

**Echanges génétiques et linguistiques entre les mondes  
occidental et oriental au nord et au sud de l'Himalaya**

JACQUESSON François

# Action "Origine de l'Homme, du Langage et des Langues"

## A. FICHE ADMINISTRATIVE

### **Titre du projet :**

Echanges génétiques et linguistiques entre les mondes occidental et oriental au nord et au sud de l'Himalaya

### **Mots-clés :**

Histoire des populations, linguistique, archéologie, histoire, ethnologie, génétique des populations, Himalaya, Asie Centrale, Inde, Assam

### **Résumé du projet :**

Il s'agit de comparer les données linguistiques et génétiques pour tenter de reconstituer l'histoire des populations dans deux zones géographiques liées l'une à l'autre, mais contrastées : le goulet qui sépare l'Inde et l'Asie du Sud-Est, et l'échelonnement de massifs montagneux qui joint les Pamirs à l'Altaï. La première zone correspond à une barrière nette entre Est et Ouest, la seconde correspond au contraire à une zone de gradation diffuse. La combinaison non seulement réunit les passages essentiels entre Est et Ouest du Continent, mais permet une typologie des zones de contact.

Il s'agit donc d'histoire des populations et des peuplements, en combinant les ressources de la linguistique et de l'ethnologie, de la paléoanthropologie, celles de l'anthropologie biologique – notamment génétique.

## 1. Responsable scientifique du projet

Nom : JACQUESSON

Prénom : François

Grade : CR1

Discipline : linguistique

Etablissement de rattachement : Laboratoire des langues et civilisations de tradition orale (LACITO)

Adresse professionnelle : 7 rue Guy Môquet, Bât. 23

Code postal : 94800

Commune : Villejuif, France

Tél. 01.49.58.37.78

Fax : 01.49.58.37.79

E-Mail : [jacquess@vjf.cnrs.fr](mailto:jacquess@vjf.cnrs.fr)

## 2. Laboratoire de rattachement de l'équipe de recherche

Laboratoire des langues et civilisations de tradition orale (LACITO)

Unité CNRS,

FRE 2204

délégation régionale de Thiais

Directeur du laboratoire : Madame Zlatka Guentcheva

Adresse : 7 rue Guy Môquet, Bât. 23

Code postal : 94800

Commune : Villejuif, France

Tél. 01.49.58.37.78

Fax : 01.49.58.37.79

### **3. Autres laboratoires ou organismes partenaires**

Laboratoire d'Anthropologie Biologique du Museum  
Laboratoire d'Immunogénétique humaine de l'Institut Pasteur  
Université Paul Sabatier de Toulouse  
Institut Français d'Etudes sur l'Asie Centrale  
Université de Franche-Comté (Besançon)  
Université du Michigan

#### A/ Laboratoire d'Anthropologie Biologique du Museum National d'Histoire Naturelle

Disciplines : biologie humaine, génétique des populations  
Directeur : Prof. André Langaney

Membres partenaires:

Langaney, André

Professeur au Museum et à l'Université de Genève

e-mail : [langaney@mnhn.fr](mailto:langaney@mnhn.fr)

Heyer, Evelyne

CR1

e-mail : [eheyer@mnhn.fr](mailto:eheyer@mnhn.fr)

Sibert, Alexandre

doctorant AMX

e-mail : [sibert@mnhn.fr](mailto:sibert@mnhn.fr)

Menecier, Philippe

technicien classe exceptionnelle

e-mail : [phm@mnhn.fr](mailto:phm@mnhn.fr)

Adresse : 17 place du Trocadéro

75116 Paris

Tél. 01.44.05.72.52

#### B/ Laboratoire d'Immunogénétique Humaine de l'Institut Pasteur

Disciplines : génétique humaine, génétique des populations

Directeur : Prof. Marc Fellous

Membres partenaires :

Fellous, Marc

Professeur

e-mail : [mfellous@pasteur.fr](mailto:mfellous@pasteur.fr)

McElreavey, Ken

Chef de laboratoire

e-mail : [kenmce@pasteur.fr](mailto:kenmce@pasteur.fr)

Quintana-Murci, Lluís

post-doctorant INSERM

e-mail : [quinatana@pasteur.fr](mailto:quinatana@pasteur.fr)

Adresse : 25 rue du Docteur Roux

75724 Cedex 15 Paris  
Tél. : 01.45.68.85.37 ou 76 Fax : 01.40.61.31.53

C/ Université Paul Sabatier de Toulouse

Discipline : génétique des populations

Directeur : Brigitte Crouau-Roy

Membre partenaire :

Crouau-Roy, Brigitte

Professeur d'université

e-mail : [bcrouau@cict.fr](mailto:bcrouau@cict.fr)

Adresse : 118, route de Narbonne  
31062 Toulouse Cedex 4, France

D/ Institut Français d'Etudes sur l'Asie Centrale, à Tachkent

Directeur : Vincent Fourniau

Membre partenaire :

Svetlana Jacquesson

secrétaire scientifique (à partir de septembre 2000)

e-mail actuel : [Jacquesson.F.S@wanadoo.fr](mailto:Jacquesson.F.S@wanadoo.fr)

Adresse : 18A rue Rakatboshi (ex-Chpilkov)  
700031 Tachkent Ouzbékistan  
Tél. (998 71) 139.47.03 Fax : (998 71) 120 66 56

E/ Université de Franche-Comté de Besançon

Discipline : histoire de l'art et archéologie

Membre partenaire :

Schiltz, Véronique

Professeur d'université

pas d'e-mail

Madame Schiltz prend sa retraite cette année

Adresse personnelle : 44 bd. de Port-Royal, 75013 Paris

F/ Université du Michigan, à Ann Arbor (USA)

Discipline : Linguistique et Anthropologie

Membre partenaire

Burling, Robbins

Professor emeritus

e-mail : [rburling@umich.edu](mailto:rburling@umich.edu)

Adresse : Université du Michigan, Dpt of Anthropology  
1020 LSA Building  
Ann Arbor, Michigan 48109, USA

## B. PROJET SCIENTIFIQUE

en Annexe : Description de l'Equipe

### 1/ Méthode et Développement de l'enquête

(les rubriques correspondent à celles du Budget)

#### A/ Phase préparatoire

##### A-1/ Documentation & contacts à distance

Quoique le groupe des linguistes, ethnologues & historiens (baptisés "groupe LEH") connaisse les lieux, une opération de cette envergure exige des travaux complémentaires pour documenter l'ensemble des deux zones : auprès des collègues compétents, français et étrangers, et des bibliothèques et instituts divers. Même chose pour les équipes de généticiens (baptisés globalement "groupe GEN"), qui doivent serrer au mieux l'état actuel des prélèvements disponibles. Il faudra par exemple, pour l'Inde, aller à Londres (India Office) et à Tübingen; et pour l'Asie Centrale, aller à Moscou et à Petersbourg. Ces voyages auront d'ailleurs d'autres fonctions.

La mise en place d'un site web présentant le Projet, son ampleur, son avancement, certaines données collectées et les synthèses partielles, les régions impliquées, les implications théoriques à plus ou moins grande échelle, enfin le rapport complet final, sera à la charge des unités participantes.

Les généticiens et les autres concourront à développer un protocole d'enquête et à définir un protocole de consentement éclairé (dont les détails dépendront des pays considérés) destinés à accompagner systématiquement les prélèvements, et à les rendre utilisables même à long terme, pour des recherches ultérieures. L'action pluridisciplinaire permet de donner à ces protocoles une finesse et une fiabilité jusqu'à présent rarement atteinte.

##### A-2/ Négociations

Une phase décisive est la prise de contact avec les institutions de santé ou d'anthropologie des pays concernés, afin d'obtenir notamment les autorisations de prélèvement, toujours délicates quand il s'agit de sang. Prévoir l'intervention des institutions de tutelle (CNRS, Pasteur, etc.) et de longues tractations en vue de l'obtention des visas, etc.

Les démarches pour les financements supplémentaires auprès de plusieurs organismes sont déjà en cours, et ressortissent au même titre. Elles nécessitent plusieurs voyages. Ceux qu'on a évoqués ci-dessus auront aussi cette fonction.

Une excellente façon d'attirer la bienveillance des autorités étrangères est d'inviter certains responsables institutionnels à un colloque d'ouverture à Paris, qui permettra de montrer le sérieux de notre Projet. Et aussi de montrer aux collègues de diverses disciplines ce que nous voulons faire.

##### A-3/ Définition des sites d'intervention

Le groupe LEH doit établir une cartographie de la distribution actuelle des populations et des langues, en observant quels traits sont imputables à une époque récente, et lesquels peuvent révéler une communauté plus ancienne. Cet état des lieux est indispensable, pour qu'ensuite les généticiens puissent cibler au mieux les prélèvements. Cette étape du choix des sites est décisive.

Nos contacts et réseaux de relations sur place, ainsi que l'appui des institutions françaises à l'étranger (IFEAC à Tachkent, CEFIPRA, EFEO et CSH en Inde) sont également utiles aux visites des généticiens comme des autres.

Concrètement, cette étape se déroulera en deux temps qui peuvent se recouper. 1/ Des explorations complémentaires LEH 2/ Des missions de reconnaissance combinées LEH+GEN. Ces missions de reconnaissance détaillées, tant pour sonder des ethnies, que les dispensaires les plus proches, seront indispensables, afin de ne pas apparaître comme de simples "quêteurs de sang" auprès de populations chez qui cette question est très sensible, d'établir un premier contact favorable et d'expliquer notre démarche. Faute de quoi, nous risquons d'avoir à faire face à des réactions très négatives.

Pendant ce temps, les généticiens s'emploieront aussi à tester les meilleures utilisations des méthodes actuelles de prélèvement, en fonction des conditions de travail envisagées (voir plus loin).

Dans le même temps, il faut rassembler tout ce qui est avéré de l'histoire de ces populations. C'est une tâche difficile, puisque dans les deux cas il s'agit de zones de passages, et de cristallisations variées. La contribution de différentes disciplines est un principe de base de notre entreprise, comme le montre la

composition de l'équipe et aussi ses correspondants nombreux (qui ne sont pas détaillés ici). C'est le moment de développer une conception écologique des peuplements humains, qui puisse rendre compte de la juxtaposition de groupes maintenus distincts ou au contraire de la diffusion de types englobants qui gommant les différences. Les questions de relief, d'hydrographie, de zones de passage, de traditions et techniques alimentaires opposant les paysans aux éleveurs, et tous aux urbains – ou tissant entre eux des relations – doivent être intégrées à notre enquête. Car, dans une certaine mesure, ce sont ces mécanismes quand ils sont avérés sur place qui permettent de comprendre la logique des déplacements ou des sédentarités d'autrefois – tout en tenant compte des différences de densité de population. Cette écologie des populations, où les frontières linguistiques sont directement impliquées, mais ne se superposent pas nécessairement aux frontières ethniques, joue un rôle important dans notre conception du projet. C'est aussi un des apports théoriques utiles.

#### A-4/ Expérimentation des méthodes de prélèvement

Il convient enfin d'expérimenter, parmi les techniques de prélèvements connues, ou qui sont en train de se mettre au point, celles qui conviendront le mieux aux différentes phases du projet, en fonctions des populations ciblées.

#### B et C/ Phases d'enquêtes

Dans un second temps, interviennent les prélèvements des généticiens (combinées avec les enquêtes LEH) qui tenteront d'explorer, à l'aide de marqueurs aussi variés que possible en fonction des publications déjà disponibles et des conditions de prélèvement, l'histoire propre de chaque population et les distances génétiques entre les populations choisies. Voir les cartes détaillées plus loin. La variété de ces marqueurs, selon des techniques qui sont en train d'être développées (notamment par les membres du groupe Musée de l'Homme de notre équipe) peut, dans une certaine mesure, permettre une chronologie relative des distances génétiques, donc une chronologie des extensions, contractions, et ruptures, des zones de peuplement; et naturellement des échanges entre elles.

Pour promouvoir un travail réellement pluridisciplinaire, il importe – puisque ce plan sur quatre ans est assez long pour le permettre – d'assurer une interaction maximale entre les résultats génétiques et leurs interprétations en fonction des réalités locales. Dans la zone Sud, notamment, où la génétique des populations est inconnue et les zones linguistiques encore obscures, il est inévitable que les découvertes génétiques impliqueront un réexamen des données ethnologiques et linguistiques, et une seconde saison d'enquêtes génétiques. La même chose se produira dans deux régions de la zone Nord : dans le Ferghana et dans l'Altai-Sayan, à peuplement complexe. C'est l'équivalent chez les archéologues de deux campagnes de fouilles.

C'est là l'étape la plus neuve du projet, car si les populations actuelles sont, sinon toujours bien connues, du moins connues, les prélèvements génétiques sont rares dans la zone Nord, et pratiquement inexistant dans la zone Sud. Ajoutons que cette entreprise, qui poursuit en un sens celle que certains membres (le groupe de l'Institut Pasteur) de notre équipe ont menée avec succès à l'ouest de l'Himalaya (Pakistan, Iran, et ouest de l'Inde), rencontre aussi le Projet assez similaire organisé par M. Laurent Sagart plus loin en Chine et aux frontières de la Chine. De sorte que l'ensemble des travaux permettrait, à terme, une première cartographie génétique cohérente d'une grande partie du continent asiatique.

Voyages de plusieurs membres du groupe dans chaque zone pour la première saison (titre B), puis dans chaque zone pour la seconde saison (titre C). Il faut ensuite compter la location de véhicules sur de très longs trajets (plusieurs milliers de km) dans la zone Nord. Les voyages vers Gauhati (capitale de l'Assam) impliquent également ensuite de louer des véhicules; les frais seront sans doute moindre dans la zone Sud, mais les séjours seront plus longs. Les frais d'assurance, de séjour, et annexes seront pris en charge par les Laboratoires ou par les participants eux-mêmes.

Le coût du matériel pour les prélèvements biologiques est actuellement évalué (traitement et analyse des données inclus) à 250 FF par personne testée si on se limite à la technique des frottis buccaux, à 400 FF par personne si l'on peut opérer des prélèvements sanguins (voir partie génétique). Ensuite, le coût dépend du nombre de personnes testées. Etant donné qu'un bon étalonnage d'un site de test suppose environ 50 tests sur des personnes sans apparemment proche pour les petites populations (mais il faut prévoir davantage pour les populations vastes, comme dans la zone nord), et que l'intérêt est de pouvoir multiplier les sites pour pouvoir les comparer, on peut estimer l'ampleur des recherches à 10 ou 12 sites en zone Nord, à 10 ou 15 sites en zone Sud. Soit un total entre 20 et 30 sites, donc un minimum de 1500 personnes testées.

Le matériel nécessaire à la conservation des prélèvements dépend du type de prélèvement, et des tests techniques qui vont être effectués pendant la première année. Comme il est expliqué en D/, les frottis buccaux ne demandent pas de précautions particulières, mais les prélèvements sanguins nécessitent l'accès à un laboratoire dans les 48h, condition *sine qua non* pour réaliser des lignées cellulaires. Il faut donc prévoir d'indemniser des

biologistes locaux (vacations) qui travailleront sur place pour le projet. C'est d'ailleurs une excellente façon de promouvoir une coopération sérieuse.

## D/ Résultats

Un troisième temps est en effet l'interprétation des données, moment qui impose une collaboration étroite entre les chercheurs linguistes, ethnologues et historiens, voire archéologues ou paléanthropologues d'une part, et généticiens d'autre part. Car lorsqu'on aura montré les différences et proximités génétiques relatives de telles et telles populations, une connaissance approfondie de celles-ci, et des vraisemblances culturelles et linguistiques, sera nécessaire pour comprendre dans un cadre historique et humain l'histoire des populations porteuses des marqueurs identifiés. L'expérience prouve qu'aucune interprétation purement génétique n'est possible, sans que d'autres disciplines ne fournissent au moins des hypothèses sur les mouvements de population. Il est vrai qu'à ce moment du projet, d'une part les généticiens auront d'autant mieux appris à connaître les pays qu'ils auront pu y venir eux-mêmes pour prendre une vue concrète des choses, et d'autre part que les linguistes, ethnologues et historiens auront compris les contraintes, les singularités, et la puissance du travail de leurs collègues. Dans tous les cas une collaboration mettant en œuvre toutes les capacités et favorisant tous les éclairages est de rigueur.

Ceci explique à la fois le nombre des membres du projet, et sa géométrie particulière (voir Description de l'Equipe).

Cependant, il nous semble heureux de réunir en fin de parcours un colloque plus ample, non seulement pour rassembler les participants français et étrangers, mais un certain nombre de collègues extérieurs à la réalisation du Projet, mais non pas indifférents à ses résultats. Le détail des faits ne pourra être fixé qu'assez tard. Il nous semble bon de prévoir un budget permettant d'inviter des savants locaux, dans les domaines couverts par le Projet, qui se seront montrés particulièrement efficaces pour l'aider et susceptibles de participer à sa mise en valeur *in fine*. Cela permet aussi d'ouvrir de nouvelles possibilités pour l'avenir.

## 2/ Justification détaillée des zones d'enquête

### la zone sud

Le passage entre l'Inde et l'Asie du Sud-Est est ancien, mais n'a pu se faire que par des voies limitées, qui sont encore pratiquées de nos jours. Le succès considérable des langues indo-aryennes dans l'Inde du Nord, qu'on présente souvent (et non sans raison) comme une défaite progressive des parlers dravidiens, a été plus difficile dans le Nord-Est indien, où les parlers austroasiatiques (munda et môn-khmer) et surtout tibéto-birmans sont très présents. Ces derniers sont représentés par une centaine de langues, et cette région est un des lieux de la planète où la densité linguistique (en nombre de langues par unité de surface) est la plus élevée. Les parlers tai paraissent d'introduction plus récente.

L'Assam bénéficie de chroniques historiques dont l'usage a été apporté par les Tais, et qui situe leur arrivée au XIIe siècle, parmi des populations tibéto-birmanes identifiées. Celles-ci étaient déjà en place au moment de l'arrivée des éléments indo-aryens issus de l'Inde du Nord, comme l'indique une toponymie largement tibéto-birmane dans un pays dont les parties basses sont aujourd'hui majoritairement aryanophone (parlers assamais et bengalis). L'influence culturelle de l'Inde du Nord est attestée dès le VIe siècle par des inscriptions et des ruines, et les premiers immigrants de haute caste, appelés pour leur prestige, n'ont pas dû être beaucoup antérieurs ; c'est pourquoi on situe d'ordinaire le début de l'immigration indo-aryenne au début de notre ère. Ce processus lent d'immigration explique que de très nombreuses ethnies tibéto-birmanes aient conservé jusqu'à nos jours leur identité et leurs langues. A vrai dire, s'il est certain que plusieurs groupes tibéto-birmans étaient présents avant le début de notre ère, il est probable que certains autres ne sont arrivés qu'ensuite par le nord ou par l'est. La diversité linguistique des Tibéto-birmans est en effet considérable. Mais ces populations anciennes avaient elles-mêmes été précédées par d'autres. La démonstration, de l'unité ancienne des parlers munda et môn-khmer au sein d'un groupement dit " austroasiatique ", a prouvé que l'Assam avait été un couloir – difficile mais réel – de communication entre l'Inde actuelle et l'Asie du Sud-Est. La question fameuse des

mégalithes, si fréquents à la fois chez les Khasi (Môn-khmers d'Assam) et les Mundas de l'est de l'Inde, est souvent évoquée dans ce contexte, mais la datation en est obscure. Dans ses grandes lignes, la chronologie des expansions linguistiques est la suivante :

Cette chronologie est, pour les périodes anciennes, plus relative qu'absolue, et on ne doit pas la prendre strictement : nous ignorons quand arrivèrent en Assam les premières populations tibéto-birmanes, et plus encore quelle est l'époque d'installation des ancêtres des parlers môn-khmers, sinon qu'elle est associée à celle des parlers mundas, qui est antérieure à l'arrivée des parlers indo-européens en Inde du Nord, vers – 2000. Il en résulte par exemple que si les parlers ancêtres du môn-khmer d'Assam (khasi) datent de – 5000, l'introduction des premiers parlers tibéto-birmans peut s'être faite quelques siècles plus tard. Nous n'en savons rien.

L'archéologie de ces régions est encore dans une phase exploratoire.

En outre, les découpages politiques actuels (la frontière avec la Chine et la Birmanie) ne peuvent pas faire oublier les relations avec les ethnies de la frontière et au-delà. Elles montrent que cette frontière connaît des passages, qui sont d'une grande importance pour l'étude des peuplements. Cette extension du regard permettrait sans aucun doute de comprendre le passé plus ancien, si – pour ce que sont nos connaissances actuelles – nous ne perdions trop en précision.

### la zone nord

A l'inverse de la zone sud, la série de massifs montagneux qui “ part ” de l'Himalaya occidental et du complexe pamirien pour rejoindre l'Altaï et les monts Sayan, est de nos jours linguistiquement homogène : il s'agit presque uniquement de parlers turks (kazakh au nord, kirghiz au sud dans les hautes montagnes), bordés sur leur frange méridionale par le domaine iranien avec quoi le turk ouzbek fait transition, et sur la frange septentrionale par le domaine mongol-toungouse, devant quoi s'interpose le groupe compliqué des parlers turk de l'Altaï-Sayan.

Au sud de cette zone, le fait fondamental est la poussée des ethnies turkes et le retrait de la frontière iranienne au cours des siècles, depuis l'époque “ scythe ”, matérialisée par les nombreuses (y compris très récemment) découvertes de kourganes (tombes à tumulus). Durant les premiers millénaires avant et après JC, la steppe (qui est d'ailleurs en partie une création des grands troupeaux) semble essentiellement occupée par des populations de langue iranienne. Les poussées turko-mongoles ultérieures vont dissoudre cette continuité des parlers iraniens anciens, mais plusieurs traces en ont été retrouvées, et de nos jours encore il en existe des descendants (du point de vue linguistique) aux deux extrémités : d'une part les Ossètes du Caucase, d'autre part les langues iraniennes des Pamirs ; ces deux groupes offrent des similitudes suffisantes pour qu'on les considère comme des différenciations d'un ancien continuum des parlers iraniens des steppes. La dissolution de ce continuum est due à plusieurs étapes de poussée turko-mongole, dont ce n'est pas le lieu de détailler l'histoire. Il nous suffit de constater que les Turks, partis d'une région mal déterminée mais correspondant à peu près à notre axe (dont la partie septentrionale conserve les plus inscriptions turkes), ont gagné vers l'est, l'ouest, et vers le sud. La langue ouzbek, langue turke à phonologie iranienne, témoigne dans ses diverses variantes d'une région de contacts.

Dans le nord du domaine, le complexe Altaï-Sayan a été le foyer de nombreuses populations / langues aujourd'hui dispersées ou disparues : (a) l'élément turk, au contact de l'élément mongol et (peut-être) de l'élément toungouse. Les plus anciennes inscriptions turkes connues sont dans cette zone; (b) l'élément iénisséien, puisqu'il est avéré que l'ethnie ket, aujourd'hui un isolat linguistique, est en partie originaire de cette région; (c) l'élément samoyède, dont une partie est également attestée dans cette région encore au siècle dernier. Cette région doit donc être considérée comme une zone clef des expansions anciennes connues. Les données archéologiques dans cette zone sont nombreuses et d'un grand intérêt.

Cet axe Pamir-Sayan est important non seulement parce qu'il est de nos jours au centre du " couloir des steppes ", mais parce qu'il témoigne du phénomène central d'opposition entre le monde des nomades des steppes et le monde des formations d'agriculteurs sédentaires au sud (Iran, Inde, Chine). L'analyse de la situation actuelle et de celles que permettent de restituer, sur trois millénaires, les documents historiques et archéologiques, offre une fenêtre particulièrement importante pour comprendre d'une part la mise en place du contraste nomade / sédentaire et la néolithisation, mais aussi – en fonction de ce que nous comprenons de l'histoire des climats – de la période qui a précédé.

Plusieurs localités sur cet axe offrent l'occasion d'étudier sur le vif la juxtaposition des peuplements écologiquement distincts à cause de l'altitude : la zone de passage entre Altaï et Sayan d'une part, la zone des piémonts pamiriens (passage Kazakh-Kirghiz et micro-systèmes du Ferghana) de l'autre. Enfin, cet axe Pamir-Sayan est la " frontière historique " des zones d'influence de la Chine, et donc une des interfaces majeures entre Occident et Orient. A ce titre, elle continue dans l'espace la confrontation dont témoigne la zone sud : le secteur Assam-Bengale.

### 3/ La génétique des populations

Depuis de nombreuses années, les généticiens étudient les variations génétiques entre les individus et les populations dans le but de comprendre ces différences et leur signification en terme d'interactions entre les peuples. Les études de ces différences génétiques, connues sous le nom de polymorphisme, peuvent apporter des informations importantes sur les relations de parenté entre individus et sur l'histoire des différentes populations. Quand il existe plus d'un allèle en un même locus, on calcule les proportions respectives de chaque allèle dans chaque population et on en dérive ainsi des représentations (par exemple géographiques) des variations génétiques.

On peut alors confronter les classifications génétiques et linguistiques, tenter de retracer les origines génétiques des individus qui parlent une même langue ou des langues différentes, examiner les échanges génétiques relatifs entre les groupes parlant des langues différentes et, dans le même contexte, mettre en évidence des processus de remplacement des langues (résultats d'une dominance culturelle, par exemple) et/ou des gènes (Hurles M.E.,1998).

Les données concernant plusieurs locus voisins permettent d'estimer de façon détaillée les caractéristiques de l'histoire de chaque région du génome. De plus, les comparaisons entre des régions différentes permettent de faire le départ entre les forces évolutives qui agissent sur des régions spécifiques (comme la sélection) et des forces qui agissent sur le génome dans son ensemble (comme la taille des populations, leur structuration, les flux géniques entre elles, etc.). Ces forces sont illustrées par le gène codant la chaîne  $\beta$  de l'hémoglobine : quelques uns des allèles mutés de ce gène confèrent une résistance spécifique au paludisme et, par conséquent, leur distribution géographique est pour partie le résultat d'un intense effet de sélection. Des observations similaires ont été rapportées pour le gène du récepteur de la mélanocortine 1, impliqué dans la pigmentation chez l'homme (peau, cheveux).

Certaines régions du génome sont transmises par un seul des parents : l'ADN mitochondrial (ADNmt), hérité de la mère, et l'essentiel du chromosome Y, hérité du père, qui se sont montrés très informatifs en génétique des populations (Quintana-Murci L., 1999). L'utilité de ces régions tient au fait que l'ADNmt et la majorité du chromosome Y échappent, en principe, à la recombinaison méiotique et sont par conséquent transmis à la génération suivante avec pour seules modifications d'éventuelles mutations. Ces systèmes haploïdes ont aussi l'avantage de présenter très peu de mutations récurrentes (dites "reverse"). Cette

caractéristique permet en particulier, en connaissant les taux de mutation, d'estimer l'âge des événements mutationnels et d'affiner la mesure des distances génétiques entre populations permettant d'inférer une partie de leur histoire. De plus, l'analyse de ces chromosomes rend possible l'étude spécifique des lignées femelle ou mâle et l'identification de comportements reproducteurs et migratoires différents selon les sexes. Enfin, le chromosome Y et l'ADNmt sont souvent considérés comme sélectivement neutres; il sera néanmoins utile de considérer que des facteurs réduisant la diversité de l'ADNmt peuvent opérer dans certaines populations.

A l'inverse, les autres régions du génome (autosomes et chromosome X), transmises par les deux parents, sont sujettes à la recombinaison. Ceci rend très difficiles certaines approches de la génétique des populations développées pour l'ADNmt ou l'Y. En revanche, il est possible d'utiliser d'autres outils fondés, par exemple, sur les mesures de déséquilibre de liaison (i.e. le fait que certains gènes ne sont pas transmis indépendamment d'une génération à la suivante). De plus, le grand nombre des polymorphismes connus pour ces gènes est intéressant.

L'étude génétique des populations des deux zones comprendra deux parties classiques (études intra-population et inter-population à l'intérieur de chaque zone) suivie d'une comparaison des deux zones, en s'appuyant sur les polymorphismes de nombreux locus (ADNmt, chr X et Y, marqueurs autosomiques), ainsi que des systèmes classiques (HLA, Gm, etc.).

Nous souhaitons, lors de l'analyse du jeu de données et dans la mesure du possible, nous associer aux équipes qui ont développé les différents tests mis en œuvre.

L'étude intra-population s'intéressera en particulier à

- La démo-génétique de chaque population (e.g. la mise en évidence d'expansions récentes) à l'aide de techniques établies (coalescence, mean pairwise differences, etc.)
- L'existence de goulots d'étranglements et leur datation.

Ce travail précédera les calculs de distances entre populations puisqu'il conditionne l'interprétation de celles-ci (Fay and Wu, 1999; Gaggiotti and Excoffier, 2000). Par ailleurs, on utilisera les informations linguistiques, ethnologiques et paléo-anthropologiques comme cadre interprétatif des résultats de ces analyses (Poloni et al., 1997). Enfin, il est possible que les données haplotypiques donnent des informations quant aux migrations des populations étudiées (Quintana-Murci et al., 1999).

L'étude inter-population permettra de

- Définir entre quels niveaux se répartit la diversité génétique (entre les populations ou les groupes de populations définis suivant des critères externes : linguistiques, archéologiques, paléo-anthropologiques), entre autres au moyen de l'AMOVA (logiciel Arlequin).
- Calculer des corrélations entre données génétiques, linguistiques et géographiques pour les différents systèmes étudiés (Poloni et al., 1997).
- Soumettre ce même jeu de données aux techniques d'auto-corrélation spatiale, pour mettre en évidence d'éventuels "clines" (Barbujani et al., 1987).

- Utiliser enfin les méthodes que l'équipe du Laboratoire d'Anthropologie de l'Université de Genève met au point actuellement et qui concernent l'évaluation de la perméabilité et la réalité de certaines frontières définies *a priori*.

Par ailleurs, il est essentiel d'essayer de dater des événements de migration mis en évidence entre les populations étudiées. Pour cela, il sera important de réfléchir à la notion de distance entre populations actuelles et à la possibilité d'interpréter ces dernières en fonction d'échelles temporelles propres à chaque système de marqueurs.

La comparaison des deux zones portera sur les deux niveaux d'analyse précédents :

- étude comparée de l'histoire des populations dans les deux zones
- étude comparée de la corrélation entre diversité linguistique et diversité génétique. On s'attend, entre autres, à l'existence de frontières plus fortes (i.e. moins perméables) dans la zone sud.

#### 4/ bibliographie succincte

##### a/ génétique :

Barbujani G. (1987). Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117 : 777-782

Comas D. et al. (1998). Trading genes along the silk road : mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 63 : 1824-1838

Fay J.C. and Wu C.I. (1999). A Human Bottleneck Can Account for the Discordance Between Patterns of Mitochondrial Versus Nuclear DNA Variation. *Mol Biol Evol* 16 : 1003-1005

Gaggiotti O.E. and Excoffier L.(2000). A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. *Proc R Soc Lond* 267 : 81-87

Hurles ME et al. (1998). Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 65 :1437-48.

Pérez-Lezaun A. et al. (1999). Sex-specific migration patterns in Central Asian populations revealed by the analysis of Y-chromosome STRs and mtDNA. *Am J Hum Genet* 65 : 208-219

Poloni E.S. et al. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61 : 1015-1035.

Quintana-Murci L. et al. (1999). Genetic evidence of an early exit of *Homo sapiens* from Africa through eastern Africa. *Nature Genetics*, 23 : 437-441

Quintana-Murci L. et al. (1999). L'ADN mitochondrial, le chromosome Y et l'histoire des populations humaines. *M/S* 15 : 974-982

Tishkoff SA et al. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271 : 1380-1387

##### b/ ethnologie, linguistique, histoire

- Burling, Robbins (1983). "The "sal" languages", *Linguistics of the Tibeto-Burman Area* (Berkeley), 7/2, 1-31.
- Fortescue, Michael. (1998). *Language Relations across the Bering Strait*. Cassell.
- Francfort, H.-P., ed. (1990). *Nomades et sédentaires en Asie Centrale*. CNRS.
- Harris, David R., ed. (1996). *The Origins and Spread of Agriculture and pastoralism in Eurasia*. University College London Press.
- Jacquesson, F. (1999). "Abrégé d'histoire de l'Assam", *Journal Asiatique* 287/1, 191-283.
- Mallory, J.P. (1989). *In Search of the Indo-Europeans*. Language, Archaeology, and Myth. Thames & Hudson.
- Masica, Colin (1991). *The Indo-Aryan Languages*. Cambridge Univ. Press.
- Menges, Karl H. (1968). *The Turkic Languages and Peoples*. Harrassowitz.
- Schiltz, Véronique (1994). *Les Scythes et les Nomades des steppes*. Gallimard.
- Sinor, Denis, ed. (1990). *Early Inner Asia*. Cambridge University Press.

## ANNEXE : DESCRIPTION DE L'EQUIPE

Nous entendons nous tenir avec peu de modifications – sauf en ce qui concerne les contacts locaux en Asie – à la liste des membres qui figure ici. Mais nous entendons aussi avoir recours à des compétences variées pour tous les aspects de l'histoire du peuplement des deux régions qui nous paraissent importants : histoire ancienne, archéologie, paléanthropologie, histoire du climat, etc. La présence dans notre équipe de plusieurs excellents russophones, sans parler de ceux qui manient les langues locales, nous évitera la plupart du temps, sauf dans les régions reculées, le recours aux interprètes. En fonction des étapes du Projet, et des lieux de travail, notre groupe agrégera donc plus ou moins d'auxiliaires. Il est souhaitable de susciter la collaboration des meilleurs savants locaux.

En outre, chaque laboratoire envisage de faire travailler sur certains aspects du Projet un ou plusieurs étudiants, soit à fin de Mémoire, soit comme une partie de thèse, ou plus simplement comme une participation active à la vie des laboratoires. C'est en effet une excellente occasion pour s'initier à la recherche. Si certains d'entre eux trouvent là l'occasion de gagner des bourses, leur présence sur leur terrain sera fort utile.

L'équipe elle-même se compose de deux versants, celui des généticiens (GEN : 7 membres), et les autres, spécialistes de linguistique, d'ethnologie, d'histoire (LEH : 5 membres).

Les notices sont rangées par ordre alphabétique des noms de famille :

Robbins Burling est le fondateur des études linguistiques dans le Nord-Est indien. Ethnologue de formation, il est venu au Meghalaya dans les années 1950, et a publié plusieurs livres connus d'ethnologie, tant sur cette région que sur des problèmes généraux. Son premier article de linguistique date de 1956. Il a publié sur le garo (langue qu'il parle) une grammaire et de nombreux articles. Il a publié en 1983 un article important sur certaines langues d'Assam, qui mettent en valeur un passage possible vers l'Asie du Sud-Est. C'est donc un éminent spécialiste du **Sud**.

Marc Fellous est Professeur de Génétique Humaine (Paris VII et l'Institut Pasteur). Depuis plus de vingt ans il travaille avec son équipe INSERM à l'Institut Pasteur sur le chromosome Y. Depuis 1985, avec les généticiens de l'Ecole de Pavie, il utilise les polymorphismes du chromosome Y pour étudier l'histoire des populations méditerranéennes. Il a contribué à décrire un Protochromosome Y commun à cette région du monde, dont le gradient de fréquence Orient-Europe renforce l'hypothèse d'une expansion néolithique humaine de la Corne de l'Afrique vers l'Europe.

Evelyne Heyer, chargée de recherche CNRS au laboratoire d'Anthropologie biologique, travaille depuis plusieurs années en génétique des populations humaines. Elle a montré dans différentes populations comment l'histoire démographique et sociologique des populations humaines influe sur l'évolution de leur pool génique; ces travaux lui ont valu une médaille de bronze du CNRS en 1999. Elle s'intéresse actuellement aux différents estimateurs de diversité génétique qui permettent de retracer l'histoire des populations humaines.

François Jacquesson, chargé de recherche CNRS au LACITO, linguiste, connaît les deux régions. **Nord** : Il a séjourné en Mongolie et en Asie Centrale, et a publié plusieurs articles sur les langues ouraliennes, altaïques, paléo-sibériennes, leurs rapports et leur histoire. Il dirige au LACITO un groupe “ Typologie des langues de l’Eurasie du Nord ”. **Sud** : Depuis cinq ans, il se rend régulièrement en Assam, où il étudie certaines langues tibéto-birmanes, sur quoi il a publié plusieurs articles; ainsi que sur l’histoire de ce pays. Son intérêt majeur est pour la théorie de l’évolution des langues, à quoi il a contribué.

Svetlana Jacquesson, qui soutiendra une thèse d’ethnologie sur l’Asie Centrale en septembre, vient d’être nommée Secrétaire scientifique de l’Institut d’Etudes sur l’Asie Centrale à Tachkent (Ouzbékistan) où elle va développer un programme de recherches sur les voies traditionnelles de migration et d’échanges chez les populations de pasteurs semi-nomades. Elle a publié des articles sur l’histoire des techniques et la tradition orale chez les Kirghizes et les Kazakhs. Elle parle russe, kirghiz, kazakh, entre autres. C’est donc la spécialiste de la zone **Nord**, et l’ancrage essentiel du Projet dans cette zone.

André Langaney, généticien, est Directeur du Laboratoire d’Anthropologie Biologique du Musée de l’Homme (Muséum National d’Histoire Naturelle) et Professeur à l’Université de Genève. Ses recherches ont principalement porté sur le polymorphisme moléculaire des populations humaines et l’histoire du peuplement humain actuel. Il s’est aussi intéressé à l’enseignement, tant universitaire que grand public, de la biologie générale et humaine. Il est notamment l’auteur de plus de 180 publications scientifiques originales et de plusieurs ouvrages de vulgarisation scientifique

Ken McElreavey, chef de laboratoire à l’Institut Pasteur, dans l’Unité d’Immunologie Humaine, travaille depuis dix ans sur la génétique et la biologie du chromosome Y humain. Il a identifié sur ce chromosome un polymorphisme qui a été utilisé pour démontrer le flux génétique des individus mâles en Espagne, à travers la barrière linguistique basque / catalan. Il fait partie du *European Diversity Project*, coordonné par C.Tyler-Smith (Oxford) et M.Jobling (Leicester Univ.), qui vise à définir l’haplotype de 5 000 chromosomes Y dans les différentes populations de l’Europe.

Philippe Menecier, technicien classe exceptionnelle au Laboratoire d’Anthropologie biologique du Musée de l’Homme, est le trait d’union entre l’équipe des généticiens, à quoi il appartient, et l’autre équipe, car il est depuis plusieurs années associé au LACITO comme linguiste. Il a soutenu et publié une thèse où il décrit le dialecte eskimo de l’est du Groenland, sur lequel il a publié plusieurs articles. Il s’intéresse aux analogies de structure des langues de l’Eurasie du nord. C’est également un excellent russisant, et traducteur, et il a vécu en Russie. Lui aussi est un homme du **Nord**.

Lluis Quintana-Murci est post-doctorant INSERM dans l’Unité d’Immunogénétique Humaine de l’Institut Pasteur. Il est impliqué depuis sept ans dans l’étude de la diversité des populations humaines grâce à l’étude de l’ADN mitochondrial et du chromosome Y. Sa thèse de doctorat, soutenue en Italie, portait sur l’utilisation de ces marqueurs uniparentaux pour la génétique des populations. Il a publié récemment un article qui suggère une deuxième “ sortie d’Afrique ” de l’Homme moderne, voici 60.000 ans, via l’Ethiopie et l’Arabie, vers l’Inde, renouvelant ainsi l’étude du peuplement de l’Asie.

Véronique Schiltz, professeur d’université, est spécialiste de l’histoire et de l’archéologie des steppes. Elle a publié et dirigé d’importants ouvrages sur les Sarmates et

Scythes, dont *L'Art des Steppes* (Gallimard, coll. "L'Univers des Formes"). Elle est aussi l'auteur de plusieurs catalogues d'exposition, parmi lesquels *l'Or des Scythes* (au Grand Palais, 1975). C'est une russisante accomplie, auteur de traductions littéraires. Elle est en contact constant avec les archéologues français et russes travaillant dans la zone **Nord**.

À la fin de ses études d'ingénieur (1995), Alexandre Sibert a effectué un trimestre de recherches au département de physique théorique de l'Indian Institute of Science (Bangalore). Depuis lors, il s'est rendu en Inde chaque année et a notamment travaillé pour un dispensaire de Médecins du Monde (1997) dans le cadre de ses études de médecine. Parallèlement, il a obtenu un DEA (mention biométrie, 1998) pour avoir étudié les relations entre prédisposition génétique et histoire familiale des maladies monogéniques. Alexandre Sibert est actuellement doctorant au laboratoire d'Anthropologie biologique du Muséum et s'intéresse à l'effet de paramètres culturels sur l'histoire génétique des populations humaines.