ancient individuals, one ancient Taino (sample 8) and one ancient Peruvian from Paucarcancha (sample 6). The Andean populations which were concordant were the Tayacaja and San Martin, both, native populations of the Peruvian Andes, (Fuselli et al., 2003). The PCA (see Fig. 3) shows the proximity between PG and the San Martin and Tayacaja groups. The sharing of specific haplotypes with Peruvian individuals makes sense, considering that NWA belongs to the Central Andean cultural complex of food producers and pastoralists. However, it can be noted that despite the large number of sequences available for comparison, the individuals from PG share very few lineages with other populations. In particular, it is interesting to observe the lack of matches in extant populations geographically close to PG, like the Coyas from Salta and Jujuy (Alvarez-Iglesias et al. 2007), or the populations of the Argentinean Gran Chaco (Ach , Mataco, Pilaga, and Toba) (Dornelles et al., 2004; Schmitt et al., 2004; Cabana et al., 2006). The analysis of the haplogroup frequencies shows more clearly a genetic affinity between the individuals of PG and the Andean populations. We can observe that the proportions of haplogroups B and D described in this ancient population are very significant. Such a high frequency of B haplogroup is characteristic of Andean populations, for both current populations as well as for ancient groups. The ancient "Peruvian highlanders" (Shinoda et al., 2006) are supposed to be of an indigenous highland origin and present a frequency of haplogroup B of about 66%. Moraga et al. (2005) has described a frequency of 42% of haplogroup B for a prehistoric population of the

Middle Horizon (1800–1300 BP) in Northern Chile. This is interesting because the period corresponds to the Candelaria cultural period found at PG. Moraga et al. (2005) concluded that the genetic diversity of this ancient Chilean group (San Pedro de Atacama, Fig. 1) reflects gene flow from the Andean highlands to the valleys of the south regions of the Andean complex. High frequencies of haplogroup B are also detected in modern populations of the central Andes area, like the Aymara, Quechua, Atacameno, Ancash, Mapuche, Huilliche (Baillet et al., 1994; Merriwether et al., 1995; Rodriguez-Delfin et al., 2001; Lewis et al., 2005). Concerning haplogroup D, such frequencies can be observed in populations of the Gran Chaco (Toba and Wichi) (Dornelles et al., 2004; Cabana et al., 2006) and also in Patagonian groups, like the Tehuelche or Mapuche (Baillet et al., 1994; Merriwether et al., 1995). Indeed, the statistical analysis carried out (see Fig. 3) based on the mitochondrial haplogroups frequencies highlights the genetic similarity of Andean populations. The remarks aforementioned are in agreement with observations made by various authors such as Fuselli et al. (2003), Luiselli et al. (2000) or Tarazona-Santos et al. (2001) who have previously shown the biological homogeneity of the Andean populations, contrasting with the heterogeneity of Amazonian populations.

Affinities of the paternal lineages are less clear, which can be due to the paucity of available data on the Y-STR of Amerindian populations limiting the comparisons which can be made (Fondevila, et al., 2003; Garc a-Bour et al., 2004; Tovar et al., 2006; Lee et al., 2007; Leite et al., 2008; Tirado et al., 2008; Toscanini et al., 2008). However, the allelic diversity and haplotype distribution observed seem to link PG with Amerindian populations like the Equatorian Kichwas (Gonzalez-Andrade et al., 2008). This population is known to be related to the Andean Quechua. The relatively low Y chromosome diversity could be explained by the small sample size, by differential patterns of male-female gene flow or by patrilineal social patterns. The comparatively low Y chromosome diversity of patrilocal societies was demonstrated by Oota et al. (2001) through a comparison with matrilineal groups.

The affinities established through the analysis of the distributions of each haplotype and haplogroup are consistent with the geographic situation of PG. As such, this study also provides independent evidence from genetics on the relationships established from material culture (influence of the Andean Bolivian culture on Candelaria material culture). Indeed, the similarities mentioned earlier for the maternal lineages, between ancient populations of North Chile and those of PG, allow us to hypothesize that exchanges and relations between these two ancient populations occurred very early (Leoni and Acuto, 2008). This can be put in relation to the complex social and cultural organization of the region from the appearance of the first civilizations such as Tiwanaku (Hastorf, 2008). It thus appears that movements of populations for trade or different exchanges led to gene flow which resulted in the linguistic, cultural, and biological homogeneity observed. These exchanges were favored by the fact that no barrier to gene flow was found in this region (Luiselli et al., 2000). It is however still difficult to assess if the observed homogeneity is only due to the extent of gene flow. It would be interesting to know if characteristics such as high frequencies of B haplogroup in the entire Andean region

are due to the numerous exchanges between populations or to an identical starting gene pool for all these populations (the first hypothesis not excluding the other one).

Structure and evolution of the PG population

Associated to these observations, we also note that despite the small sample size of our study, the genetic diversity (based on the mtDNA haplogroup frequencies) observed seems comparable to that of modern Amerindian populations, and in particular to Andean populations (see Fig. 3). Nevertheless, the distribution of haplotypes, for mtDNA as well as for Y chromosome, seems specific to PG. Indeed, very few haplotype correspondences were discovered, despite the large comparative sample. Different hypotheses can explain this difference in haplotypic variability with the contemporary Andean populations. The first explanation relates to the absence of thorough sampling of these lineages; the studied populations of NWA were small, which can cause an important bias in the study. However, given the size of the comparative databases, about 5000 mitochondrial sequences and 8000 Y chromosome haplotypes, other explanations can be evoked. Indeed, a loss of genetic diversity could explain the peculiarity of the haplotypes discovered at PG. These lineages could have been lost in the Andean populations, due, for example, to recent demographic or historical events. The establishment of the Inca civilization followed by the arrival of European colonists would have had a major impact on the demography of the indigenous populations, and it could be associated to the disappearance of these lineages. The revolt of certain NWA populations against the rule of the Incas would have led to important social tensions in the region (Acuto, 2008). Furthermore, during the arrival of the Europeans, these populations would have strongly resisted, leading to a significant decrease in population size. However, several authors (Marrero et al., 2005; Wang et al., 2008) showed that the region of the NWA, and particularly the province of Salta, is one of the regions with the largest contribution of native Americans to the current gene pool. The native populations of the highlands and of the more inaccessible regions were able to preserve their genetic variability over time. Moreover, the Andean region has maintained a large effective population size, which favored the preservation of genetic diversity. Therefore, a more widespread sampling of the current populations of NWA could confirm or not the presence of particular lineages in this ancient group from PG, so deducing the correct hypothesis.

CONCLUSION

Successful ancient DNA extraction and amplification of various complementary genetic markers allowed us to characterize from a biological point of view individuals from the ancient population of PG. The PG samples present genetic similarities to other Andean populations, in particular when considering the frequencies of mitochondrial haplogroups. In spite of the Candelaria culture of PG being local and specific, we can hypothesize that gene flow between Andean populations, facilitated by an important cultural network, allowed the genetic similarity between populations of the region to be maintained. We can also hypothesize a common starting gene pool for all the populations of the Andean region.

The population of PG (400–650 AD) presented similarities concerning haplogroup frequencies, but presented a specific haplotypic diversity in comparison with contemporary Andean populations. This is very interesting because it could show a loss of genetic diversity in the modern population caused by more recent events such as the arrival of the Europeans. However, we cannot assert that this observation is not simply due to the lack of data concerning the populations of this region.

ACKNOWLEDGMENTS

We thank the Museum of the University of La Plata for the availability of the samples.

LITERATURE CITED

- Acreche N. 2003. *Microevolución de Poblaciones Andinas*. (dissertation) Universidad Nacional de Salta.
- Acuto FA. 2008. Experiencing Inca domination in Northwestern Argentina and the Southern Andes. In: Silverman H, Isbell WH, editors. *Handbook of South American archaeology*. New York: Springer. p 845–864.
- Albeza MV, Picornell A, Acreche N, Tomas C, Castro JA, Ramon MM. 2002. Genetic variability at 14 STR loci in the Puna population of north western Argentina. *Int J Leg Med* 116:126–132.
- Alvarez-Iglesias V, Jaime JC, Carracedo A, Salas A. 2007. Coding region mitochondrial DNA SNPs: targeting east Asian and Native American haplogroups. *Forensic Sci Int Gen* 1:44–55.
- Baeza C, Guzman R, Tirado M, Lopez-Parra AM, Rodriguez T, Mesa MS, Fernandez E, Arroyo-Pardo E. 2007. Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Sci Int* 173:214–219.
- Baffi EI, Torres MF, Cocilovo JA. 1996. La población prehispánica de Las Piguas (Salta, Argentina) Un enfoque integral. *Revista Argentina de Antropología Biológica* 1:204–218.
- Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO. 1994. Founder mitochondrial haplotypes in Amerindian Populations. *Am J Hum Genet* 54:27–33.
- Baldini M, Baffi EI, Togo J. 1998. Abrigos y cavernas que hacen historia : los hallazgos de Las Piguas (Pampa Grande, Salta). In: *Fac. de Filosofía y Letras, editors. Homenaje a Alberto Rex Gonzalez*. Buenos Aires: FADA, UBA. p 343–359.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.
- Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang YP. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 67:512–524.
- Barrot-Feixat C, Gonzalez-Martin A, Ortega M, Sanchez C, Gorostiza A, Huguet E, Brandt-Casadevall C, Corbella J, Géné M. 2004. Hna hnu Amerindian population from La Sierra Otomi-Tepehua (Mexico) characterised by 14 STR-PCR polymorphisms. *Int Congr Ser* 1261:179–181.
- Bert F, Corella A, Géné M, Pérez A, Turbon D. 2001. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum Biol* 73: 1–16.
- Bert F, Corella A, Géné M, Pérez-Perez A, Turbon D. 2004. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. *Ann Hum Biol* 31:9–28.
- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS. 1998. Characterization of ancestral and derived Y-chromo-

ANCIENT DNA FROM PAMPA GRANDE

461

- some haplotypes of New World native populations. *Am J Hum Genet* 63:1862–1871.
- Cabana GS, Merriwether DA, Hunley K, Demarchi DA. 2006. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol* 131:108–119.
- Dejean CB, Crouau-Roy B, Goicoechea AS, Avena SA, Carnese FR. 2004. Genetic variability in Amerindian populations of Northern Argentina. *Genet Mol Biol* 27:489–495.
- Demarchi DA, Panzetta-Dutari GM, Motran CC, Lopez de Basualdo MA, Marcellino AJ. 2001. Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. *Am J Phys Anthropol* 115:199–203.
- Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320:784–786.
- Dipierrri JE, Alfaro A, Martínez-Marignac VL, Baillet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001–1010.
- Dixon EJ. 2001. Human colonization of the Americas: timing, technology and process. *Quaternary Sci Rev* 20:277–299.
- Dornelles CL, Battilana J, Fagundes NJR, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479–488.
- Fondevila M, Jaime JC, Salas A, Lareu MV, Carracedo A. 2003. Y-chromosome STR haplotypes in Córdoba (Argentina). *Forensic Sci Int* 137:217–220.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–945.
- Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 20:1682–1691.
- Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. 2001. Improved mtDNA sequence analysis of forensic remains using “mini-primer set” amplification strategy. *J Forensic Sci* 46:247–253.
- García-Bour J, Pérez-Pérez A, Álvarez S, Fernández E, López-Parra AM, Arroyo-Pardo E, Turbón D. 2004. Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-Chromosome STR characterization. *Am J Phys Anthropol* 123:361–370.
- Goicoechea AS, Carnese FR, Dejean C, Avena SA, Weimer TA, Franco MH, Callegari-Jacques SM, Estalote AC, Simões ML, Palatnik M, Salzano FM. 2001. Genetic relationships between Amerindian populations of Argentina. *Am J Phys Anthropol* 115:133–143.
- Gonzalez AR. 1972. Descubrimientos arqueológicos en las serranías de Las Pirguas, Prov. de Salta. *Revista de la Universidad Nacional de La Plata* 24:388–392.
- Gonzalez-Andrade F, Bolea M, Martínez-Jarreta B, Sanchez D. 2006. Genetic analysis of autosomic and Y-chromosome STR-PCR polymorphisms of the Huaorani, the last nomad tribe in Ecuador. *Int Congr Ser* 1288:331–333.
- Gonzalez-Andrade F, Roewer L, Willuweit S, Sanchez D, Martínez-Jarreta B. 2008. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waorani. (Internet). *FSIGEN* 388:1–9. Available at: <http://www.forensic-source.org/periodicals/fsigen/inpress>.
- Gonzalez-Andrade F, Sanchez D, Martínez-Jarreta B. 2003. Genetic profile of the Ecuadorian Mestizo population (Ecuador-South America) by using the Power Plex 16 System Kit. *Forensic Sci Int* 135:64–66.
- Gonzalez Martin A, Sanchez C, Barrot C, Ortega M, Brandt-Casadevall C, Gorostiza A, Huguet E, Corbella J, Gené M. 2004. Huastecos Amerindian population (Mexico) characterised by 12 STR-PCR polymorphisms. *Int Congr Ser* 1261:182–184.
- Goulet Lanes Chula F, Rodenbusch R, Schumacher S, Grandi T, Tosi Michelon C, Zoratto Gastaldo A, Costia C, Carvalho B, Dornelles da Silva CM. 2008. 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil. *Forensic Sci Int* 3:35–38.
- Hastorf CA. 2008. The formative period in the Titicaca basin. In: Silverman H, Isbell WH, editors. *Handbook of South American archaeology*. New York: Springer. p 545–562.
- Heredia O. 1969. Consideraciones sobre el conversero y la cronología de la cultura Candelaria. *Ciencia e Investigación XXV*, 3:384–406.
- Ivanov PL, Wadhams MJ, Holland MM, Weedn VW, Parsons TJ. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12:417–420.
- Jaime JC, Fondevila M, Salas A, Lareu MV, Carracedo A. 2004. Nine autosomal STRs genotype profiles in a sample from Córdoba (Argentina). *Forensic Sci Int* 139:81–83.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838.
- Keyser-Tracqui C, Crubézy E, Ludes B. 2003. Nuclear and mitochondrial analysis of a 2000-year-old necropolis in the Egiyn Gol Valley of Mongolia. *Am J Hum Genet* 73:247–260.
- Lalueza-Fox C, Calderon FL, Calafell F, Morera B, Bertranpetit J. 2001. MtDNA from extinct Tainos and the peopling of the Caribbean. *Ann Hum Genet* 65:137–151.
- Lee J, Ewis AA, Hurler ME, Kashiwazaki H, Shinka T, Nakahori Y. 2007. Y chromosomal STRs haplotypes in two populations from Bolivia. *Legal Med* 9:43–47.
- Leite F, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CHF, Raimann PE, Schwengber SP, Sortica VA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH. 2008. Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol* 20:359–363.
- Leoni JB, Acuto FA. 2008. Social landscapes in pre-Inca Northwestern Argentina. In: Silverman H, Isbell WH, editors. *Handbook of South American archaeology*. New York: Springer. p 587–603.
- Lewis CM, Tito RY, Lizarraga B, Stone AC. 2004. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol* 127:351–360.
- Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. 2000. Genetic structure of Quechua-Speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* 118:5–17.
- Marrero AR, Bravi C, Stuart S, Long JC, Pereira das Neves Leite F, Peres LM, Kommers TC, Carvalho BA, Cruz IM, Pena SD, Ruiz-Linares A, Salzano FM, Bortolini MC. 2005. Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from Southern Brazil. *Hum Hered* 64:160–171.
- Merriwether DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native American suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98:411–430.
- Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal, mtDNA, and Y-chromosome Diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am J Hum Genet* 67:1277–1286.
- Moeder KP, Schurr TG, Bamforth FJ, Bazaliiski VI, Savel'ev NA. 2006. Population affinities of Neolithic Siberians: a snapshot from Prehistoric lake Baikal. *Am J Phys Anthropol* 129:349–361.
- Moraga M, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvalho P. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the Southern cone of the continent. *Am J Phys Anthropol* 113:19–29.
- Moraga M, Santoro CM, Standen VG, Carvalho P y Rothhammer F. 2005. Microevolution in Prehistoric Andean Populations: chronologic mtDNA variation in the Desert Valleys of Northern Chile. *Am J Phys Anthropol* 127:170–181.
- Oota H, Settheetham-Ishida M, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y chromosome variation is

- correlated with matrilineal and patrilineal residence. *Nat Genet* 29:20–21.
- Paez-Riberos LA, Munoz-Valle JF, Figuera LE, Nuno-Arana I, Sandoval-Ramirez L, Gonzalez-Martin A, Ibarra B, Rangel-Villalobos H. 2006. Y-linked haplotypes in Amerindian chromosomes from Mexican populations: genetic evidence to the dual origin of the Huichol tribe. *Legal Med* 8:220–225.
- Paredes M, Galindo A, Bernal M, Avila S, Andrade D, Vergara C, Rincon M, Romero RE, Navarrete M, Cardenas M, Ortega J, Suarez D, Cifuentes A, Salas A, Carracedo A. 2003. Analysis of the CODIS autosomal STR loci in four main Colombian regions. *Forensic Sci Int* 137:67–73.
- Ricaud FX, Keyser-Tracqui C, Bourgeois J, Crubézy E, Ludes B. 2004. Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. *Hum Biol* 76:109–125.
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241–260.
- Rodriguez-Delfin LA, Rubin-de-Celis VE, Zago MA. 2001. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. *Hum Hered* 51:97–106.
- Rothhammer F, Moraga M, Rivera M, Santoro CM, Standen VG, Garcia F y Carvallo P. 2003. Análisis de ADNmt de restos esqueléticos del sitio arqueológico de Tiwanaku y su relación con el origen de sus constructores. *Chungara* 35:269–274.
- Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* 74:223–263.
- Sans M, Merriwether DA, Hidalgo PC, Bentancor N, Weimer TA, Franco MH, Alvarez I, Kemp BM, Salzano FM. 2006. Population structure and admixture in Cerro Largo, Uruguay, based on blood markers and mitochondrial DNA polymorphisms. *Am J Hum Biol* 18:513–524.
- Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. *Ann Hum Biol* 31:87–94.
- Schneider S, Roessli D, Excoffier L. 2000. ARLEQUIN version 2000: a software for population genetic data analysis. Geneva: Genetics and Biometry Laboratory, University of Geneva.
- Shinoda K, Adachi N, Guillen S, Shimada I. 2006. Mitochondrial DNA analysis of ancient Peruvian highlanders. *Am J Phys Anthropol* 131:98–107.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, Stefano de GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SDJ, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68:1485–1496.
- Tarrago MN. 2000. Las sociedades del Sudeste andino. In: Rojas Rabiela T, Murra JV, editors. *Historia general de América Latina I: Las sociedades originarias*. Paris: Ediciones UNESCO. p 465–480.
- Tirado M, Lopez-Parra AM, Baeza C, Bert F, Corella A, Perez-Perez A, Turbon D, Arroyo-Pardo E. 2008. Y-chromosome haplotypes defined by 17 STRs included in AmpFISTR Yfiler PCR amplification kit in a multi ethnical population from El Beni department (North Bolivia). *Leg Med* 11:101–103.
- Toscanini U, Gusmao L, Berardi G, Amorim A, Carracedo A, Salas A, Raimondi E. 2008. Y chromosome microsatellite genetic variation in two Native American populations from Argentina. *Forensic Sci Int Genet* 2:274–280.
- Tovar F, Chiurillo MA, Borjas L, Ilander N, Ramirez JL. 2006. Chromosome Y haplotypes database in a Venezuelan population. *Int Congr Ser* 1288:246–248.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genetic* 4: e1000037.

Application of the iPLEX(TM) Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: Benefits for ancient population studies.

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

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

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

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

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

³INAPL, CONICET, Ciudad de Buenos Aires, Argentina

Received September 20, 2010

Revised November 16, 2010

Accepted November 17, 2010

Research Article

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

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

Keywords:

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



1 Introduction

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

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

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

Correspondence: Fanny Mendisco, Institut de Médecine Légale, Laboratoire d'Anthropologie Moléculaire EA 4438, 11 rue Humann, 67085 Strasbourg Cedex, France

E-mail: fanny.mendisco@neuf.fr

Fax: +33-3-68-85-33-62

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

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

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

2 Materials and methods

2.1 Samples

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

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

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

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

2.2 DNA extraction and quantification

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

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

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

2.3 Standard STR and mtDNA typing

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

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

2.4 SNP selection

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

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

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

2.5 iPLEX[®] Gold SNP genotyping method

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

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

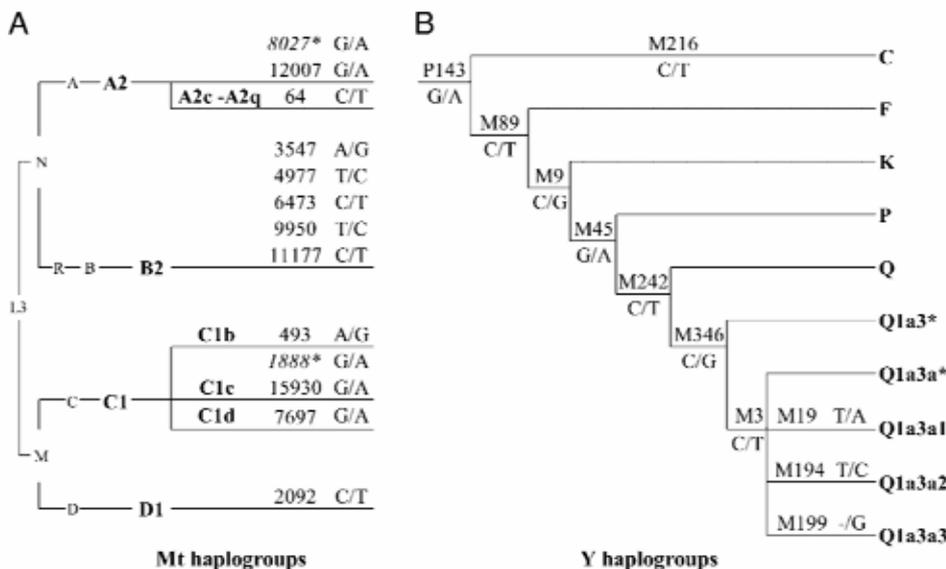


Figure 1. (A) From left to right: the characterized mitochondrial haplogroups (in bold), the positions of the selected markers on the mitochondrial DNA and the polymorphisms corresponding to these positions. The two markers in italics were rejected during the design by the software assay design. (B) Phylogenetic tree representing the 11 Y-chromosome markers selected, and the corresponding polymorphisms. The Y haplogroups characterized are indicated in bold.

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

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

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

2.6 Primer design

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

3 Results

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

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

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

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

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

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

3.2 Validation of the SNP genotyping assays

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

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

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

3.3 SNP genotyping of ancient samples

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

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

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

3.4 Genetic variation observed in the Los Amarillos ancient samples

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

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

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

4 Discussion

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

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

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

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

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

	PLEX_2																							Haplogroups	
	DF_P143	K_M9	P_M45	D_M242	D143_M3	D143_M346	D143_M3	D143_M19	D143_M194	D143_M199	C1e_19330	C1d_7997	A2_12007	A2_64	B2_3547	B2_4877	B2_6473	B2_9160	B2_1177	C1b_483	D1_2082	F_M189	C_M316		
	PLEX_1																						mt	Y	
Ancient male samples																									
LA3	A	G	A	T	G	T	T	T	T	T	G	G	G	C	A	T	C	T	C	A	A	T	T	T	D1
LA4	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA7	A	G	A	T	G	T	T	T	T	T	G	G	A	C	A	T	C	T	C	A	A	C	T	T	A2e-A2b
LA9	A	G	A	T	G	T	T	T	T	T	G	G	C	A	T	C	T	C	T	A	A	T	T	T	D1
LA10	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA12	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA16	A	G	A	T	G	T	T	T	T	T	G	G	C	A	T	C	T	C	T	A	A	T	T	T	D1
LA17	A	G	A	T	G	T	T	T	T	T	G	G	C	A	T	C	T	C	T	A	A	C	T	T	B2
LA20	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA21	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA22	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
LA24	A	G	A	T	G	T	T	T	T	T	G	G	A	T	A	T	C	T	C	A	A	C	T	T	A2
Ancient female samples																									
LA1																									A2
LA8																									A2
LA13																									A2
LA14																									A2
LA18																									A2
LA23																									C1b
Ancient samples of unknown sex																									
LA2	A																								A2
LA5	A																								/
LA6	A																								/
LA11	A																								/
LA19	A																								/

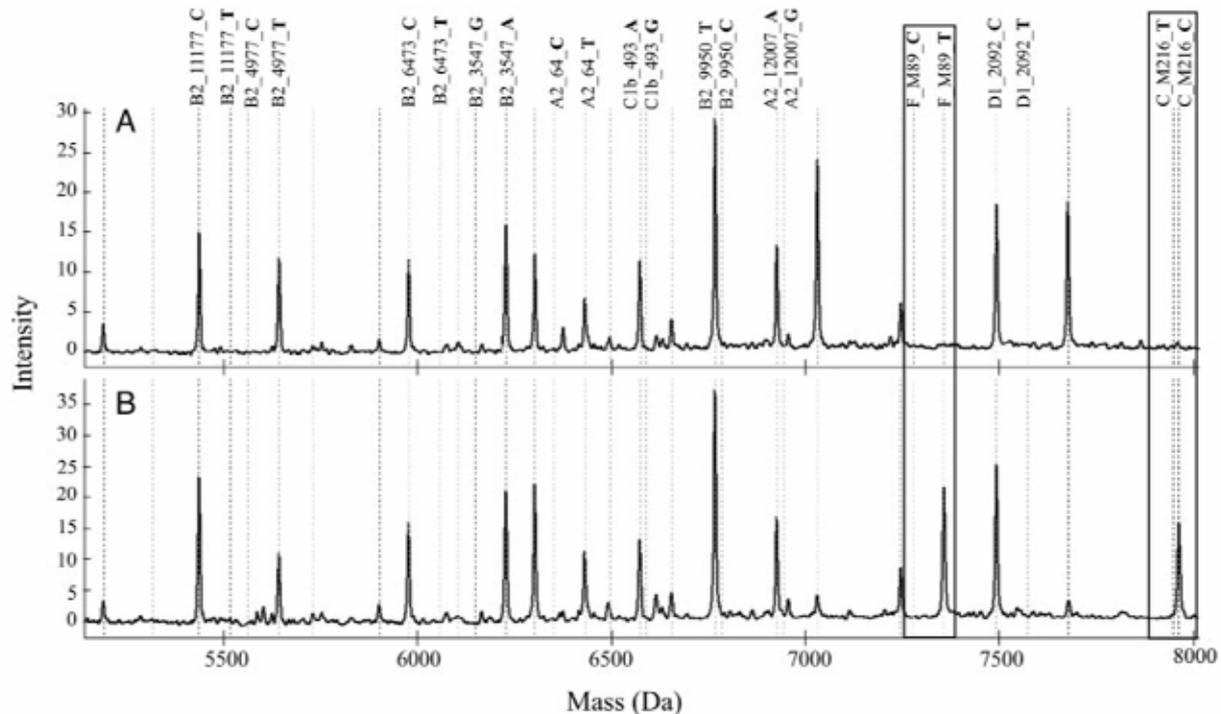


Figure 2. Example of the spectra obtained for the ancient female sample LA2 (A), and the ancient male sample LA4 (B), showing the peaks corresponding to the alleles of the “plex 2.” Both bordered markers are Y-chromosome SNP.

4.2 Benefits for ancient population studies

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

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

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

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

The authors have declared no conflict of interest.

5 References

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

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

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

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

Forensic Science International: Genetics Supplement Series (sous presse).

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

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

^a Laboratoire d'Anthropologie Moléculaire, Institut de Médecine Légale, Université de Strasbourg, CNRS UMR 5288, Strasbourg, France

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

Abstract.

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

Keywords: ancient DNA; insertion/deletion polymorphism; kinship; multiplex PCR

1. Introduction

Nowadays, Short Tandem Repeats (STRs) typing is considered as the method of choice in forensic science. The benefits of STRs are multiples and well-known. However, the analysis of STRs from highly degraded DNA sometime failed to resolve informative profile due to the relatively large size of amplicons. In order to solve this problem, miniSTR kit (Applied Biosystems) has been developed to reduce amplicon size. Nevertheless, the power of discrimination is low (10^{-11}) compared to STR typing kits (10^{-18} at least). As an alternative to STRs, SNPs have been successfully employed for highly degraded DNA due to the small size of amplicons (less than 150 bp), although their use in forensic kinship investigation is discussed (Amorim and Pereira, 2005). Another form of genetic variation on the human genome has recently received an increasing focus: insertion/deletion polymorphisms (indels). Their characteristics seem particularly interesting in forensic analysis: (i) their wide distribution throughout the genome (Weber *et al.*, 2002); (ii) the small size of amplicons which facilitates multiplex amplifications in highly degraded DNA; (iii) the possibility of typing them with a simple dye labeling electrophoretic approach.

In the present study, indels were analyzed through the Investigator DIPplex indels typing kit (Qiagen), in order to evaluate the performance of the kit for the analysis of ancient DNA samples and the informativeness of this kit for kinship investigation in anthropological fields where the trio mother-father-son is not always available. First tested on modern samples, the assay was secondarily applied on ancient specimens selected due to their putative kinship deduced from classical markers previously analyzed such as autosomal and Y-chromosomal STRs and mitochondrial DNA.

2. Materials and methods

Four modern DNA samples (buccal swabs), for which an appropriate informed consent was obtained, and nine ancient samples (teeth) from different age and location were analyzed (Table 1). Previous genetic analyses (unpublished observation) to estimate close parental relationship were carried as published in Lacan *et al.* (Lacan *et al.*, 2011).

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

3. Results and discussion

The results obtained on modern DNA samples are successful since the 30 indels loci included in this kit were amplified. Nonetheless, the analysis of the DNA profiles revealed that the intensity of the peak differs depending on the dye used. Indeed, the intensity was always the lowest with the BTR dye (Fig. 1). Moreover, these samples showed an imbalanced allele for heterozygous genotype which could be a significant limit for the use of this kit in DNA mixture analysis.

Regarding ancient DAN samples, a clear correlation between DNA quantity in the samples and the number of amplified loci was observed. The sensitivity of this assay allows obtaining workable profiles from degraded samples, although, as described by the manufacturer, the profiles are often partial and the results inconsistent, with less than 100 pg of DNA. In addition, the loss of signal

* Corresponding author at: 11 rue Humann, 67085 Strasbourg Cedex, France.

intensity with the high molecular weight markers from standard STRs typing kit is reduced due to the small size of amplicons (< 150 bp).

Then, the objective was to study if the results of indels typing obtained from ancient DNA samples confirmed the putative kinships deduced from previous autosomal/Y-STRs typing and mtDNA sequencing. In three cases (I1 and I2, O1 and O2, L9 and L20), the two individuals under study seemed linked by a parent/child relationship (Table 1). Indels typing gave confirmation to the hypothesis in two cases (O1/O2 and L9/L20) since the two linked individuals shared a common allele at each indels locus typed. The hypothesis is challenged in one case (I1 and I2). Indeed, the two individuals did not share any allele for one indels locus. This result allows excluding a parent/child relationship, reinforcing the hypothesis given by the statistical analysis conducted with the DNA view software (Brenner, 1997) ($LR=104 \times 10^{12}$). For these individuals, thanks to the analysis of the different types of markers used, the brother/sister link appears as the privileged hypothesis.

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

In conclusion, although this study as just focused on a few number of archaeological samples, it provides a first idea of the value of the investigator DIPplex for the analysis of ancient DNA samples. Indels typing was successful and highly informative for most of ours samples. This kit in combination with standard markers improves the discriminatory power and provides potential supplementation for anthropological studies.

7. Conflict of interest

None.

8. References

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

Table 1: Data on the ancient specimens tested.

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

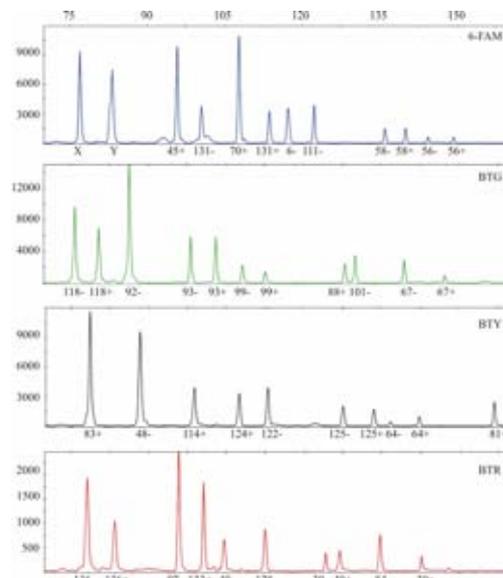


Fig. 1. Electropherogram obtained with the Investigator DIPplex kit using 332 pg DNA sample from an ancient specimen.