Zeitschrift: Archives des sciences [2004-ff.]
Herausgeber: Société de Physique et d'histoire Naturelle de Genève
Band: 61 (2008)
Heft: 1

Artikel: Evidences génétiques du peuplement de l'Asie orientale
Autor: Sanches-Mazas, Alicia
DOI: https://doi.org/10.5169/seals-738302

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 03.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Evidences génétiques du peuplement de l'Asie orientale

Alicia SANCHEZ-MAZAS[1]

Ms. reçu le 6 janvier 2008, accepté le 20 mars 2008

Abstract

Genetic evidence on the peopling of East Asia. – *This article is a current synthesis on the main genetic hypotheses related to human peopling history of East Asia, based on the analysis of different genetic systems and their comparison to archaeological and linguistic data. The general introduction (chapter 1) explains how genetics uses different approaches to analyze human diversity and interpret the results in relation to peopling history. Chapter 2 gives an overview of the results related to the first human settlements of this continent, with an explanation on how our ideas evolved since the first population genetics analyses until today. Chapter 3 describes the principal theories on more recent human migrations (like the Neolithic period) founded on the confrontation of genetic data and linguistic information.*

Keywords: *population genetics, linguistics, archaeology, prehistory, human peopling history, human evolution*

Résumé

Cet article est une synthèse actuelle des principales hypothèses génétiques sur l'histoire du peuplement humain en Asie orientale, basées sur l'analyse de plusieurs systèmes génétiques et leur confrontation aux données de l'archéologie préhistorique et de la linguistique historique. L'introduction générale (chapitre 1) explique les différentes approches par lesquelles la génétique est capable d'analyser la diversité humaine et d'interpréter les résultats dans le contexte du peuplement humain. Le chapitre 2 décrit les résultats relatifs aux tout premiers peuplements préhistoriques de ce continent, en expliquant comment nos idées ont évolué depuis les premières analyses génétiques de populations jusqu'à nos jours. Le chapitre 3 expose les théories sur les migrations humaines plus récentes (comme la période néolithique) émanant de la confrontation de la génétique avec des hypothèses linguistiques.

Mots clefs: *génétique des populations, linguistique, archéologie, préhistoire, histoire du peuplement humain, évolution humaine.*

1. Introduction: contribution de la génétique à l'histoire du peuplement humain

Nos connaissances sur l'histoire du peuplement du monde par les humains modernes (*Homo sapiens*) se sont enrichies, depuis quelques décennies, grâce aux résultats obtenus dans le domaine de la génétique. Du fait que l'ADN, la molécule de la vie qui porte nos gènes, se transmet continuellement de parents à enfants au cours des générations successives depuis l'origine de notre espèce, tous les humains actuels détiennent au fond de leurs cellules une information biologique susceptible de raconter l'histoire de leurs ancêtres. A plus large échelle, les populations actuelles de tous les continents portent aussi les traces biologiques de leur passé: elles sont, génétiquement, plus ou moins distinctes les unes des autres, et cette diversité génétique est le résultat de différenciations qu'elles ont subies au cours de leur histoire et de leurs migrations géographiques planétaires depuis l'émergence de la toute première population humaine moderne. L'étude des variations génétiques aujourd'hui observées dans les populations est donc des plus utiles pour reconstituer l'histoire du peuplement des continents.

[1] *Laboratoire d'anthropologie, génétique et peuplements (AGP), Département d'anthropologie et d'écologie, Université de Genève, 12, rue Gustave-Revilliod, CH-1211 Genève 4* – alicia.sanchez-mazas@unige.ch – http://agp.unige.ch

1.1. Approche populationnelle

La génétique aborde en fait les reconstitutions de notre passé selon deux approches très différentes. La première approche, de type «populationnel», se base sur les proportions – ou *fréquences* – de diverses variantes de gènes observées dans les populations actuelles. De nombreux gènes, comme ceux qui codent pour les groupes sanguins ABO, présentent en effet des variations entre individus sans que cela n'affecte (sauf dans des cas particuliers) la survie ou les facultés de procréation. De tels gènes sont appelés des *polymorphismes génétiques*, et il est possible, an analysant le sang ou, plus directement, l'ADN d'un ensemble d'individus d'une population donnée, de déterminer les fréquences des diverses variantes de chaque polymorphisme dans la population en question. Des «distances génétiques» entre populations peuvent ensuite être calculées à partir de ces fréquences et interprétées à la lumière d'autres connaissances.

Cette approche populationnelle consiste un peu à reconstruire des généalogies de populations, l'histoire ayant conduit celles-ci à se différencier génétiquement les unes des autres à partir d'une population ancestrale commune. Les fréquences des variantes géniques varient en effet au cours des générations sous l'effet du hasard, de telle sorte que la distance génétique entre deux populations ayant divergé à partir de la même population initiale peut devenir très grande au bout d'un certain temps. Mais le temps n'est pas le seul facteur qui intervient dans ces différenciations; la démographie joue aussi un rôle important. Les changements de fréquences géniques sont, en effet, particulièrement rapides dans les petites populations, surtout si celles-ci ont peu d'échanges avec le monde extérieur. Le cas de milieux physiquement isolés, comme nous le verrons pour l'île de Taïwan, illustre bien ce phénomène de *dérive génétique*. Il conduit en général à ce qu'une population perde de la diversité génétique, car les variantes géniques les plus rares ont plus de chances de disparaître. Mais l'isolement d'une population n'est pas forcément dû à une barrière géographique comme un bras de mer (cas de Taïwan), une zone désertique ou une montagne. Il peut aussi avoir une cause culturelle: des différences de comportement socio-culturel entre populations peuvent représenter des obstacles aux contacts et aux unions, et, par là-même, aux échanges de gènes. Pour mieux comprendre l'histoire des populations, il est alors intéressant d'étudier en parallèle leurs degrés de ressemblance ou de différence aux niveaux génétique et culturel. Or, un critère très utilisé aujourd'hui pour estimer l'éloignement culturel plus ou moins important entre populations est leur apparentement, plus ou moins grand, au niveau linguistique. Si les langues parlées par deux populations appartiennent à la même famille (comme la famille sino-tibétaine, à laquelle appartient le chinois), l'histoire des différenciations de cette famille et l'étude de ses relations avec d'autres familles linguistiques peuvent servir à reconstituer l'histoire des différenciations des populations concernées. Aujourd'hui, généticiens et linguistes s'associent ainsi fréquemment pour tenter d'évaluer l'impact d'éventuelles barrières culturelles sur les patrimoines génétiques des populations, ou, à l'inverse, de tester diverses hypothèses alternatives de reconstruction des familles de langues à l'aide des résultats de la génétique.

Opposé à l'idée de barrière, un autre mécanisme évolutif qu'il convient d'invoquer est le *flux génique*, ou échanges génétiques entre populations, conduisant, cette fois, à ce que des populations génétiquement différentes se ressemblent davantage. Plusieurs populations géographiquement voisines, si les échanges génétiques se font librement entre elles, seront d'autant plus semblables génétiquement qu'elles seront plus proches géographiquement, et d'autant plus différentes qu'elles seront plus éloignées. Le mécanisme en jeu est appelé *isolement par la distance géographique*, ou, plus simplement, *isolement par la distance*. Il est fréquent, en outre, et nous le verrons dans le cas de l'Asie orientale, d'observer des variations génétiques progressives (ou *continues*) le long d'un axe géographique uniquement (par exemple selon la latitude). Cela sera généralement interprété par des échanges (*inter-migrations*) ou des déplacements (*expansions géographiques*) de populations le long de cet axe.

Les phénomènes évolutifs mentionnés ci-dessus ont été étudiés sur le plan théorique dès la première moitié du XXe siècle par des généticiens de populations tels que Sewall Wright, aux Etats-Unis, et Gustave Malécot, en France. Ces développements ont trouvé des applications à l'histoire du peuplement humain dès les années 1960-1970 grâce à l'étude des systèmes génétiques «classiques» comme les groupes sanguins ABO, Rhésus et MNSs (parmi d'autres), et les enzymes et protéines, pour lesquels de très nombreuses données se sont accumulées pendant toute la seconde moitié du XXe siècle.

Parmi ces systèmes «classiques», nous parlerons abondamment, dans ce chapitre, des systèmes GM et HLA qui se sont révélés particulièrement informatifs pour reconstruire l'histoire du peuplement humain. Le système GM consiste en des variations moléculaires situées au niveau des molécules d'anticorps circulant dans le sang des individus, et résultant de variations, au niveau de l'ADN, que l'on appelle les «haplotypes GM». Au niveau des populations, les fréquences de ces haplotypes sont variables et semblent

en effet refléter l'histoire de leurs différenciations passées. Le système HLA est un autre système de type «immunitaire», comme GM, mais ses variations sont situées, cette fois, au niveau de molécules attachées à presque toutes les cellules de l'organisme, et qui ont pour rôle de reconnaître des virus ou autres pathogènes que l'organisme souhaite éliminer. La plus grande part des variations de fréquences des variantes génétiques HLA observées entre populations est due, là aussi, à l'histoire des différenciations humaines.

1.2. Approche phylogéographique

La deuxième approche est populairement connue sous le nom de «phylogéographie», ou géographie des lignages. Elle a un rapport étroit avec une approche de type généalogique (ou *phylogénétique*), mais appliquée aux gènes. Son principe consiste à reconstruire l'arbre généalogique le plus probable de toutes les variantes d'un gène (ou fragment d'ADN) observées chez des individus. L'âge de cette généalogie – c'est-à-dire le temps nécessaire pour que toutes les variantes observées aient pu dériver par mutation à partir d'un gène ancestral commun – peut alors être estimé, moyennant des hypothèses préliminaires sur la vitesse d'évolution de ces molécules. En outre, en tenant compte de l'abondance de chaque *lignage* (ou ensemble des variantes moléculaires généalogiquement rattachées) dans les populations humaines actuelles, il est possible de proposer des lieux d'origine de chaque branche de la généalogie et ses éventuels «parcours» géographiques dans d'autres régions du monde; par exemple, si quelques variantes d'un lignage *A* sont observées en Afrique alors que le lignage *A* dans son ensemble est particulièrement fréquent en Asie, l'on peut supposer que l'origine de ce lignage se situe en Asie et que les variantes trouvées en Afrique en ont dérivé. Combiné à une estimation de l'âge du lignage en question, ce «parcours» moléculaire d'Asie en Afrique est alors également daté, et éventuellement attribué à une migration de population particulière; ainsi, si l'on déduit, des analyses moléculaires, qu'un lignage *B* a eu sa source en Anatolie aux alentours de 10000 ans et s'est propagé en Europe, l'on sera tenté d'associer cet événement aux migrations des premiers agriculteurs venus du Proche-Orient tel que l'atteste l'archéologie.

Cette approche est intéressante mais elle a, bien entendu, ses défauts et ses limites: les âges des lignages sont souvent estimés avec de grandes erreurs, et l'attribution d'une origine géographique à un lignage donné est souvent très hypothétique. De plus, la construction d'arbres phylogénétiques ne s'applique pas aussi facilement à n'importe quel fragment d'ADN étudié. Il existe en effet diverses catégories de «marqueurs» génétiques, selon leur type de transmission de parents à enfants: les marqueurs à transmission *monoparentale*, et les marqueurs à transmission *biparentale*. L'ADN mitochondrial, petite molécule d'ADN présente dans les mitochondries des cellules (et non dans leur noyau comme la presque totalité de l'ADN), n'est transmis que des mères aux enfants, les mitochondries paternelles ne pénétrant pas l'ovule au moment de la fécondation. Le chromosome Y, quant à lui, responsable du sexe génétique masculin, n'est transmis que de père en fils. Tout gène ou fragment d'ADN présent sur l'ADN mitochondrial ou le chromosome Y n'étant transmis, d'une génération à l'autre, que par l'un des deux parents, les généalogies de ces molécules sont assez facilement reconstituées. En revanche, les gènes présents dans le noyau de la cellule et non localisés sur le chromosome Y (comme, par exemple, les gènes de l'hémoglobine), sont transmis, eux, par les deux parents à la fois. Dans ce cas, les généalogies deviennent très compliquées car le matériel génétique des deux parents se recombine à chaque génération, de telle sorte que l'on est incapable de distinguer précisément ce qui provient du lignage paternel de ce qui provient du lignage maternel au cours des générations.

Cette approche «phylogéographique», appliquée par certains généticiens pour reconstruire l'histoire génétique de notre espèce, est donc surtout appliquée à l'analyse des marqueurs à transmission monoparentale. Mais, même dans ce cas, il faut rester prudent sur les interprétations. En effet, les arbres phylogénétiques que l'on étudie sont des généalogies de molécules d'ADN qui ne coïncident pas forcément avec des généalogies de populations. Une phylogénie basée sur l'ADN mitochondrial n'a d'ailleurs aucune raison de coïncider avec une phylogénie du chromosome Y, les deux types de molécules étant transmises de manière indépendante d'une génération à l'autre. Par conséquent, lorsque des relations sont établies, par certains chercheurs, entre des migrations de lignages moléculaires et des migrations de populations diffusant une certaine langue ou une certaine technologie, il s'agit, dans tous les cas, d'hypothèses très spéculatives.

2. Premiers peuplements de l'Asie orientale par les humains modernes

2.1. Contexte paléontologique

La question des premiers peuplements de l'Asie orientale par des humains modernes est à rattacher aux hypothèses d'émergence de notre espèce, *Homo sapiens*, que les vestiges paléontologiques, suite aux

nouvelles datations des fossiles d'Omo Kibbish, en Ethiopie, situent aujourd'hui aux alentours de195000 ans (McDougall et al. 2005). Presque tous les paléontologues, tout comme les généticiens, s'accordent aujourd'hui pour considérer un événement de spéciation unique des humains modernes à partir de types humains antérieurs, dont les représentants classiquement reconnus sont des *Homo erectus* africains anciens parfois appelés *Homo ergaster*. Hormis la date très ancienne retenue pour l'Ethiopie, les premières dates attestées, de l'ordre de 100000 ans, concernent des vestiges retrouvés à la fois en Afrique de l'est, en Afrique du sud et au Proche-Orient. Cette dernière région a notamment livré les restes spectaculaires de Skuhl et de Qafzeh, et figure ainsi parmi les toutes premières régions du monde à avoir accueilli des humains modernes.

En revanche, l'Asie orientale n'a livré que peu de témoignages fossiles des premiers peuplements d'*Homo sapiens*: en Inde, à Narmada, a été exhumé un fossile âgé d'environ 50000 ans, mais mal caractérisé du point de vue anatomique (Kennedy et al. 1991). La date très ancienne de 68000 ans attribuée au fossile *Homo sapiens* découvert à Liujiang, dans le Guangxi, en Chine (Shen et al. 2002), reste controversée du fait que la couche sédimentaire dans laquelle il aurait été retrouvé est incertaine. La grotte de Tianyuan, à Zhoukoudian, également en Chine, a livré des restes de 42000 à 39000 ans (Shang et al. 2007), mais dont la morphologie est très archaïque. Sur la base des documents paléontologiques, on est donc contraint, pour le moment, d'inférer les dates des tout premiers peuplements de l'Asie orientale continentale par les humains modernes à partir des vestiges retrouvés en Asie orientale insulaire, qui témoignent de manière fiable des premières occupations par *Homo sapiens* vers 45000-40000 ans, par exemple à Bornéo (Barker et al. 2007).

2.2. Théorie génétique du «first split»

Face à ces lacunes de connaissances dans le domaine de la paléontologie, les généticiens ont cherché à proposer des scénarios pour les premiers peuplements humains. Ainsi, sur la base d'analyses utilisant des distances génétiques entre populations, et prenant en compte plusieurs systèmes génétiques «classiques» (plus particulièrement plusieurs types de groupes sanguins), l'équipe de Luca Cavalli-Sforza, à Stanford, et la nôtre, à Genève, ont proposé, dans les années 1970-1980, que l'espèce humaine s'était initialement différenciée en deux groupes, un groupe «occidental» et un groupe «oriental», puis que chacun de ces deux groupes s'était ultérieurement scindé pour donner naissance aux populations ancestrales des Européens, Proche-Orientaux et Africains, d'une part (à partir du groupe occidental), et des Asiatiques, Océaniens et Amérindiens, d'autre part (à partir du groupe oriental). Selon cette thèse, l'origine géographique de l'humanité se situait quelque part entre ces deux grands groupes, c'est-à-dire dans une zone géographique comprenant le Proche-Orient, l'Inde et l'Afrique de l'est. Cela correspondait bien au fait que les populations de ces trois régions présentaient les plus hauts degrés de diversité génétique observés à l'échelle mondiale; ainsi, les profils génétiques des populations d'Afrique sub-saharienne, d'Europe et d'Asie orientale s'expliquaient facilement, par simple dérive génétique, à partir de ce «pool génétique» commun (Sanchez-Mazas et Langaney 1988).

Le patrimoine génétique de l'ensemble des populations d'Asie orientale apparaissait donc comme très similaire à celui des populations d'Océanie et d'Amérique, et un scénario de peuplement progressif d'ouest en est depuis l'Inde ou le Proche-Orient en était l'explication la plus simple. Ces résultats se sont néanmoins révélés contradictoires avec ceux émanant de l'étude d'autres systèmes génétiques «classiques» – des protéines et des enzymes – par l'équipe de Matasho Nei et collaborateurs (Nei et Roychoudhury 1993). Ces chercheurs trouvaient en effet que l'espèce humaine se structurait d'abord, génétiquement, en un groupe «africain» et un groupe «non-africain», puis que le groupe «non-africain» donnait naissance aussi bien aux Européens qu'aux populations d'Asie orientale, d'Amérique et d'Océanie. Dans ce scenario, l'Europe semblait ressembler davantage à l'Asie orientale qu'à l'Afrique.

Outre le fait que différents systèmes génétiques subissent des modes d'évolution distincts et ne donnent pas forcément les mêmes résultats, l'une des principales causes de cette apparente contradiction est de nature méthodologique: en effet, non seulement les populations comparées n'étaient pas les mêmes et ne représentaient même pas des régions géographiques semblables, mais les représentations des relations génétiques entre populations sous forme d'arbres dont les embranchements sont dichotomiques (méthode utilisée dans les travaux de Nei et collaborateurs) sont inappropriées, car elles ne peuvent pas refléter les liens génétiques complexes entre populations (contrairement à des espèces différentes, les populations d'une même espèce, comme *Homo sapiens*, peuvent continuer à échanger des gènes après leur différenciation). Il a donc fallu explorer d'autres approches pour se rendre compte que le schéma global des différenciations génétiques humaines ressemblait plus à un réseau superposable à la géographie qu'à un arbre, ce qui suggérait que l'humanité s'était différenciée selon un modèle d'isolement par la distance (dû à des différenciations géo-

graphiques avec des inter-migrations entre populations voisines) à l'échelle mondiale. Dans un tel réseau, l'origine géographique de la première population humaine est impossible à déterminer. Seule la paléontologie pouvait être prise comme référence pour supposer que l'origine se trouvait vers l'Afrique de l'est ou le Proche-Orient. Fait intéressant, ces régions étaient celles dans lesquelles les populations présentaient le degré le plus élevé de diversité génétique, ce qui s'accordait très bien avec l'idée qu'elles renfermaient le pool génétique le plus ancien de l'humanité.

2.3. Théorie génétique «out-of-Africa»

Comme nous l'avons dit plus haut, l'analyse directe de l'ADN des individus a conduit ensuite à aborder l'étude de la diversité génétique humaine selon une approche très différente, la construction de phylogénies moléculaires. Dans la fin des années 1980, la publication du premier travail sur l'ADN mitochondrial, basé sur la comparaison de fragments d'ADN de longueurs variables selon les individus, révélait une phylogénie montrant une séparation initiale de types d'ADN observés uniquement chez des Africains (Cann et al. 1987). A partir de cette étude, l'idée que la génétique avait démontré l'origine africaine des humains modernes s'est imposée. L'âge d'environ 200000 ans calculé pour cette phylogénie semblait d'ailleurs confirmer les résultats de la paléontologie pour les premiers humains modernes. Pourtant, nombreuses ont été les critiques vis-à-vis de cette analyse qui s'est même avérée inadéquate sur le plan méthodologique. Au niveau des interprétations, les auteurs ont pris un raccourci scientifiquement incorrect: rappelons que la généalogie des lignages moléculaires ne reflète pas la généalogie des populations, et que l'âge des molécules n'a aucune raison de correspondre à l'âge des différenciations humaines. D'autres études du même genre ont néanmoins suivi, basées, non plus sur des fragments de longueur variable, mais sur les séquences partielles ou complètes de l'ADN mitochondrial ainsi que sur des variantes du chromosome Y. Si tous ces travaux ont confirmé l'ancienneté de quelques lignages moléculaires observés seulement chez certains individus africains, ils restent erronés dans l'affirmation que ces résultats apportent les évidences génétiques de l'origine africaine d'*Homo sapiens*.

Bien que les phylogénies moléculaires ne démontrent pas l'origine africaine de l'homme moderne, et soient inaptes à la dater, d'autres arguments permettent en revanche de supposer que l'Afrique de l'Est a joué un rôle important dans la diversification du pool génétique humain. En effet, le taux de diversité génétique actuel y est généralement supérieur à celui que l'on mesure dans d'autres régions du monde. Ceci peut s'expliquer soit par une plus grande ancienneté des populations, soit par un plus grand nombre d'individus ayant accumulé cette diversité. De plus, la génétique des populations permet de dire si les populations ont subi une (ou plusieurs) expansion(s) démographique(s) par le passé – et, le cas échéant, d'en estimer la période – ou si elles sont restées démographiquement stationnaires. Or, des analyses ont conclu que les populations est-africaines ont subi les plus anciennes expansions démographiques de la planète aux alentours de 100000 ans (Excoffier et Schneider 1999).

2.4. Théorie génétique des routes nord et /ou sud empruntées vers l'Asie

L'hypothèse «out-of-Africa» étant finalement prise, malgré toutes les mises en garde, comme modèle de base par la plupart des généticiens, la multiplication des données sur l'ADN mitochondrial et le chromosome Y a conduit, ces dernières années, à discuter de la route empruntée par les premières populations sorties d'Afrique pour atteindre l'autre extrémité de l'Ancien Monde, soit l'Australie, en passant par l'Asie orientale du sud. Deux scénarios ont été envisagés:

- une migration depuis l'Afrique de l'est vers le Sinaï le long de la Vallée du Nil et le passage du Sinaï au Levant, suivi de migrations vers l'Inde puis l'Asie du sud-est et l'Australie («route nord»);
- une migration depuis l'Afrique de l'est directement vers l'Arabie par le Détroit de Bab-el-Mandeb, puis le long des côtes de l'Océan Indien vers l'Asie du sud-est et l'Australie («route sud» ou «côtière»).

Si l'on s'en tient aux données de la paléontologie, la route nord est favorisée du fait que le Proche-Orient a clairement livré des fossiles d'humains modernes âgés de 10000 ans, soit les plus anciens après ceux d'Afrique de l'est, alors qu'aucun fossile ancien n'a été retrouvé sur la «route sud». Ce scénario est également compatible avec les résultats des approches populationnelles, comme décrit plus haut. Néanmoins, certains généticiens (Macaulay et al. 2005; Thangaraj et al. 2005) ont favorisé la route sud. En effet, des lignages moléculaires très anciens et uniques de l'ADN mitochondrial (les lignages M31 et M32) observés dans les îles Andaman situées dans le Golfe de Bengale, seraient directement issus du lignage «fondateur» M prédominant en Eurasie et dont l'âge est estimé à 63000 ans (M serait lui-même dérivé de l'haplogroupe L3 originaire d'Afrique de l'est, dont l'âge est estimé à 84000 ans). Dans ce scénario, le peuplement des îles Andaman daterait

d'au moins 45000 ans, ce qui s'expliquerait par une migration rapide d'ouest en est le long de la route sud. Du point de vue du polymorphisme du chromosome Y, les lignages moléculaires observés en Asie orientale proviennent d'un ancêtre commun unique M168, ultérieurement subdivisé en 3 lignages YAP, M130 et M89 (Underhill 2005). YAP (dérivé ensuite en M174 en Asie orientale) aurait eu son origine en Afrique et serait représentatif des premiers arrivants en Asie orientale par la «route sud». En effet, selon les auteurs, le fait de le trouver à hautes fréquences aux îles Andaman (voir aussi Chandrasekar (2007)) ainsi qu'au Tibet et au Japon indiquerait qu'il aurait été «poussé» dans les régions périphériques sous l'arrivée plus récente de migrants portant d'autres lignages. Le lignage M130 est supposé d'origine asiatique très ancienne et présenterait aussi une distribution compatible avec une route côtière.

Cependant, l'interprétation des résultats de l'ADN mitochondrial en faveur d'une «route sud» a récemment été contestée. L'analyse moléculaire de l'ADN mitochondrial de nombreuses populations indiennes a en effet révélé qu'elles partageaient une partie de ces lignages avec les Andamanais et que l'origine de ces lignages daterait, tout au plus, de 24000 ans. L'âge des variants observés uniquement aux îles Andaman serait, quant à lui, inférieur à 12000 ans (Barik et al. 2008). Selon d'autres auteurs encore (Abu-Amero et al. 2008), la diversité de l'ADN mitochondrial dans la Péninsule arabique ne porte pas de traces de cette expansion (absence de lignages M et N autochtones). Enfin, l'archéologie ne livre aucun témoignage d'occupation humaine des îles Andaman avant le 3e millénaire avant notre ère.

Ces exemples nous montrent déjà (et nous le verrons plus loin sous un autre angle) qu'aucun scénario de peuplement initial de l'Asie orientale par les humains modernes n'est à ce jour définitivement démontré. Ils nous conduisent également à nous défaire de certains préjugés profondément ancrés dans les représentations collectives depuis près de deux siècles (et influençant parfois les interprétations de la génétique), à savoir que des populations de phénotype particulier – peau foncée, petite stature –, comme les Andamanais, les «Negritos» de Malaisie et des Philippines ou les Aborigènes Australiens, tous situés sur la «route sud», seraient automatiquement des représentantes de populations «primitives». Ces interprétations sont infondées sur le plan scientifique. Rappelons, ici, que les caractères phénotypiques évoluent très vite sous l'effet de pressions adaptatives, et que les convergences sont fréquentes entre populations géographiquement éloignées vivant dans des environnements relativement semblables, comme les forêts tropicales.

2.5. Origine et diffusion du pool génétique asiatique

Dans les scenarios décrits ci-dessus, il n'est question que de migrations anciennes des humains modernes de l'Afrique vers l'Asie du sud-est, l'Asie du nord-est n'étant pas mentionnée. Or, les chercheurs ont aussi beaucoup discuté des contributions génétiques possibles d'ouest en est, par l'Asie centrale, vers le nord-est.

Initialement, cette question a été abordée grâce à des résultats obtenus sous forme de projections bidimensionnelles de populations. Le problème des représentations en arbres de données populationnelles, que nous avons discuté plus haut, a en effet conduit la plupart des généticiens à choisir d'autres types de représentations des relations génétiques entre populations, notamment sous forme de projections construites d'après les distances génétiques calculées (analyses d'échelonnement multidimensionnel, analyses en composantes principales, etc). Or, dans de tels schémas appliqués aux systèmes génétiques classiques à l'échelle mondiale, la projection des populations rappelle presque systématiquement la localisation géographique de celles-ci dans les différents continents. Par exemple, la position génétique de l'Asie orientale coïncide bien avec sa position géographique intermédiaire entre le Proche-Orient, d'une part, et les continents océanien et amérindien, d'autre part. De plus, ces analyses ont révélé un autre résultat particulièrement intéressant concernant les populations est-asiatiques: celles-ci se différencient selon un axe nord-sud: les populations d'Asie du nord-est, comme les Sibériens, Japonais, Coréens et Chinois du nord, se ressemblent génétiquement entre elles. D'un autre côté, les Chinois du sud et les populations d'Asie du sud-est sont génétiquement très semblables. Soulignons cependant que ces analyses ne révèlent pas de séparation nette entre les populations du nord et du sud de l'Asie, mais montrent plutôt une continuité génétique entre populations selon la latitude.

Du point de vue des marqueurs «classiques», le peuplement de l'Asie orientale peut ainsi s'expliquer par des migrations depuis l'Asie de l'ouest de part et d'autre de la chaîne himalayenne, c'est-à-dire d'une part, par l'Asie centrale, et, d'autre part par le sud de l'Asie orientale (Sanchez-Mazas 1990; Cavalli-Sforza et al. 1994). L'Himalaya aurait donc constitué une barrière géographique au flux génique, suffisamment importante pour conduire à la différenciation des populations d'Asie orientale en deux groupes, nord et sud. Des inter-migrations auraient, par la suite, recréé une continuité génétique entre les populations du nord et du sud de l'Asie orientale. Ce scénario a été repris sous le nom de modèle de la «pince», et considère également un plus grand apport génétique vers l'Asie

du nord-est que vers l'Asie du sud-est depuis l'Asie centrale. En effet, les populations d'Asie du nord-est présentent un taux plus élevé de diversité génétique que les populations d'Asie du sud-est. Ainsi, le pool génétique asiatique aurait, non pas une seule, mais deux sources occidentales. Bien entendu, ces apports n'auraient pas forcément été simultanés.

Ce modèle d'origine multiple du pool génétique asiatique a également été défendu sur la base d'une étude du chromosome Y (Karafet et al. 2001) dont les résultats ont contredit les travaux précédents sur le même système, mais basés sur un nombre plus restreint de populations (Su et al. 1999). D'après Karafet et collaborateurs, il existe une relation génétique étroite entre les populations d'Asie centrale et celles d'Asie du nord-est. Dans ce sens, l'idée que l'Himalaya a constitué une barrière au flux génique entre le nord et le sud est soutenue par le fait que les populations du plateau tibétain n'ont pratiquement pas reçu de contribution génétique (tout au moins pour le chromosome Y) des populations indiennes situées un peu plus au sud, dans la plaine (Gayden et al. 2007). De plus, les populations du nord-est seraient plus diversifiées que celles du sud-est du point de vue moléculaire, et leurs profils génétiques ne dériveraient donc pas directement d'un pool génétique sud-est asiatique (Karafet et al. 2001). L'examen détaillé des haplogroupes n'exclut cependant pas que des migrations bidirectionnelles se soient produites en Asie orientale: par exemple, M175 pourrait avoir diffusé vers le nord à partir d'une origine méridionale, alors que des haplogroupes partagés entre les populations d'Asie centrale et d'Asie du Nord-Est auraient diffusé du nord au sud (Karafet et al. 2001). Le débat à ce sujet est resté polémique (Shi et al. 2005).

L'étude de l'ADN mitochondrial met également en évidence des différences génétiques entre les populations du nord et celles du sud de l'Asie orientale: au nord, les haplogroupes les plus fréquents sont surtout du lignage M (M8, M9, D et G), alors qu'au sud, la plupart des haplogroupes fréquents sont du lignage N (R, B, R9, F), sauf quelques exceptions (M7, du lignage M, est plus fréquent au sud, et A, du lignage N, plus fréquent au nord). L'examen plus approfondi de ces variations révèle néanmoins des résultats contradictoires. En effet, certains auteurs considèrent que les différenciations sont continues du nord au sud; en effet, des gradients de fréquences sont observés pour certains haplogroupes, comme F1, B et D4 (Yao et al. 2002). D'autres auteurs (sur la base de plusieurs systèmes génétiques indépendants) suggèrent un simple modèle d'isolement par la distance dans lequel les populations sont d'autant plus éloignées, génétiquement, qu'elles sont plus éloignées, géographiquement (Ding et al. 2000). Ces deux premiers modèles sont d'ailleurs compatibles l'un avec l'autre. Enfin, certains chercheurs soutiennent en revanche qu'il existe une ou plusieurs discontinuité(s), ou «barrière(s) génétique(s)», entre le nord et le sud (Xue et al. 2008). Mais, alors que pour l'ADN mitochondrial, la discontinuité génétique la plus importante se situerait le long de la rivière Huai dans les montagnes Qin, au nord du fleuve Yangtse, pour les systèmes «classiques», elle se situerait le long du Yangtse (Xue et al. 2008). De plus, aucune barrière significative n'a pu être mise en évidence pour les marqueurs du chromosome Y (Xue et al. 2008). Ces différences peuvent éventuellement être expliquées par des histoires démographiques distinctes entre les hommes (laissant une signature sur la diversité du chromosome Y) et les femmes (laissant une signature sur la diversité de l'ADN mitochondrial). Rappelons néanmoins que les résultats sont très dépendants de l'échantillonnage considéré et des méthodes d'analyse utilisées pour détecter des continuités ou discontinuités génétiques.

Il existe donc, globalement, une variation génétique marquée suivant la latitude en Asie orientale, avec des irrégularités dans certaines régions ou pour certaines populations particulières. Deux exemples illustrent bien ce dernier cas: les populations du Yunnan se distinguent par une fréquence élevée de l'haplogroupe mitochondrial M7b1 (près de 14%), et celles du Zhanjiang se différencient beaucoup de celles du Guangzhou malgré la proximité géographique de ces deux régions (Yao et al. 2002).

Cette variation géographique des systèmes moléculaires étant mise en évidence, les chercheurs se questionnent aussi, comme pour les systèmes classiques, sur l'origine géographique, au nord ou au sud, de l'ensemble des populations d'Asie orientale. L'âge de la plupart des lignages moléculaires de l'ADN mitochondrial observés en Asie orientale a été estimé à plus de 50000 ans, les plus anciens (R9 et B, de 81000 ans et 75000 ans, respectivement) étant prédominants dans le Sud (Yao et al. 2002). Bien que l'on ne puisse pas transposer directement ces âges à ceux des populations, l'hypothèse majoritaire reste que toutes les populations asiatiques ont leur origine dans le sud du fait que l'on y observerait aujourd'hui une plus grande diversité génétique que dans le nord, et que les profils génétiques des populations du nord dériveraient de ceux du sud. Mais ce résultat, rappelons-le, est opposé à celui que révèlent les marqueurs classiques et ceux du chromosome Y. Par conséquent, si l'on admet une origine première des populations est-asiatiques en Asie du sud, suivie de migrations du sud au nord (avec réduction de la diversité par dérive génétique), il faut alors obligatoirement considérer des apports génétiques ultérieurs importants au nord, par exemple depuis l'Asie centrale: pour pouvoir expliquer la diversité élevée observée

au niveau des lignages paternels du chromosome Y, et, dans une moindre mesure, des systèmes à transmission biparentale, ce flux génique aurait dû être majoritairement masculin.

En fait, les populations d'Asie centrale présentent un degré particulièrement élevé de diversité génétique pour le chromosome Y, qui peut être expliqueé par une origine très ancienne de populations (Wells et al. 2001). Mais, là aussi, plusieurs interprétations sont possibles: soit l'Asie centrale a constitué un réservoir ancien de variabilité à partir duquel des expansions se seraient produites à la fois vers l'est, l'ouest et le sud, comme le soutiennent ces auteurs, soit des populations de diverses origines ont migré conjointement dans cette région, créant un pool diversifié par la mise en commun de profils génétiques distincts (Comas et al. 1998; Comas et al. 2004). Plus récemment, un scénario intermédiaire a aussi été proposé sur la base de simulations par ordinateur de processus évolutifs appliqués aux marqueurs de l'ADN mitochondrial et du chromosome Y (Chaix et al. 2008): des migrations de populations auraient eu lieu d'est en ouest à l'époque paléolithique, d'abord de l'Asie orientale vers l'Asie centrale, puis de l'Asie centrale à l'Europe, ce qui expliquerait les dates d'expansions démographiques estimées, plus anciennes à l'est qu'à l'ouest. L'Asie centrale aurait donc été à la fois une région d'accueil et d'expansion de populations, ces migrations d'est en ouest étant éventuellement liées à des innovations technologiques, telles que la diffusion de la culture aurignacienne.

3. Différenciations génétiques des populations d'Asie orientale

3.1. Contexte linguistique

Au cours de la dernière décennie, de nombreuses collaborations ont vu le jour entre généticiens et linguistes pour tâcher de reconstruire l'histoire du peuplement humain. En effet, la linguistique historique propose des hypothèses d'apparentement entre les langues humaines en familles et super-familles (ou *phyla*) dont elle essaie de déduire l'histoire des différenciations en parallèle avec les expansions géographiques et/ou démographiques des populations. Les scénarios de peuplement énoncés par la génétique sont donc très utiles aux linguistes, et inversement. Comme de telles approches ont fait l'objet d'études récentes sur l'Asie, nous rappellerons d'abord, ici, brièvement, les classifications majeures des langues est-asiatiques.

L'Asie orientale compte une multitude de populations culturellement distinctes et dont les langues appartiennent à plusieurs grands *phyla* (voir *l'Ethnologue*, pour des classifications détaillées). Le phylum altaïque compte 66 langues réparties dans les familles mongole, toungouze et turque et occupant une aire géographique très vaste en Eurasie du nord, de la Turquie à la Sibérie orientale. Le japonais, le coréen et l'aïnou constituent trois familles ou isolats à part entière, parfois associés dans un même phylum, ou encore affiliés aux langues altaïques, selon les auteurs. Le sino-tibétain inclut quelques 365 langues dont le chinois (et ses dialectes), le tibétain et le birman. Son extension géographique inclut la Chine (y compris le Tibet), le Laos, la Birmanie, l'Inde, le Népal et le Bhutan. L'austro-asiatique compte 168 langues dont le khmer, le mon et le vietnamien, en Asie du sud-est, ainsi que le munda et le khasi, au nord-est de l'Inde. Le tai-kadai, ou daïque, avec ses 70 langues, est parlé en Chine, en Thaïlande, au Laos, en Birmanie et au Vietnam. Le hmong-mien, ou miao-yao, est un petit phylum de 32 langues constituant de petites enclaves linguistiques en Chine du sud ainsi qu'au Laos, en Thaïlande et au Vietnam. Enfin, l'austronésien est un phylum particulièrement important à la fois du point de vue du nombre de langues qu'il inclut (1262) que de son extension géographique, puisqu'il couvre l'ensemble du Pacifique depuis Taiwan (à l'exception de la Nouvelle-Guinée dans sa presque totalité, d'un certain nombre d'îles mélanésiennes, et de l'Australie), ainsi qu'une partie du Vietnam du sud, et Madagascar. Les hypothèses d'apparentement entre ces grandes familles sont encore ardemment débattues par les linguistes (voir Sagart et al. 2005), pour un résumé des diverses hypothèses sur les phyla du sud).

3.2. Gènes et langues en Asie orientale

Comme nous l'avons vu plus haut, l'analyse de divers systèmes génétiques à l'échelle mondiale et au niveau des différents continents a montré que les populations asiatiques présentaient un degré élevé de diversité génétique et qu'elles se structuraient, génétiquement, selon une composante géographique. Néanmoins, cette structure correspond aussi, dans une certaine mesure, à la structure linguistique des populations, comme l'indiquent les travaux réalisés sur les systèmes GM (Poloni et al. 2005), HLA (Sanchez-Mazas et al. 2005), et les marqueurs du chromosome Y (Karafet et al. 2001). Pour GM, par exemple, les populations de groupes linguistiques différents se distinguent globalement beaucoup plus entre elles que les populations prises à l'intérieur de chaque groupe.

Cependant, lorsque les groupes linguistiques sont comparés deux à deux de manière plus précise, ou qu'on les représente sur des projections bidimensionnelles (Figs. 1, 2 et 3), des relations complexes appa-

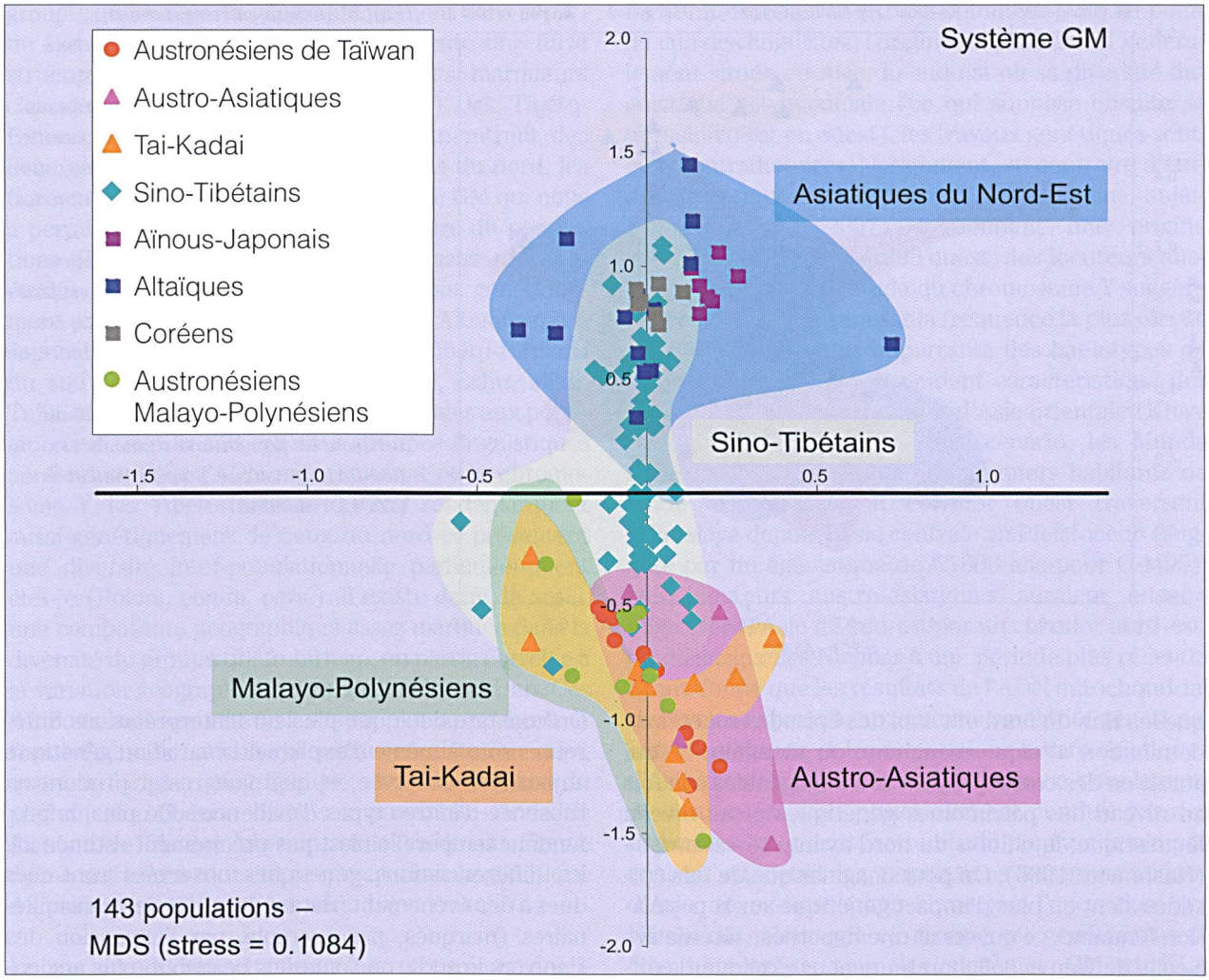


Fig. 1: Analyse d'échelonnement multidimensionnel («multi-dimensional scaling analysis», ou MDS) de 143 populations d'Asie orientale testées pour le système GM des immunoglobulines. Chaque point représente une population et son symbole la famille linguistique à laquelle la langue de cette population est affiliée. La disposition des points sur la projection reflète les distances génétiques entre populations. Les plages de couleur unissent les populations d'une même famille linguistique. La valeur de stress (0.1084) indique la fiabilité de la représentation (bonne ici). Source: résultats personnels et partiellement publiés (Poloni et al. 2005).

raissent: en effet, les groupes Tai-Kadai, Austro-asiatiques et Hmong-Mien du sud de l'Asie se superposent sur les projections, aussi bien pour le système GM (Fig. 1) que pour l'ADN mitochondrial (Fig. 2), tandis que les Han du nord se distinguent des Han du sud de manière significative (Wen et al. 2004; Poloni et al. 2005 et comm. pers.).

La situation est un peu différente pour les marqueurs du chromosome Y (Fig. 3). D'une part, les groupes Hmong-Mien, Austro-asiatiques et Tai-Kadai se distinguent génétiquement les uns des autres de façon plus marquée, et, d'autre part, les Han du nord et du sud forment ensemble un groupe génétiquement plus homogène et qui se distingue également mieux des trois premiers (Wen et al. 2004). Sur la base de ces résultats, Wen et collaborateurs soutiennent l'hypothèse d'une diffusion démique (autrement dit, avec déplacement de populations, contrairement à une diffusion culturelle), des Han du nord vers le sud, conduite principalement par des hommes, et entraînant une assimilation génétique de femmes des groupes locaux du sud, lors des deux derniers millénaires: ceci expliquerait que les Han du sud se confondent avec les groupes locaux pour l'ADN mitochondrial mais pas, ou peu, pour le chromosome Y.

Cependant, selon nous, il faut aussi tenir compte des relations existant, au nord, entre les Han et les populations altaïques: nous avons en effet montré que ces populations étaient étroitement liées sur le plan génétique, aussi bien pour les marqueurs classiques (Fig. 1 et Sanchez-Mazas et al. (2005)) que pour l'ADN mitochondrial (Poloni, comm. pers.). On sait

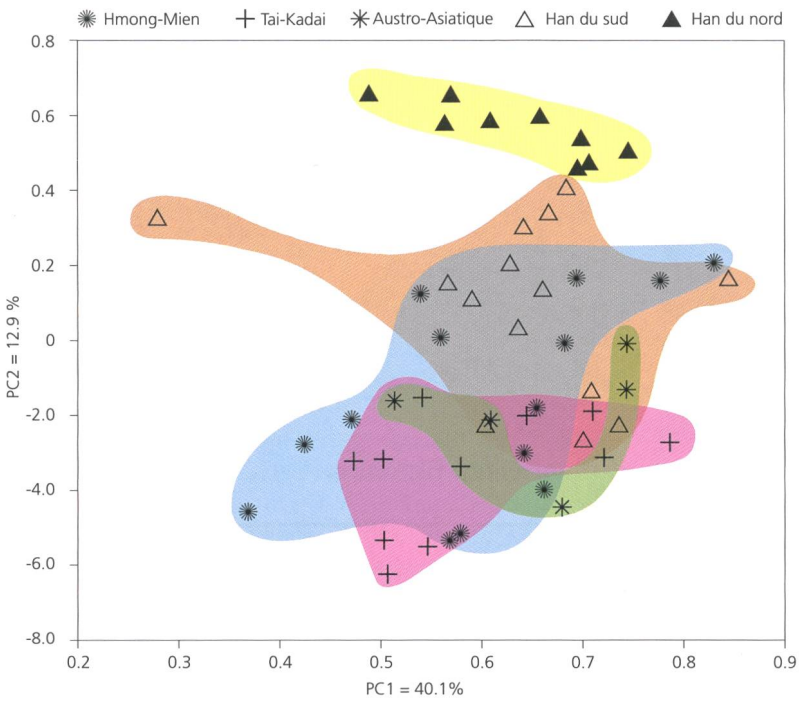


Fig. 2: Analyse en composantes principales (PCA) de 52 populations d'Asie orientale testées pour l'ADN mitochondrial. Chaque point représente une population et son symbole la famille linguistique à laquelle la langue de cette population est affiliée. La disposition des points sur la projection reflète les distances génétiques entre populations (mais ici, l'axe horizontal «PC1» est plus informatif que l'axe vertical «PC2», ces deux axes représentant 40.1% et 12.9% de la variabilité génétique totale, respectivement). Les plages de couleur unissent les populations d'une même famille linguistique. Source: Wen et al. (2004), redessiné.

que les Han du nord ont subi des épisodes répétés de domination altaïque manchoue ou mongole, ce qui aurait eu des conséquences non négligeables à la fois au niveau des patrimoines génétiques et au niveau linguistique, le chinois du nord ayant été «altaïcisé» (Hashimoto 1986). On peut imaginer que de tels épisodes aient eu plus d'impact génétique sur la population féminine, ce qui serait une hypothèse alternative à celle de Wen et collaborateurs pour expliquer la différenciation génétique plus importante, pour l'ADN mitochondrial que pour le chromosome Y, entre Han du nord et Han du sud. Mais cela reste bien sûr à démontrer.

On constate donc que plusieurs interprétations différentes sont à même d'expliquer la variation génétique aujourd'hui observée, et qu'il faut rester prudent en l'absence d'autres types d'évidences. De plus, la profondeur temporelle n'est pas précisément estimée ici: les différenciations génétiques observées sont-elles dues à des événements datant des deux derniers millénaires (marqués, par exemple, par l'expansion des Han vers le sud), ou sont-elles beaucoup plus anciennes? Rappelons que les langues chinoises sont apparentées aux langues tibéto-birmanes dans la famille sino-tibétaine. Or, la position génétique des populations tibéto-birmanes est particulière. En effet, ce

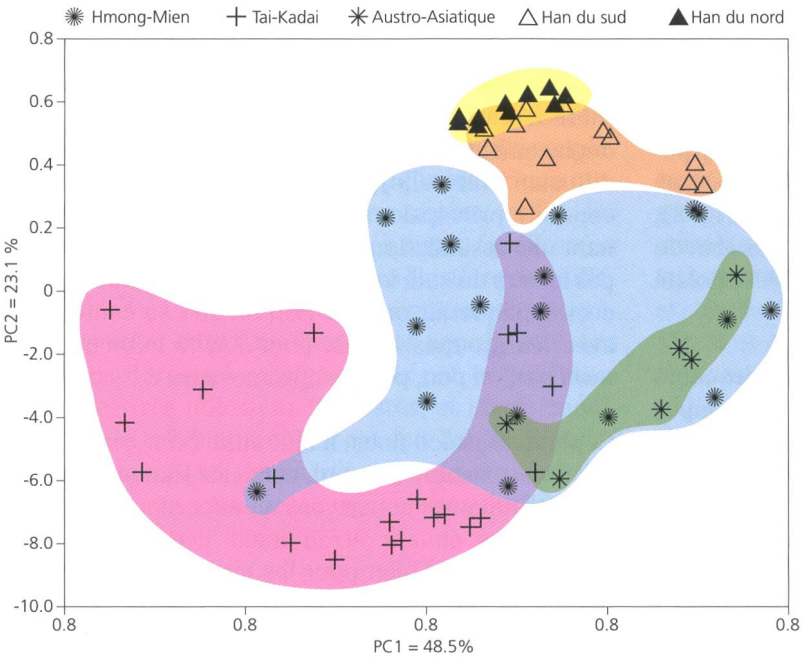


Fig. 3: Analyse en composantes principales (PCA) de 66 populations d'Asie orientale testées pour le chromosome Y. Chaque point représente une population et son symbole la famille linguistique à laquelle la langue de cette population est affiliée. La disposition des points sur la projection reflète les distances génétiques entre populations (mais ici, l'axe horizontal «PC1» est plus informatif que l'axe vertical «PC2», ces deux axes représentant 48.5% et 23.1% de la variabilité génétique totale, respectivement). Les plages de couleur unissent les populations d'une même famille linguistique. Source: Wen et al. (2004), redessiné.

groupe, qui se répartit géographiquement dans la partie sud-ouest de l'Asie orientale, présente une forte structure interne. Pour l'ensemble des marqueurs classiques (Cavalli-Sforza et al. 1994), les Tibéto-Birmans du nord (Tibet et Bhoutan) montrent des liens génétiques étroits avec les Chinois du nord, les Coréens et les Japonais. Pour le système GM qui nous a permis d'inclure un plus grand nombre de populations (Poloni et al. 2005, Sanchez-Mazas présente étude), les Tibétains et Bhoutanais sont génétiquement proches des Chinois du nord, des Altaïques, des Japonais et des Coréens, alors que les Tibéto-Birmans du sud (comme les Kachari, Sonowal, Lahu, Mikir, Tujia, et Yi) sont génétiquement semblables aux populations d'Asie du sud-est, tous groupes linguistiques confondus. Pour l'ADN mitochondrial et le chromosome Y, les Tibéto-Birmans du sud se démarquent aussi génétiquement de ceux du nord et présentent une diversité inter-populationnelle particulièrement élevée (Poloni, comm. pers.). Il existe donc, là aussi, une composante géographique assez marquée dans la diversité du groupe tibéto-birman, en partie corrélée à la variation géographique du pool génétique Han. Ces résultats mériteraient un examen plus approfondi. A l'heure actuelle, aucun consensus n'existe, en effet, quant à l'origine géographique et à l'ancienneté du phylum sino-tibétain dans son ensemble: l'Inde, le Sichuan, le Plateau tibétain, et la Vallée du Fleuve Jaune ont tour à tour été proposés, mais la génétique n'a pas encore d'hypothèse définitive à ce sujet.

Le cas du phylum altaïque n'est pas mieux documenté. Il est néanmoins intéressant de constater que, pour GM et HLA, les populations de langues «altaïque-propre» se différencient génétiquement de manière significative des Japonais et Coréens, en accord avec les théories linguistiques selon lesquelles ces groupes ne seraient pas rattachés phylogénétiquement entre eux, ou qui les considèrent comme très éloignés au sein d'un grand phylum macro-altaïque ou «eurasiatique». Considérées à part, les populations parlant des langues de la famille «altaïque-propre» incluant le turc, le mongol, le manchou et le toungouze sont, en fait, très différenciées entre elles pour le système GM (Fig. 1) et pour le chromosome Y (Xue et al. 2006; Karafet et al. 2008). Il semble que la dérive génétique ait joué un rôle majeur dans la différenciation de ces populations réparties, pour un certain nombre d'entre elles, de manière discontinue et avec de faibles densités dans les vastes régions de Sibérie (Karafet et al. 2008).

En revenant sur les groupes du sud de l'Asie orientale, l'origine des populations de langues austro-asiatiques est actuellement très discutée, étant donné la distribution particulière de ce groupe linguistique entre le nord-est de l'Inde (Munda), l'Asie du Sud-Est (Mon-Khmer), et plusieurs enclaves géographiquement intermédiaires (Khasi-Khmuique). Si, du point de vue des linguistes, l'origine du phylum est généralement située en Asie du sud-est où sa diversité linguistique est maximale (ce qui suppose ensuite sa diffusion d'est en ouest), les travaux génétiques sont, eux, contradictoires. Notamment, au contraire d'études génétiques précédentes sur le même sujet, Kumar et al. (2007) soutiennent une origine indienne, autrement dit à l'ouest, des locuteurs austro-asiatiques. Leur étude du chromosome Y suggère que les Munda présentent la fréquence la plus élevée et la diversité la plus importante des haplotypes de l'haplogroupe O-M95 également caractéristique des populations austro-asiatiques d'Asie orientale (Khasi et Mon-Khmers). Selon leur scenario, les Munda seraient les descendants des premiers habitants de l'Inde arrivés par un corridor ouest traversant l'Himalaya depuis l'Asie centrale au Pleistocène (suggéré par un âge estimé de 65000 ans pour O-M95). Des locuteurs austro-asiatiques auraient ensuite migré vers l'Asie du sud-est par un corridor nord-est, et, de là, aux îles Nicobar à une période plus récente. Etant donné que les résultats de l'ADN mitochondrial ne révèlent, quant à eux, aucune similarité génétique entre les Munda et les populations austro-asiatiques d'Asie du sud-est, il est suggéré que ce peuplement d'ouest en est était principalement conduit par des hommes. Cependant, cette étude ponctuelle n'a pas analysé en détail les populations austro-asiatiques d'Asie du sud-est (mon-khmer) à l'exception des Nicobarais, alors que ce groupe est génétiquement très hétérogène pour les systèmes GM et HLA (Poloni et al. 2005; Sanchez-Mazas et al. 2005), ainsi que pour l'ADN mitochondrial (Poloni, comm. pers.). De plus, les Munda sont génétiquement très éloignés des Mon-Khmer aussi bien pour le chromosome Y (Kumar et al. 2007) que pour l'ADN mitochondrial (Metspalu et al. 2004) et selon nos résultats sur la diversité génétique HLA-A, -B et -DRB1 (Riccio 2008). Si l'origine austro-asiatique de certaines populations d'Inde du nord-est, et plus particulièrement les Khasi situés plus à l'est que les Munda, dans la région du Meghalaya, a peut-être laissé une certaine signature génétique (Reddy et al. 2007), il est en tout cas prématuré d'affirmer une quelconque origine géographique de ce groupe sur une base génétique.

Les autres groupes d'Asie du sud-est – Hmong-Mien et Tai-Kadai – sont à la fois génétiquement proches entre eux et proches du groupe austro-asiatique, tous trois formant ainsi un ensemble de populations globalement apparentées. Cependant, l'accumulation récente des données génétiques a révélé une importante diversité génétique entre populations à l'intérieur de chacun de ces groupes: la variabilité inter-populationnelle des Tai-Kadai et des Hmong-Mien est apparemment bien supérieure à celle observée parmi les Han, d'après les données du chromosome Y (Wen et al. 2005), ce qui

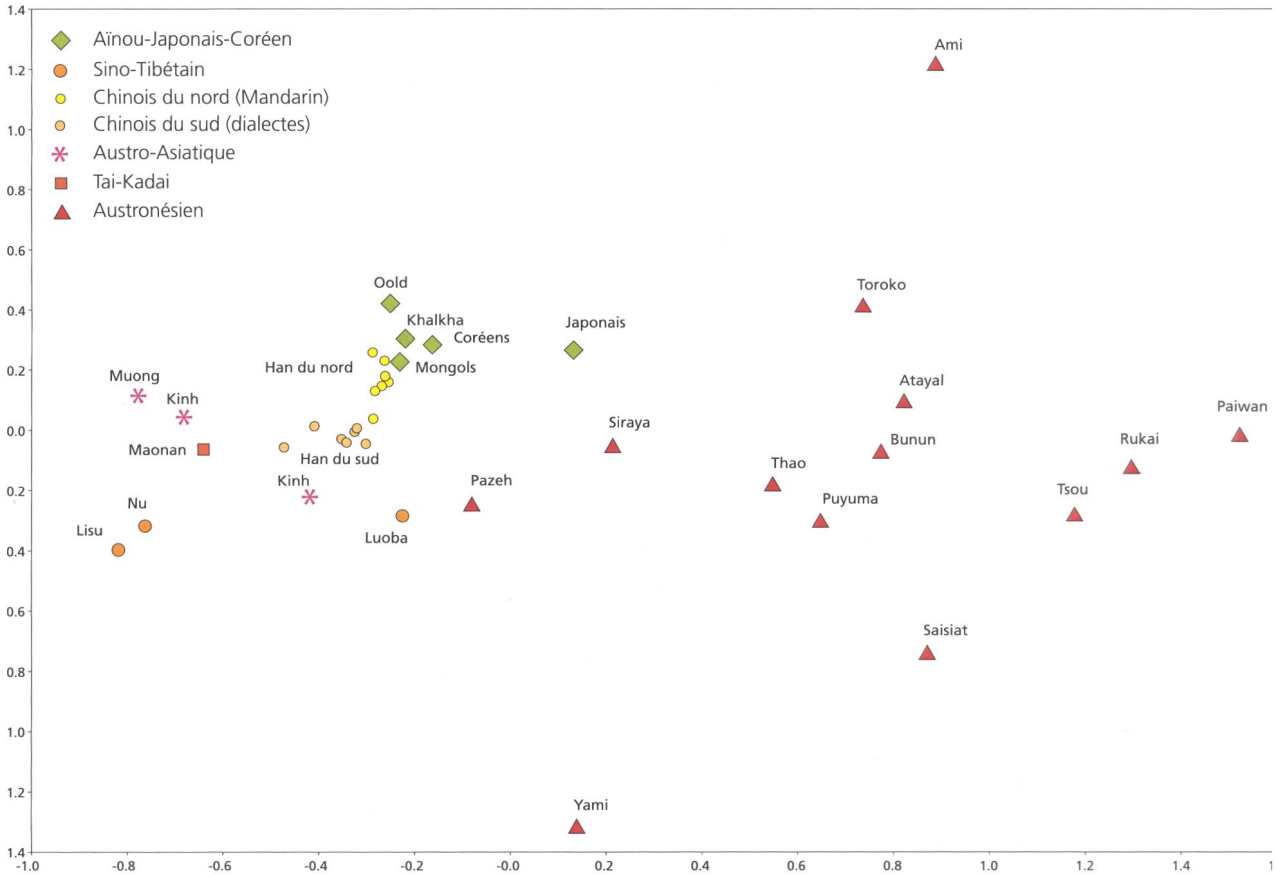
serait aussi vrai pour les Hmong-Mien d'après les données de l'ADN mitochondrial (Wen et al. 2005, et Poloni, comm.pers.) (Fig. 2). Pour les Hmong-Mien, ce résultat se justifie bien si l'on considère les effets d'une dérive génétique rapide due à leur morcellement géographique. Les Tai-Kadai, quant à eux, ressemblent beaucoup aux Austro-Asiatiques du point de vue de l'ADN mitochondrial (Li et al. 2007, Poloni comm.pers.), même s'ils ont la particularité d'une fréquence plus élevée de l'haplogroupe M7. Sur la base de tels résultats, il est difficile, à l'heure actuelle, d'émettre des hypothèses cohérentes sur l'origine et les différenciations de ces différents groupes sur une base purement génétique. Force est de constater que l'Asie du sud-est représente un véritable foyer de diversité génétique qui reste à être exploré.

3.3. Hypothèse «Out-of Taiwan»

Une théorie communément admise pour le peuplement de l'Asie du sud-est insulaire et le Pacifique est une expansion de populations à partir de Taiwan vers la moitié de l'Holocène, soit à partir de 4000 ans environ. Les bases de cette hypothèse sont à la fois de nature linguistique et de nature archéologique:

- d'un point de vue linguistique, le phylum austronésien inclut, d'une part, neuf branches très distinctes entre elles et parlées uniquement à Taiwan par diverses populations d'aborigènes, et, d'autre part, une dixième branche, le malayo-polynésien, qui s'étend sur toute la région couverte par ce phylum, des Philippines jusqu'à Madagascar, à l'ouest, et jusqu'à l'île de Pâques, à l'est. La diversité linguistique de l'austronésien étant la plus élevée à Taiwan: c'est donc sur cette île que ce phylum aurait eu son origine.
- d'un point de vue archéologique, Taiwan a livré des vestiges de riz et de millet fossilisés témoignant d'une domestication de ces céréales il y a 5000 ans au sud-ouest de l'île, et associés à un type de poterie ressemblant, d'une part, à celles que l'on retrouve en Chine du sud, et, d'autre part, à celles que l'on retrouve de façon continue, à partir de 4000 ans, dans les îles Batan, aux Philippines et en Indonésie. Peter Bellwood

Fig. 4: Analyse d'échelonnement multidimensionnel («multi-dimensional scaling analysis», ou MDS) de 39 populations d'Asie orientale testées pour le système HLA d'histocompatibilité humaine (ici moyenne des loci HLA-A, HLA-B et HLA-DRB1). Chaque point représente une population et son symbole la famille linguistique à laquelle la langue de cette population est affiliée. La disposition des points sur la projection reflète les distances génétiques entre populations. La valeur de stress de 0.10 indique que la fiabilité de la représentation est bonne. Source: Da Di (communication personnelle).



décrit dans le chapitre précédent (et d'autres références) les évidences archéologiques soutenant cette hypothèse «Out-of-Taiwan». L'idée que ce peuplement a été très rapide a été reprise par Diamond sous le nom, devenu populaire, d'«Express Train to Polynesia». Mais Bellwood lui-même n'approuve pas cette appellation: les Austronésiens auraient passé 1500 ans à Taiwan avant de peupler le reste de l'Océanie!

Si cette théorie reste la plus connue, d'autres hypothèses ont été avancées pour expliquer les profils génétiques particuliers des populations polynésiennes. Selon la théorie «slow boat» (Kayser et al. 2000; Kayser et al. 2006; Kayser et al. 2008b), des contacts prolongés auraient eu lieu entre les ancêtres des Polynésiens et des populations mélanésiennes en particulier au niveau des îles situées au large des côtes de Nouvelle-Guinée (Kayser et al. 2008a). Un autre scenario, repris du modèle «entangled bank» énoncé par Terrell (1988) inclut, lui, des interactions encore plus complexes entre Polynésiens, Mélanésiens et Asiatiques du sud-est.

L'interprétation des résultats relatifs aux différents systèmes génétiques est, il est vrai, encore difficile à concilier. Du point de vue des marqueurs classiques, les populations aborigènes de Taiwan se différencient considérablement les unes des autres, comme le montre la Fig. 4 sur le système HLA.

Les populations des plaines de la côte ouest, les Pazeh et Siraya, sont en général les plus proches, génétiquement, des populations du continent, alors que les populations du sud et de l'est de l'île (Paiwan, Ami, Puyuma), ainsi que les Yami (ou Tao) de l'île Lan-Yu sont souvent les plus divergentes et présentent un degré de diversité génétique plus réduit. D'après les résultats des analyses du système GM des immunoglobulines, le répertoire génétique s'appauvrit également du nord-ouest au sud-est de Taiwan (Sanchez-Mazas et al. 2008). Ces observations s'accordent avec l'hypothèse linguistique énoncée par Sagart (2008) à propos des innovations successives observées du nord au sud et sud-est de Taiwan le long de la côte ouest, et suggérant un peuplement progressif de l'île selon cet itinéraire. Depuis la côte est, une population aurait ensuite migré vers les Philippines, l'Indonésie, la Mélanésie et le Pacifique, donnant naissance, linguistiquement, à la branche malayo-polynésienne de l'austronésien. Une migration aurait également eu lieu de Taiwan vers l'Asie du sud-est, où les langues tai-kadai se seraient différenciées à partir de l'austronésien (Sagart 2004, 2005).

L'étude de l'ADN mitochondrial (Trejaut et al. 2005) confirme le rôle de Taiwan (ou de l'ensemble de la région comprenant Taiwan, les Philippines et Bornéo) dans l'histoire du peuplement du Pacifique: les lignages les plus fréquents sont identiques à ceux que l'on trouve sur le continent et suggèrent que Taiwan a été peuplée depuis la Chine. De plus, l'origine d'un lignage caractéristique (car très fréquent) des Polynésiens – le lignage B4a1a portant une délétion de 9 paires de bases, ou «motif polynésien» – peut être retracée à Taiwan durant l'Holocène. La grande diversité des profils génétiques observés chez les aborigènes taiwanais suppose néanmoins une assez longue période d'isolement et de diversification dans cette île, ce qui est en parfait accord avec les hypothèses archéologiques. Par ailleurs, il n'est pas exclu que ces populations aient hérité de gènes présents dans les populations établies sur l'île depuis la période pléistocène, avec lesquelles elles se seraient mélangées (Trejaut et al. 2005).

L'hypothèse d'origine des Austronésiens à Taiwan est cependant discutée par certains auteurs: notamment, Hill et al. (2007), tout en reconnaissant, d'après quelques lignages particuliers (D5, Y2, F1a3, F1a4, et, de manière plus controversée, M7c1c), qu'une migration «out-of-Taiwan» vers le sud a pu se produire au milieu de l'Holocène, soutiennent que celle-ci n'aurait eu qu'un impact mineur sur les patrimoines génétiques des populations d'Asie du sud-est insulaire par rapport à des expansions de populations qui se seraient produites après le dernier maximum glaciaire, soit à une période bien plus reculée, depuis la région de Bornéo, à la fois vers le nord (jusqu'à Taiwan) et vers l'est. Bien entendu, toutes ces interprétations se basent sur l'âge estimé et l'origine géographique supposée des nombreux lignages moléculaires observés dans cette région, et il faut donc les considérer avec beaucoup de prudence. Il est utile de relever, ici, que l'analyse multidimensionnelle basée sur les fréquences des haplogroupes mtDNA des populations de cette région montre une relation étroite avec la géographie, tout à fait compatible avec l'hypothèse «Out-of-Taiwan» (Fig. 5).

4. Conclusions

Cet exposé soulève des questionnements plutôt que d'apporter des réponses. On constate que la génétique explore avec acharnement l'histoire du peuplement de l'Asie orientale depuis quelques années, mais une vision d'ensemble est encore loin d'être atteinte. L'accès récent à des données génétiques chinoises ou d'autres régions asiatiques encore peu connues, comme le Vietnam, en est en partie la cause. Une deuxième raison, que nous ne pouvons que déplorer, est l'absence d'un programme scientifique commun qui rassemblerait l'ensemble des informations relatives aux divers systèmes génétiques connus et qui adopterait des approches similaires.

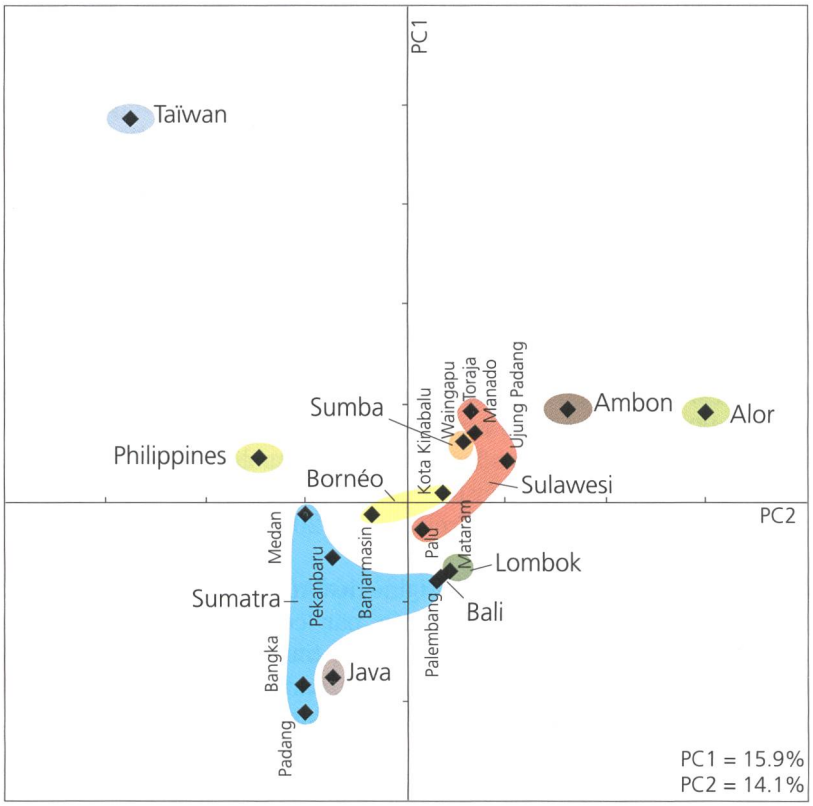


Fig. 5: Analyse en composantes principales (PCA) de 19 populations de taïwan et d'Asie du sud-est insulaire (Philippines, Indonésie) testées pour l'ADN mitochondrial. Chaque point représente une population. La disposition des points sur la projection reflète les distances génétiques entre populations (mais ici, l'axe horizontal «PC1» est légèrement plus informatif que l'axe vertical «PC2», ces deux axes représentant 15.9% et 14.1% de la variabilité génétique totale, respectivement). Les plages de couleur unissent les populations d'une même région géographique. Source: Hill et al. (2007), redessiné.

Malgré tout, des progrès ont été réalisés dans nos connaissances. Nous avons en effet appris que l'Asie orientale présentait une diversité génétique considérable de même ordre de grandeur que le continent africain, et que cette diversité était probablement l'héritage d'un peuplement très ancien, dont les traces paléontologiques n'ont, ma foi, pas été trouvées à ce jour. Nous savons aussi que le principal facteur ayant gouverné les différenciations génétiques des populations asiatiques est la géographie, qui a clairement signé les discontinuités génétiques de part et d'autre de l'Himalaya et les continuités génétiques à travers les autres régions, y compris le Pacifique. Parmi les pistes prometteuses qui se dessinent à présent, nous voyons que l'étude conjointe des gènes et des langues peut nous aider à comprendre la complexité des résultats obtenus dans l'un et l'autre des deux domaines impliqués. De plus, nous avons en quelque sorte la «chance» d'être dotés d'un patrimoine génétique complexe qui se transmet par les deux sexes mais aussi, en partie, par les lignées féminines et masculines indépendamment. Nous l'avons vu, la diversité génétique des populations actuelles ne nous raconte pas forcément la même histoire pour tous les marqueurs, et cela ne peut que nous encourager à poursuivre des approches qui tiennent compte de l'ensemble de cette information.

Remerciements

Je tiens à remercier la Dr Estella Poloni, généticienne, et le Dr Laurent Sagart, linguiste, avec lesquels je partage une collaboration étroite pour l'étude de l'histoire du peuplement de l'Asie orientale, ainsi que Maria-Eugenia Riccio et Di Da qui travaillent consciencieusement sur ce sujet dans le cadre de leur thèse de doctorat. Nos travaux bénéficient des soutiens financiers du Fonds National Suisse (FNS) pour la Recherche Scientifique (subside #3100A0-112651) ainsi que du CNRS dans le cadre de son programme Origine de l'Homme, du Langage et des Langues (OHLL). Je remercie également les Editions Einaudi pour avoir autorisé la publication de la version française de cet article avant sa parution dans leur ouvrage: *The Chinese Civilization: from its Origins to Contemporary Times*. Grandi Opere Einaudi, General editor Maurizio Scarpari, Volume I, tome 1: *Prehistory and the Roots of Civilization in China*, Edited by Roberto Ciarla and Maurizio Scarpari, 2009 in press.

Références

- **Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM.** 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evol Biol. 8: 45.
- **Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, Sarkar BN, Bhattacharya S, Annapoorna PC, Sreenath J, Sun D, Sanchez JJ, Ho SY, Chandrasekar A, Rao VR.** 2008. Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. Am J Phys Anthropol. 136: 19-27.
- **Barker G, Barton H, Bird M, Daly P, Datan I, Dykes A, Farr L, Gilbertson D, Harrisson B, Hunt C, Higham T, Kealhofer L, Krigbaum J, Lewis H, McLaren S, Paz V, Pike A, Piper P, Pyatt B, Rabett R, Reynolds T, Rose J, Rushworth G, Stephens M, Stringer C, Thompson J, Turney C.** 2007. The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). J Hum Evol. 52:243-261.
- **Cann R, Stoneking M, Wilson AC.** 1987. Mitochondrial DNA and human evolution. Nature, 325: 31-36.
- **Cavalli-Sforza LL, Menozzi P, Piazza A.** 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, New Jersey.
- **Chaix R, Austerlitz F, Hegay T, Quintana-Murci L, Heyer E.** 2008. Genetic traces of east-to-west human expansion waves in Eurasia. Am J Phys Anthropol. 136: 309-317.
- **Chandrasekar A, Saheb SY, Gangopadyaya P, Gangopadyaya S, Mukherjee A, Basu D, Lakshmi GR, Sahani AK, Das B, Battacharya S, Kumar S, Xaviour D, Su D n, et Rao VR.** 2007. YAP insertion signature in South Asia. Ann Hum Biol. 34: 582-586.
- **Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J.** 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am J Hum Genet. 63: 1824-1838.
- **Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J.** 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet. 12: 495-504.
- **Ding YC, Wooding S, Harpending HC, Chi HC, Li HP, Fu YX, Pang JF, Yao YG, Yu JG, Moyzis R, Zhang Y.** 2000. Population structure and history in East Asia. Proc Natl Acad Sci USA, 97: 14003-14006.
- **Excoffier L, Schneider S.** 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proceedings of the National Academy of Science USA, 96: 10597-10602.
- **Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ.** 2007. The Himalayas as a directional barrier to gene flow. Am J Hum Genet. 80: 884-894.
- **Hashimoto MJ.** 1986. The altaicization of Northern Chinese. In: J. McCoy, et T. Light eds), Contributions to Sino-Tibetan studies. E.J. Brill, Leiden, pp 76-97.
- **Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M.** 2007. A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet. 80: 29-43.
- **Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF.** 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. Am J Hum Genet. 69: 615-628.
- **Karafet TM, Osipova LP, Hammer MF.** 2008. The effect of history and lifestyle on genetic structure of North Asian populations. *In*: Sanchez-Mazas A, Blench R, Ross M, Peiros I, Lin M (eds), Past Human Migrations in East Asia:matching archaeology, linguistics and genetics. Routledge, London and New York, pp 395-415.
- **Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, Moyse-Faurie C, Rutledge RB, Schiefenhoevel W, Gil D, Lin AA, Underhill PA, Oefner PJ, Trent RJ, Stoneking M.** 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Mol Biol Evol. 23: 2234-2244.
- **Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M.** 2000. Melanesian origin of Polynesian Y chromosomes. Curr Biol. 10: 1237-1246.
- **Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Trent RJ, Suarkia D, Schiefenhovel W, Stoneking M.** 2008a. The Impact of the Austronesian Expansion: Evidence from mtDNA and Y Chromosome Diversity in the Admiralty Islands of Melanesia. Mol Biol Evol. 25: 1362-1374.
- **Kayser M, Lao O, Saar K, Brauer S, Wang X, Nurnberg P, Trent RJ, Stoneking M.** 2008b. Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. Am J Hum Genet. 82: 194-198.
- **Kennedy KA, Sonakia A, Chiment J, Verma KK.** 1991. Is the Narmada hominid an Indian Homo erectus? Am J Phys Anthropol. 86: 475-496.
- **Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L, Reddy BM.** 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. BMC Evol Biol. 7: 47.
- **Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, Liu W, Liu Y, Pan S, Qian J, Tan CC, Jin L.** 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. Am J Phys Anthropol. 134: 481-488.
- **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, Richards M.** 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science, 308: 1034-1036.
- **McDougall I, Brown F. H., Fleagle JG.** 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature, 433: 733-736.
- **Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R.** 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 5: 26.

- **Nei M, Roychoudhury AK.** 1993. Evolutionary relationships of human populations on a global scale. Mol Biol Evol. 10: 927-943.
- **Poloni ES, Sanchez-Mazas A, Jacques G, Sagart L.** 2005. Comparing linguistic and genetic relationships among East Asian populations: a study of the RH and GM polymorphisms. *In*: Sagart L, Blench R, Sanchez-Mazas A (eds), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. RoutledgeCurzon, London & New York, pp 252-272.
- **Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Reddy AN, Meka A, Reddy AG, Thangaraj** K, Singh L. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. PLoS ONE, 2: e1141.
- **Riccio MA.** 2008. Polymorphismes moléculaires des loci HLA-A, HLA-B et HLA-DRB1 dans des populations austro-asiatiques du Vietnam et du Sud-Est Asiatique. Anthropology and Ecology. University of Geneva, Geneva.
- **Sagart L.** 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. Oceanic Linguistics, 43: 411-444.
- **Sagart L.** 2005. Tai-Kadai as a subgroup of Austronesian. *In*: Sagart L, Blench R, Sanchez-Mazas A (eds), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. RoutledgeCurzon, London & New York, pp 177-181.
- **Sagart L.** 2008. The expansion of Setaria farmers in East Asia: a linguistic and archaeological model. *In*: sanchez-Mazas A, Blench R, Ross M, Peiros I, Lin M (eds), Past Human Migrations in East Asia: matching archaeology, linguistics and genetics. Routledge, London and New York pp. 133-157.
- **Sagart L, Blench R, Sanchez-Mazas A.** 2005. Introduction. In: L. Sagart, R. Blench, et A. Sanchez-Mazas eds), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. RoutledgeCurzon, London & New York, pp 1-14.
- **Sanchez-Mazas A.** 1990. Polymorphisme des Systèmes Immunologiques Rhésus, Gm et HLA et Histoire du Peuplement Humain. University of Geneva.
- **Sanchez-Mazas A, Langaney A.** 1988. Common genetic pools between human populations. Hum Genet. 78: 161-166.
- **Sanchez-Mazas A, Osipova L, Dugoujon J-M, Sagart L, Poloni ES.** 2008. The GM genetic polymorphism in taiwan aborigines. *In*: Sanchez-Mazas A, Blench R, Ross M, Peiros I, Lin M (eds), Past Human Migrations in East Asia: matching archaeology, linguistics and genetics. Routledge, London and New York, pp 273-296.
- **Sanchez-Mazas A, Poloni ES, Jacques G, Sagart L.** 2005. HLA genetic diversity and linguistic variation in East Asia. *In*: Sagart L, Blench R, Sanchez-Mazas A (eds), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. RoutledgeCurzon, London & New York, pp 273-296.
- **Shang H, Tong H, Zhang S, Chen F, Trinkaus E.** 2007. An early modern human from Tianyuan Cave, Zhoukoudian, China. Proc Natl Acad Sci USA, 104: 6573-6578.
- **Shen G, Wang W, Wang Q, Zhao J, Collerson K, Zhou C, Tobias PV.** 2002. U-Series dating of Liujiang hominid site in Guangxi, Southern China. J Hum Evol. 43: 817-829.
- **Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B.** 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am J Hum Genet. 77: 408-419.
- **Su B, Xiao J., Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L.** 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am J Hum Genet. 65: 1718-1724.
- **Terrell J.** 1988. History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. Antiquity, 62: 642-657.
- **Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L.** 2005. Reconstructing the origin of Andaman Islanders. Science, 308:996.
- **Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M.** 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. 3: e247.
- **Underhill PA.** 2005. A synopsis of extant Y chromosome diversity in East Asia and Oceania. *In*: Sagart L, Blench R, Sanchez-Mazas A (eds), The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. RoutledgeCurzon, London & New York, pp 297-313.
- **Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF.** 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA, 98: 10244-10249.
- **Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Deka R, Jin L.** 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol. 22: 725-734.
- **Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Deka R, Su B, Chakraborty R, Jin L.** 2004. Genetic evidence supports demic diffusion of Han culture. Nature, 431: 302-305.
- **Xue F, Wang Y, Xu S, Zhang F, Wen B, Wu X, Lu M, Deka R, Qian J, Jin L.** 2008. A spatial analysis of genetic structure of human populations in China reveals distinct difference between maternal and paternal lineages. Eur J Hum Genet. 16: 705-717.
- **Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, Yang H, Tyler-Smith C.** 2006. Male demography in East Asia: a north-south contrast in human population expansion times. Genetics, 172: 2431-2439.
- **Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP.** 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet. 70: 635-651.