HAL
open science

Du système de parenté à la diversité génétique dans les populations humaines d'Asie du Sud-Est

Goki Ly

To cite this version:

Goki Ly. Du système de parenté à la diversité génétique dans les populations humaines d'Asie du Sud-Est. Génétique humaine. Museum national d'histoire naturelle - MNHN PARIS, 2017. Français. NNT : 2017MNHN0007 . tel-01814180

HAL Id: tel-01814180

https://theses.hal.science/tel-01814180v1

Submitted on 13 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MUSEUM NATIONAL D'HISTOIRE NATURELLE
Ecole Doctorale Sciences de la Nature et de l'Homme – ED 227

Année 2017

N°attribué par la bibliothèque
|_|_|_|_|_|_|_|_|_|_|_|_|

THESE

Pour obtenir le grade de

DOCTEUR DU MUSEUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Génétique des populations humaines

Présentée et soutenue publiquement par

Goki LY

Le 12 décembre 2017

Du système de parenté à la diversité génétique dans les populations humaines d'Asie du Sud-Est

---

Sous la direction de **:** Monsieur Frédéric Austerlitz, Directeur de Recherche – CNRS
Et la co-direction de Madame Raphaëlle Chaix, Chargé de Recherche CNRS
et Monsieur Samuel Pavard, Maître de Conférence – MNHN

JURY :

M. Lounès Chikhi	**Directeur de Recherche - CNRS**	**Rapporteur**
M. Olivier Hardy	**Maître de Recherche – FNRS**	**Rapporteur**
M. Olivier Evrard	**Chargé de Recherche – IRD**	**Examinateur**
Mme. Emmanuelle Génin	**Directeur de Recherche - INSERM**	**Examinateur**
M. Michael Houseman	**Directeur d'Etudes – EPHE**	**Examinateur**

Remerciement

Tout d'abord, je remercie tous les membres du jury pour avoir accepté de lire ce manuscrit : Olivier Evrard, Emmanuelle Génin et Michael Houseman, les examinateurs. Je remercie tout particulièrement Lounès Chikhi et Olivier Hardy d'avoir accepté d'être les rapporteurs de cette thèse.

Je remercie Raphaëlle Chaix pour ses conseils avisés, pour son encadrement tout au long de cette thèse. Mais surtout pour sa présentation à l'ENS qui m'a donné envie de continuer dans cette voie et de rejoindre son équipe de recherche.

Je remercie Samuel pour sa patience, son sourire et son enthousiasme débordant qui m'ont permis de m'accrocher jusqu'au bout.

Je remercie Fredéric Austerlitz pour m'avoir donné l'occasion de réaliser cette thèse et aussi pour sa solidarité sur l'importance de la ponctualité pour les repas du midi.

Je souhaiterais remercier les membres de mon comité de thèse : Olivier Evrard, Emmanuelle Génin, Sylvie Le Bomin et Bruno Toupance pour leur encouragement, leur remarques et pour m'avoir empêché de partir dans toutes les directions.

Un grand merci à toutes les personnes ayant accepté de participer à l'échantillonner et d'avoir répondu aux questionnaires ethno-démographiques dans le cadre du projet SoGen sans qui ce projet de thèse n'aurait pas pu voir le jour. Merci à Chou Monidarin et Gérard Diffloth pour leur précieuse aide lors des missions. Merci à Frédéric Bourdier pour son accueil et son aide lors de la mission de retour au Cambodge.

Je remercie également tous mes collègues du Musée de l'Homme. Un merci tout particulier à Romain Laurent et Sophie Lafosse pour leur contribution inestimable du terrain à la bio-informatique en passant par la paillasse mais aussi à pour avoir été toujours disponible pour répondre à mes interrogations. Merci à mes co-bureaux du bureau 204 : Lou, Jérémy, Valentin, Christophe, Vanessa, Emma, Delphine, et les anciens co-bureaux Jonathan, Florence et Nastassia pour avoir égayé par votre présence, notre magnifique bureau. Merci aussi à Myriam et Priscille, mes anciennes co-bureaux au Jardin des Plantes, pour m'avoir accueilli dans la forêt amazonienne. Merci à Nina pour son soutien infaillible depuis le TP droso, et à Bérénice pour son énergie et son entrain sans limite. Merci à Evelyne, Céline, Flora, Laure S., Marie-Claude,

César, Claire, Laure G., Paul, Franz pour les enrichissantes discussions scientifiques, mais surtout extra-scientifiques,et leur convivialité.

Je remercie tous mes amis pour leur bonne humeur et tous les moments qu'on a passés ensemble. Merci aux amis de l'ENS, particulièrement à Toto pour m'avoir fait découvrir l'escalade, à Popo et Lolo pour les après-midi jeu et les après-midi raclette. Merci à tous les grimpeurs de l'ASP6 pour ne m'avoir jamais laissé tomber. Merci aux amis du LFJT, surtout à Sylvain pour avoir organisé nos réunions.

Enfin je souhaiterais remercier ma famille que je ne remercierais jamais assez pour leur soutien inconditionnel m'ayant permis de finir ce long périple.

Abstract

In humans, evolution is not only biological but also cultural. In addition, biological and cultural processes interact with each other. Kinship system is particularly interesting for population geneticists since it conditions when, where and with whom men and women reproduce and raise their children. It is therefore a key factor in the genetic evolution of human populations. However, most studies in human population genetics do not take into account the influence of social structures.

The aim of this Phd thesis was to deepen our understanding of the influence of kinship system on genetic diversity. We undertook a pluridisciplinary and quantitative approach by collecting genetic and ethno-demographic data from 12 Southeast Asian populations exhibiting a wide variety of descent (matrilineal, patrilineal, or cognatic) and residence (matrilocal, patrilocal, or multilocal) rules.

We first showed that kinship systems influence ethno-demographic variables that impacts the evolution of genetic diversity, notably sex-specific migrations. We found that patrilineal and matrilineal systems are not the symmetric opposite of each other. There was a higher residence rule flexibility in patrilineal populations compared to matrilineal populations. In consequence, male migration rates were similar between kinship systems whereas female migration rates were higher in patrilineal populations compared to matrilineal populations. In addition, we showed that matrilineal populations and cognatic populations with predominant matrilocal residence had a higher village endogamy compared to patrilineal populations. The ethnological reasons for these observations were discussed, in particular in the light of the matrilineal puzzle hypothesis.

We then tested to which extent such ethno-demographic differences between populations following different kinship systems impact their uniparental genetic diversity. We could detect the impact of the higher residence rule flexibility in patrilineal populations: indeed, Y chromosome diversity followed the male migration pattern, and was similar between kinship systems, whereas mitochondrial DNA diversity followed the female migration pattern, and was higher in patrilineal populations compared to matrilineal populations.

Finally, we focused on the influence of kinship systems on autosomal diversity, more specifically on inbreeding levels. We demonstrated that, due to larger village endogamy, inbreeding level was higher in matrilineal and cognatic populations compared to patrilineal populations

Together these results showed that the kinship system has to be considered as the combination of a set of crossing and interacting rules (descent, residence and alliance), whose effects on genetic diversity can be disentangled only by going beyond categorizations and performing a quantitative assessment of relevant ethno-demographic variables.

Key-words: genetic diversity, kinship system, migration, endogamy, matrilineal puzzle, flexibility, descent group, inbreeding, social organization, lineage, clan

Résumé

L'évolution humaine n'est pas seulement génétique, elle est aussi culturelle, et les processus culturels et génétiques interagissent entre eux. Plus particulièrement, le système de parenté, en déterminant quand, où et avec qui les individus se reproduisent et élèvent leurs enfants, est un facteur clé de l'évolution génétique des populations humaines. Cependant la grande majorité des études de génétique des populations humaines ignorent l'existence de ces structures sociales.

L'objectif de cette thèse est de remédier à ce manque en explorant par une approche pluridisciplinaire et quantitative l'influence des systèmes de parenté sur la diversité génétique de 12 populations d'Asie du Sud-Est présentant des règles de filiation patrilinéaire, matrilinéaire et cognatique associées à des règles de résidence patrilocale, matrilocale et multilocale.

Nous avons tout d'abord mis en évidence que les systèmes de parenté se répercutent sur les variables ethno-démographiques d'importance pour l'évolution génétique des populations, et notamment sur les migrations maritales sexe-spécifiques. En particulier nous avons observé que les systèmes patrilinéaires et matrilinéaires ne sont pas symétriques. Il existe une plus grande flexibilité de la règle de résidence chez les populations patrilinéaires par rapport aux populations matrilinéaires. Cette différence a pour conséquence des taux de migrations d'hommes similaires entre les systèmes de parenté alors que les taux de migrations de femmes sont plus élevés chez les populations patrilinéaires que matrilinéaires. En outre, nous avons montré que les populations matrilinéaires et cognatiques avec résidence matrilocale prédominante ont une endogamie de village plus élevée que les populations patrilinéaires. Les raisons ethnologiques de ces observations sont discutées, particulièrement en lien avec l'hypothèse du « puzzle matrilinéaire ».

Puis, nous avons exploré l'impact de ces différences ethno-démographiques entre populations suivant des systèmes de parenté différents sur leur diversité génétique uniparentale. Nous avons pu observer l'effet de la plus grande flexibilité de la règle de résidence chez les populations patrilinéaires : en effet, la diversité du chromosome Y suit le patron de migration des hommes, et est similaire entre les systèmes de parenté alors que celle de l'ADN mitochondrial suit le patron de migration de femmes et est plus élevée chez les populations patrilinéaires que matrilinéaires.

Enfin, nous nous sommes intéressés à l'effet des systèmes de parenté sur la diversité autosomale et plus spécifiquement sur la consanguinité. Nous avons montré que le taux de consanguinité est plus élevé dans les populations matrilinéaires et cognatiques que dans les populations patrilinéaires, ce qui s'explique par la différence d'endogamie de village entre les systèmes de parenté.

Pris ensemble, ces résultats montrent qu'il est nécessaire de prendre en compte le système de parenté comme une combinaison de règles (de filiation, de résidence et d'alliance) qui se croisent et interagissent, et dont l'effet sur la diversité génétique ne peut être appréhendé que par une analyse quantitative des variables ethno-démographiques pertinentes.

Mot clés : diversité génétique, système de parenté, migration, endogamie, puzzle matrilinéaire, flexibilité, groupe de filiation, consanguinité, organisation sociale, lignage, clan

Table des matières

Introduction 1

1. L'influence du système de parenté sur la diversité génétique 5

1.1 L'influence de la résidence sur la diversité génétique 5

1.2 L'influence de la filiation sur la diversité génétique 8

1.3 L'influence de l'alliance sur la diversité génétique 9

2. Vers une intégration plus complète du système de parenté dans les études de génétique des populations humaines 10

3. Objectif de la thèse 14

Chapitre 1 - Des systèmes de parenté aux variables ethno-démographiques 19

1. Introduction 19

2. Matériels et méthodes 21

2.1 Echantillonnage des populations 21

2.2 Exploitation des questionnaires 21

2.3 Estimation des variables ethno-démographiques d'intérêt pour l'évolution génétique des populations 24

2.4 Comparaison des systèmes de parenté 28

3. Résultats 31

3.1 Analyse en composantes principales et matrice de corrélation 31

3.2 Comparaison des variables ethno-démographiques entre systèmes de parenté 37

4. Discussion 45

5. Conclusion 51

Chapitre 2 - Des variables ethno-démographiques à la diversité uniparentale 55

1. Introduction 55

2. Article : Residence rule flexibility and descent groups dynamics shape uniparental genetic diversities in South East Asia 57

3. Discussion 85

3.1 L'influence de la flexibilité de la règle de résidence sur la diversité uniparentale 85

3.2 La difficulté de départager les influences de la résidence et de la filiation 87

3.3 L'influence des fissions de clans sur la diversité uniparentale 87

Chapitre 3 - Des variables ethno-démographiques à la diversité autosomale 91
1. Introduction 91
2. Matériels et méthodes 94
2.1 Description du jeu de données autosomal 94
2.2 Estimation de la différenciation génétique entre populations 96
2.3 Estimation de *l'allele sharing distance* entre individus 96
2.4 Estimation de la taille efficace des populations 97
2.5 Estimation de l'hétérozygotie haplotypique 98
2.6 Estimation du coefficient de consanguinité par FEstim 99
2.7 Détection des segments d'homozygotie 100
3. Résultats 101
3.1 Différenciation génétique entre populations 101
3.2 Distance génétique entre individus 102
3.3 Taille efficace des populations 104
3.4 Hétérozygotie haplotypique 105
3.5 Coefficients de consanguinité 106
3.6 Longueurs des segments d'homozygotie 109
4. Discussion 112
Conclusion 119
Annexe 127
Annexe 1 - Questionnaire ethno-démographique 127
Annexe 2 - Simulation de la dynamique de clan dans un village 132
1. Introduction 132
2. Matériels et méthodes 132
3. Résultats 134
4. Conclusion 136
Annexe 3 - Distribution des clans dans les villages Pacoh et Prai 137
Annexe 4 - Article : Matrilineal puzzle shapes genetic inbreeding patterns in human populations 140
Annexe 5 - Contrôle qualité effectué par Romain Laurent 162
Annexe 6 - Longueur totale des ROHs court et moyen 163
Références bibliographiques 164

Introduction

L'objectif de la génétique des populations est d'étudier les variations génétiques qui existent entre individus au sein d'une espèce mais aussi entre espèces proches. Dans l'espèce humaine, l'information génétique est codée sous forme d'ADN (acide désoxyribonucléique). Chaque individu, à l'exception des jumeaux monozygotes, possède un génome qui lui est propre. Ainsi, on retrouve en moyenne un *single nucleotide polymorphism* ou SNP tous les 1000 paires de base entre deux individus pris au hasard (International Human Genome Sequencing, 2001; The 1000 Genome Project Consortium, 2010). Ces différences entre individus sont issues de l'évolution de notre espèce au cours des 200 000 dernières années (McDougall et al., 2008). La diversité génétique est générée par les mutations. Les mutations sont des modifications spontanées de la séquence d'ADN. Elles peuvent être classés en deux grands catégories : les substitutions de nucléotide et les variations structurelles (Feuk et al., 2006).

Les substitutions aussi appelées *single nucleotide polymorphism* ou SNP sont les mutations les plus présentes dans le génome humain. Ce sont des variations d'un nucléotide (l'unité de base formant la chaîne d'ADN) à un locus spécifique dans l'ADN. La base de données dbSNP qui recense les SNPs pour l'espère humaine possède plus de 135 millions d'entrées validées lors de leur dernière mise à jour (version 150 datant du 3 février 2017). Les deux principaux mécanismes de formation des SNPs sont les erreurs de réplication de l'ADN et les erreurs de réparation de l'ADN après endommagement (Friedberg, 2003). Ces mutations surviennent en moyenne à une fréquence de 10^{-8} substitutions par site par génération (The 1000 Genome Project Consortium, 2010; Kong et al., 2012). Cependant, le taux de mutation n'est pas uniforme le long du génome et certaines régions du génome ont des taux de mutation plus élevés (Campbell et al., 2012).

Les variations structurelles sont des mutations impliquant des réarrangements de la séquence génétique : inversion, insertion, délétion, translocation et duplication. La longueur de ces réarrangements peut varier de quelques paires de base pour des duplications de petit motif d'ADN à des chromosomes entiers. Les variations structurelles sont plus rares que les substitutions, mais, de par l'importance de la modification apportée à la séquence d'ADN, peuvent avoir un effet important sur le phénotype (Stankiewicz et Lupski, 2010).

Les différentes mutations ainsi créés verront leur fréquence varier au cours du temps. Ces variations de fréquences alléliques résultent de l'action de différentes forces : la dérive génétique, la démographie, les migrations et la sélection (encadré 1).

Encadré 1 : Dérive génétique, démographie, migration et sélection

Dans une population de taille finie, les allèles (variants génétiques) portés par les individus formant la nouvelle génération est un sous-échantillonnage des allèles de la génération précédente. En l'absence de toutes autres forces évolutives (mutation, sélection, migration), les fréquences alléliques dans une population vont donc fluctuées de manière aléatoire d'une génération à l'autre, c'est la dérive génétique. L'ampleur de ces variations dépend de la taille efficace de la population, Ne. La taille efficace d'une population représente la taille théorique d'une population idéale (de type Wright-Fisher) pour laquelle l'évolution du coefficient de consanguinité serait identique à celle de la population étudiée (Henry et Gouyon, 2003). Plus une population à une taille efficace faible, plus les variations des fréquences alléliques seront importantes d'une génération à une autre et la vitesse et la probabilité de fixation/disparition d'un allèle sera importante. Dans une population avec une taille efficace de Ne, la vitesse moyenne de fixation d'un nouvel allèle est de 4Ne et la probabilité de fixation 1/2Ne pour les marqueurs diploïdes. A chaque génération, les allèles présents étant un sous-échantillonnage des allèles de la génération précédente, et s'il n'y a pas d'apport de nouveaux allèles par mutation, la dérive génétique va tendre à réduire la diversité génétique dans la population au cours du temps.

Ainsi, l'histoire démographique d'une population aura un effet sur la diversité génétique de cette population. Lors d'une réduction de la taille de population ou goulot d'étranglement, la dérive génétique sera plus forte et entraine une diminution de la diversité génétique. Au contraire, lors d'expansions démographiques, la taille efficace de la population augmente ce qui réduit la dérive génétique et augmente ainsi la diversité génétique dans la population.

La migration correspond au déplacement d'un individu ou d'un groupe d'individus d'une population vers une autre. Si les populations échangeant des migrants, les migrations peuvent réintroduire des allèles perdues par dérive et donc permet d'augmenter la diversité génétique dans chacune des populations.

Enfin, la sélection peut aussi modifier les fréquences alléliques dans une population. Un individu portant un trait lui procurant un avantage sélectif dans son environnement lui permet d'avoir plus de descendants qu'un individu n'ayant pas ce trait. Par conséquence, si ce trait est héritable, il augmentera en fréquence dans la population au fil des générations. Au niveau génétique, la sélection naturelle tend à augmenter la fréquence des allèles donnant un avantage au porteur et diminuer la fréquence des allèles conférant un désavantage au porteur. Ainsi, contrairement aux forces démographiques qui agissent sur tout le génome, la sélection naturelle n'influence que certaines régions du génome portant des mutations qui influence le succès reproducteur des individus.

Une des particularités de l'espèce humaine est que son évolution n'est pas seulement génétique, elle est aussi culturelle : les humains développent de nouvelles technologies, modifient leur régime alimentaire et créent des organisations sociales complexes. Par ailleurs, les deux évolutions ne sont pas indépendantes, les processus génétiques et culturels interagissent entre

eux (Laland et al., 2010). En particulier, le système de parenté (encadré 2) détermine quand, où et avec qui les individus se reproduisent et élèvent leurs enfants. C'est donc un facteur clé de la transmission des gènes entre générations mais aussi entre les populations humaines. De plus, les systèmes de parenté seraient apparus dès la formation de sociétés humaines. Comme Lévi-Strauss l'écrit : « dès les premiers débuts de la vie en société, nos ancêtres ont dû connaître et appliquer des règles de mariage très strictes » (1983). Ainsi, les systèmes de parenté auraient orienté le choix de partenaire dans les populations humaines tout au long de l'évolution humaine. Leur prise en compte est donc fondamentale pour comprendre l'évolution génétique des populations humaines.

Encadré 2 : Le système de parenté

Le système de parenté d'une population peut être caractérisé par trois composantes principales (Godelier, 2004) : l'alliance, la résidence et la filiation. Nous décrivons ici succinctement les principales règles que nous pouvons retrouver dans les populations humaines.

L'alliance

Les règles d'alliance d'une population regroupent toutes les préférences, prescriptions positives ou négatives concernant le choix d'un conjoint. Nous pouvons calculer que dans 32% des populations mondiales répertoriées dans le Standard Cross-Cultural Sample (SCCS), le mariage avec une cousine de premier degré est le mariage préférentiel ce qui confirme les résultats de (Murdock et White, 1969). Au contraire, les mariages avec les cousines de premier degré sont interdits dans 52% des populations.

Ainsi, un des types de mariages fréquemment retrouvé est le mariage avec la cousine croisée matrilatérale (la fille du frère de la mère) ou mariage MBD (mother's brother's daugther). 38% des populations ayant une préférence de mariage pour une cousine pratique ce type de mariage (Murdock et White, 1969). L'autre type de mariages entre cousins est le mariage avec la cousine parallèle patrilatérale (la fille du frère du père) appelé aussi le mariage arabe car caractéristique du système de parenté arabe (Cuisenier, 1962; Chelhod, 1965).

Tout comme les mariages avec la cousine, le taux de polygamie varie entre les populations mondiales (Murdock et White, 1969; Marlowe, 2000). Il a été estimé que 17% des populations mondiales sont monogames et ne pratiquent pas la polygamie. Dans 51% des populations, la polygynie (un mari pour plusieurs femmes) serait pratiquée par une faible proportion des hommes (moins de 20%), et dans 31% des populations, elle serait pratiquée par plus de 20% des hommes. Enfin, 1% des populations seraient polyandres (une femme pour plusieurs maris).

La résidence

La règle de résidence détermine le lieu de résidence d'un couple après leur mariage. La résidence est patrilocale dans 70% des populations mondiales présente dans l' « Ethnographic Atlas » de Murdock (Murdock, 1967; Levinson et Malone, 1980). Après le mariage, la femme vient s'installer avec ou à proximité de la famille de son mari. A l'inverse dans les populations matrilocales, qui représente 20% des populations de l' « Ethnographic Atlas », le mari vient s'installer avec ou à proximité de la famille de sa femme. Le reste des populations mondiales possèdent une règle de résidence multilocale (les conjoints vivent alternativement avec la famille du mari et de la femme pour un certain nombre d'années

avant de s'installer définitivement à proximité d'une des familles), ou une règle de résidence néolocale (les conjoints s'installent dans un nouveau lieu de résidence après leur mariage).

La filiation

La filiation est « le principe gouvernant la transmission de la parenté » (Ghasarian 1996 p57). Elle détermine à quel groupe de parenté l'individu est affilié. Nous pouvons distinguer trois grands types de filiation : unilinéaire, bilinéaire, et cognatique. Lorsque la filiation est unilinéaire, un seul des sexes transmet sa filiation à sa descendance. Dans le cas d'une filiation patrilinéaire ce sera le père, dans le cas d'une filiation matrilinéaire ce sera la mère qui transmet à ses enfants l'appartenance à son groupe de filiation (lignage, clan et/ou tribu). Ces groupes de filiation sont alors constitués d'individus qui se reconnaissent un ancêtre commun en lignée paternelle ou maternelle dans les populations patrilinéaires et matrilinéaires respectivement. A l'échelle du lignage, les individus peuvent généralement retracer la généalogie jusqu'à leur ancêtre commun, ce qui n'est pas le cas pour le clan ou la tribu, pour lesquels l'ancêtre commun peut être mythique (Ghasarian 1996). Dans le système bilinéaire, les enfants appartiennent à la fois au groupe patrilinéaire du père et au groupe matrilinéaire de la mère. Enfin, dans un système de filiation cognatique, il n'y a pas de groupe de filiation et les enfants sont reconnus comme appartenant à la famille du père et de la mère. Le système cognatique peut aussi être appelé indifférencié ou bilatéral. La filiation unilinéaire est estimée être présente chez 57% des populations mondiales, 45% seraient patrilinéaires et 12% matrilinéaires. La filiation bilinéaire serait présente chez 4% et la filiation cognatique chez 39% des populations mondiales (Godelier, 2004).

Cependant, la majorité des études s'intéressant à l'évolution génétique des populations humaines se fondent sur des modèles de population panmictique et suppose que les individus se reproduisent aléatoirement (Goldstein et Chikhi, 2002). Elles ne prennent donc pas en compte l'influence potentielle des systèmes de parenté des populations étudiées sur leur diversité génétique (Wakeley, 2004; Schaffner et al., 2005; Schiffels et Durbin, 2014). Il existe néanmoins un certain nombre d'études qui se sont intéressées à l'influence des systèmes de parenté sur la diversité génétique.

1. L'influence du système de parenté sur la diversité génétique

1.1 L'influence de la résidence sur la diversité génétique

Au cours de ces deux dernières décennies, un nombre croissant d'études se sont intéressées à l'impact du système de parenté sur la diversité génétique des populations. En particulier, l'influence de la règle de résidence a été le centre d'un vif intérêt de la part des généticiens des populations. Cet intérêt a émergé suite à l'observation de différences génétiques plus importantes entre populations à l'échelle mondiale pour le chromosome Y que pour l'ADN mitochondrial (Seielstad et al., 1998). Cette observation a été mise sur le compte de la prépondérance de la patrilocalité dans les populations humaines. En effet, environ 70% des populations humaines suivent une règle de résidence patrilocale (Murdock, 1967; Levinson et Malone, 1980). Dans une population patrilocale, la femme s'installe avec ou à proximité de la famille de son mari. La patrilocalité restreint donc le déplacement des hommes, favorise celui des femmes et crée ainsi un biais de migration entre les sexes. Ainsi à l'échelle mondiale, les différences génétiques plus importantes entre les populations pour le chromosome Y que pour l'ADN mitochondrial pourrait donc être dû à cette prévalence de la patrilocalité (encadré 3), en accord avec les observations de Seielstad et al. (1998).

Cependant, cette différence entre le chromosome Y et l'ADN mitochondrial n'a pas été retrouvée en utilisant un autre panel de populations mondiales (Wilder et al., 2004). Il a été proposé que le manque de congruence entre ces deux études pourrait être due à une différence de distances géographiques entre les populations échantillonnées dans ces deux études (Wilkins et Marlowe, 2006). Les auteurs ont montré par une approche de simulation que plus l'échantillonnage des populations est dense (i.e. les distances géographiques entre les populations échantillonnés sont faibles), plus les estimations de distances génétiques inter-populationnelles sont influencées par les patrons de migrations récentes. Ainsi, l'étude de Seielstad et al. (1998) ayant un échantillonnage dense mettrait en évidence au niveau génétique l'effet des migrations récentes (après la transition néolithique), lorsque la patrilocalité aurait été adoptée par une majorité de populations. A l'inverse, l'étude de Wilder et al. avec un échantillonnage plus éparse montrerait la trace de migrations plus anciennes lorsque les populations étaient encore des chasseurs-cueilleurs suivant en majorité une règle de résidence multilocale (Marlowe, 2004; Kramer et Greaves, 2011), ce qui expliquerait l'absence de consensus entre ces deux études. Par ailleurs, la différence entre marqueurs utilisés dans ces

deux études pourrait aussi être une source des différences observées (Wilder et al., 2004). L'utilisation de *single nucleotide polymorphisms* ou SNPs prédéterminés pour le chromosome Y dans l'étude de Seielstad et al. pourrait surestimer les distances génétiques entre populations par rapport à celles obtenues en utilisant des données de séquençage comme dans l'étude de Wilder et al. En effet, les SNPs prédéterminés tendent à être plus anciens que les variants détectés par séquençage. Du fait de la dérive, ils ont donc pu atteindre des fréquences alléliques plus contrastées entre populations et donc pourraient entrainer une surestimation des distances génétiques (Lachance et Tishkoff, 2013).

Encadré 3 : Hérédité uniparentale et comportements asymétriques selon le sexe

Les chromosomes sexuels déterminent le sexe de l'individu : les hommes possèdent un chromosome Y et un chromosome X, les femmes deux chromosomes X. Ainsi, le chromosome Y n'est présent que chez les hommes dans l'espèce humaine et n'est transmis que de père en fils.

L'ADN mitochondrial se trouve dans les mitochondries, des organites dont la fonction principale est de produire l'énergie utilisée dans les fonctions cellulaires. Lors de la reproduction sexuée, seul le gamète femelle transmet ses mitochondries à l'embryon (les mitochondries du spermatozoïde ne pénétrant pas l'ovule). Ainsi, l'ADN mitochondrial est transmis de mère en filles au cours des générations.

Le chromosome Y et l'ADN mitochondrial n'étant transmis que par un des sexes, ce sont des marqueurs dits uniparentaux, témoins génétiques de l'histoire d'un des sexes. Ce sont donc des marqueurs de choix pour étudier l'histoire spécifique des hommes ou des femmes. Les comportements culturels sexe-spécifiques tels que la règle de résidence, la règle de filiation, la polygynie, l'âge à la reproduction ou la transmission du succès reproducteur vont alors influencer ces marqueurs uniparentaux différemment (révisée dans Heyer et al., 2012). Par exemple, la patrilocalité en favorisant les migrations des femmes va avoir tendance à homogénéiser la diversité génétique mitochondriale entre les populations. A l'inverse, en réduisant les migrations d'hommes, elle va avoir tendance à augmenter les différences entre populations pour le chromosome Y.

Une étude plus récente, utilisant le panel CEPH-HGDP (Cann et al., 2002) et se fondant sur des données de séquençage a reconfirmé l'existence de différences plus importantes entre populations humaines pour le chromosome Y que pour l'ADN mitochondrial (Lippold et al., 2014). Cette étude a analysé la séquence complète de l'ADN mitochondrial ainsi que 500 kilobase (kb) de la partie non recombinante du chromosome Y chez 623 hommes échantillonnés dans 51 populations de par le monde. L'échantillonnage des populations étant géographiquement dense, cette étude mettrait en évidence la signature du passé proche tout comme l'étude de Seielstad et al. Par ailleurs, cette étude est fondée sur des données de séquence pour le chromosome Y et l'ADN mitochondrial, limitant ainsi les éventuels biais liés aux marqueurs génétiques utilisés.

L'effet de la règle de résidence sur la diversité uniparentale a également été retrouvé dans des études menées à l'échelle locale. Des distances génétiques plus importantes entre populations pour le chromosome Y que pour l'ADN mitochondrial ont été observées chez des populations patrilocales (Perez-Lezaun et al., 1999; Destro-Bisol et al., 2004; Malyarchuk et al., 2004; Nasidze et al., 2004; Pemberton et al., 2012b). Par ailleurs, une comparaison de trois populations patrilocales à trois populations matrilocales de Thaïlande montre une corrélation entre la diversité uniparentale et la règle de résidence : la diversité du chromosome Y est plus importante chez les populations matrilocales que patrilocales, et à l'inverse, la diversité de l'ADN mitochondrial est plus importante chez les populations patrilocales que matrilocales (Oota et al., 2001).

Cependant, d'autres études récentes n'ont pas retrouvé cet effet de la règle de résidence. Ainsi, des diversités uniparentales similaires ont été observées entre populations patrilocales et matrilocales Hmong-Mien en Thaïlande (Besaggio et al., 2007) et en Inde (Kumar et al., 2006). Un effet de la matrilocalité sur l'ADN mitochondrial, mais non de la patrilocalité sur le chromosome Y, a été observé sur l'île de Sumatra (Gunnarsdóttir et al., 2011). Différentes explications pour ce manque de consensus ont été proposées. Dans le cas de l'Inde, une absence de migration d'hommes et de femmes entre populations en lien avec la forte endogamie du système de castes (Kumar et al., 2006). A Sumatra, la patrilocalité serait trop récente (suite à une transition depuis un système matrilocal) pour avoir pu impacter la diversité du chromosome Y de la population patrilocale (Gunnarsdóttir et al., 2011). Il a été également proposé dans le cas de Sumatra que la règle de résidence soit plus flexible dans la population patrilocale : une proportion de mariages matrilocaux plus élevée qu'attendue compte tenu de la règle patrilocale limiterait l'impact de la règle de résidence sur la diversité génétique du chromosome Y (Gunnarsdóttir et al., 2011).

1.2 L'influence de la filiation sur la diversité génétique

Malgré la prédominance des études focalisées sur de la règle de résidence, l'influence de la règle de filiation sur la diversité génétique a néanmoins fait l'objet de plusieurs travaux de recherche (Chaix et al., 2004, 2007; Ségurel et al., 2008; Sanchez-Faddeev et al., 2013; Montinaro et al., 2016). Dans les populations unilinéaires, l'appartenance à un groupe de filiation (tribu, clan, lignage) est transmise par le père (patrilinéarité) ou par la mère (matrilinéarité) à ses enfants. Les groupes de filiation suivent donc un mode de transmission similaire au chromosome Y et à l'ADN mitochondrial chez les populations patrilinéaires et matrilinéaires respectivement (encadré 2). Nous nous attendons donc à ce que la structuration en groupe de filiation soit aussi reflétée dans une structuration génétique de la population pour les marqueurs uniparentaux. C'est ce qui a été montré dans le cas des populations patrilinéaires où deux hommes du même clan sont plus apparentés au niveau du chromosome Y entre eux que deux hommes pris au hasard dans la population (Chaix et al., 2004) ; et cela même en tenant compte de la distance géographique pouvant séparer les différents clans (Montinaro et al., 2016). Une autre étude montre que les clans sont génétiquement différenciés pour le chromosome Y mais pas pour l'ADN mitochondrial dans un village patrilinéaire au Ghana (Sanchez-Faddeev et al., 2013). Cette sous-structuration de la population augmente la force de la dérive génétique dans chaque groupe de filiation. Ces groupes de filiation étant transmis de manière unilinéaire, il n'y a priori pas d'apport de nouveaux haplotypes au sein des groupes de filiation. Ainsi, les haplotypes existant augmentent rapidement en fréquence, ce qui crée des « noyaux d'identité », des haplotypes majoritaires dans chaque groupe de filiation (Chaix et al., 2007). Chaque groupe de filiation étant de taille finie, la stochasticité démographique mène inévitablement à des extinctions de groupe comme prédit par le modèle Single Large or Several Small (SLOSS) utilisé en biologie de la conservation (Diamond, 1975; Frankham et al., 2002), ce qui réduit la diversité génétique globale (Chaix et al., 2007). Cette réduction de la diversité est accentuée par le mécanisme de fission linéaire lors de la formation des groupes de filiation. Ainsi, lorsque les groupes de filiation se divisent pour former deux nouveaux groupes, la séparation des individus peut être aléatoire, ou linéaire quand les individus les plus apparentés se regroupent entre eux. Ce dernier type de fission réduit la taille efficace dans les groupes de filiation ainsi créés comparé à une fission aléatoire (Smouse et al., 1981). Les fissions linéaires associées à des extinctions des groupes de filiation vont alors réduire la taille efficace sexe spécifique. Ainsi chez les populations patrilinéaires d'Asie Centrale, la taille efficace des hommes est inférieure

à celle des femmes, alors que chez les populations cognatiques, les tailles efficaces d'hommes et de femmes sont similaires (Ségurel et al., 2008).

1.3 L'influence de l'alliance sur la diversité génétique

Certaines études se sont aussi intéressées à l'influence de la règle d'alliance sur la diversité génétique (Watkins, 2004; Billari et al., 2007; Guillot et al., 2015). En définissant des préférences pour certains types de partenaires, les règles d'alliance écartent la population du mélange panmictique et ont potentiellement une influence sur la diversité génétique (Jacquard, 1968, 1969). Des études menées à l'échelle globale ont observé l'existence de fortes proportions d'individus consanguins dans certaines régions du monde où des préférences pour des mariages entre apparentés existent (Moyen-Orient, Asie du Sud, Amérique du Sud), attestant ainsi du potentiel impact que peuvent avoir les règles d'alliance (McQuillan et al., 2008; Leutenegger et al., 2011; Pemberton et al., 2012a; Gazal et al., 2015). Une étude théorique basée sur des simulations a montré l'effet que pourrait avoir la préférence du mariage avec la cousine croisée matrilatérale sur la diversité génétique (Guillot et al., 2015). Un respect strict de cette préférence dans une population patrilocale réduirait la diversité génétique des autosomes, du chromosome X et de l'ADN mitochondrial mais pas du chromosome Y. Cependant, lorsque la règle de mariage n'est pas suivie de façon stricte, il n'y a pas de réduction attendue de la diversité génétique.

Par ailleurs, la polygynie a été également mentionnée comme un des facteurs expliquant la réduction de la diversité génétique du chromosome Y (Salem et al., 1996; Kayser et al., 2003; Heyer et al., 2012; Guillot et Cox, 2014) ou la réduction de la diversité des autosomes par rapport au chromosome X (Hammer et al., 2008). En effet, la polygynie crée une plus grande variance du succès reproducteur pour les hommes que pour les femmes. Or, une grande variance du succès reproducteur réduit la taille efficace des populations (Kimura et Crow, 1963). Ainsi, la polygynie peut réduire la taille efficace des hommes par rapport à celle des femmes ce qui a une influence sur la diversité uniparentale (révisée dans Heyer et al., 2012).

2. Vers une intégration plus complète du système de parenté dans les études de génétique des populations humaines

Ces différents travaux ont contribué à montrer l'importance de la prise en compte du système de parenté dans l'étude de l'évolution génétique des populations humaines. Cependant, il nous est apparu que cette prise en compte reste lacunaire et l'intégration du système de parenté dans les études de génétique des populations humaines peut être améliorée. Ma thèse s'inscrit donc dans cette volonté de développer une vision plus complète de l'influence du système de parenté sur la diversité génétique.

Nous avons tout d'abord noté que peu d'études prennent en compte le système de parenté dans sa globalité, c'est-à-dire les différentes composantes du système de parenté de concert. En particulier, rares sont les études s'intéressant à la règle de filiation. En effet, la majorité des études sur la règle de résidence précédemment citées ne mentionne pas la règle de filiation des populations étudiées (Perez-Lezaun et al., 1999; Oota et al., 2001; Destro-Bisol et al., 2004; Nasidze et al., 2004; Kumar et al., 2006; Besaggio et al., 2007; Gunnarsdóttir et al., 2011). Or, si la règle de résidence peut influencer les migrations sexes spécifiques, la règle de filiation peut impacter les tailles efficaces sexe spécifiques (Smouse et al., 1981; Ségurel et al., 2008). De plus, il existe différentes combinaisons de règle de filiation et de résidence possibles. Dans les populations avec un système de parenté dit « harmonique », la règle de filiation et la règle de résidence coïncident, c'est-à-dire la filiation est patrilinéaire et la résidence patrilocale ou la filiation est matrilinéaire et la résidence matrilocale. Dans ces cas, nous pouvons penser que la filiation et la résidence ont des effets synergiques sur la diversité génétique. Ainsi, il est possible que dans une population patrilinéaire et patrilocale, la filiation patrilinéaire réduise la taille efficace des hommes par rapport à celle des femmes et la résidence patrilocale réduise les migrations d'hommes par rapport à ceux des femmes. Les deux règles combinées diminuent alors la diversité génétique du chromosome Y. Cependant, il existe aussi des populations avec des systèmes de parenté dit « disharmoniques » (par exemple, une filiation matrilinéaire associé à une résidence patrilocale). Dans ce cas, les effets de la filiation et de la résidence pourraient être antagonistes.

Par ailleurs, la pratique de la polygynie peut également avoir un effet synergique ou antagoniste avec les autres composantes du système de parenté. La polygynie a été associée à la patrilocalité pour expliquer la réduction de la diversité du chromosome Y par rapport à l'ADN mitochondrial (Salem et al., 1996; Kayser et al., 2003). Mais les populations matrilocales peuvent aussi être

polygynes, et dans ce cas, nous pouvons nous attendre à un effet antagoniste entre la polygynie et la matrilocalité (Murdock et White, 1969). Ainsi, l'absence de consensus sur l'effet de la règle de résidence pourrait provenir du fait que ces études n'ont pas pris en compte le système de parenté dans sa totalité (Destro Bisol et al., 2012). Il nous semble donc important de prendre en compte conjointement la résidence et la filiation, mais aussi les autres aspects du système de parenté comme la polygynie pour mieux comprendre l'influence du système parenté sur la diversité génétique.

Par ailleurs, la plupart de ces études se sont intéressées au système de parenté en termes de règle (de résidence, de filiation…) et ont fondé leurs attendus sur cette règle et non sur le comportement démographique réel des individus au sein des populations étudiées. Or, la règle est fondée sur l'idéal selon les informateurs ou le comportement majoritaire des individus de la population étudiée (Murdock, 1949; Walker et al., 2013). Elle ne reflète donc pas forcément le comportement réel des individus. Ainsi, les études sur l'influence de la résidence sur la diversité génétique ont considéré les populations patrilocales et matrilocales comme des opposés symétriques de l'un par rapport à l'autre sur la base de leur règle de résidence : dans un cas les femmes migrent, dans l'autre ce sont les hommes qui migrent. Or, des asymétries peuvent exister entre populations matrilocales et patrilocales.

Ainsi, la règle de résidence peut ne pas être nécessairement respectée avec la même rigueur dans les populations patrilocales et matrilocales. Une flexibilité de la règle de résidence peut créer des différences au niveau des taux de migration d'hommes et de femmes par rapport à l'attendu. Comme nous l'avons vu, cette hypothèse a été proposée par Gunnarsdottir et al. (2011) pour expliquer l'absence de différence de diversité pour le chromosome Y entre populations patrilocales et matrilocales de Sumatra : les populations patrilocales pourraient suivre leur règle de résidence avec moins de rigueur que les populations matrilocales, et donc il n'y aurait pas de différences de migration d'hommes entre ces deux populations (Gunnarsdóttir et al., 2011). Un des processus qui pourraient engendrer une asymétrie entre populations patrilocales et matrilocales est ce que les anthropologues appellent le « puzzle matrilinéaire » (Richards, 1950; Schneider et Gough, 1961). Dans les sociétés matrilinéaires, la filiation est transmise en ligne maternelle. Cependant, dans la majorité des sociétés, même matrilinéaires, les hommes possèdent une autorité dans leurs familles natale et conjugale, mais ont aussi un rôle politique dans leur communauté. Ainsi, dans les populations matrilinéaires, « les hommes sont pris entre leur rôle de frère et de mari, entre leurs attaches et devoirs envers leur propre lignage, leurs sœurs et leurs neveux/nièces utérins, d'une part, et la « nécessité »

(sic) d'asseoir leur autorité sur l'épouse et les enfants dans l'unité conjugale, d'autre part » (Mathieu, 2007). C'est ce qui a été nommé le « puzzle matrilinéaire » (Richards, 1950; Schneider et Gough, 1961). L'existence de ce « puzzle matrilinéaire » pourrait engendrer des asymétries entre les systèmes de parenté unilinéaires harmoniques (patrilinéaire et patrilocale, matrilinéaire et matrilocale).

En effet, dans les populations matrilinéaires et matrilocales, lors du mariage, l'homme s'installe avec la famille de sa femme. Or, un certain nombre d'hommes pourraient rester dans leur village natal en se mariant patrilocalement afin d'exercer les fonctions d'autorité au sein de leur groupe de filiation natal. Il n'existe pas de raison similaire chez les populations patrilinéaires et patrilocales pouvant augmenter la proportion de couples matrilocaux. Nous nous attendons alors à avoir une flexibilité de la règle de résidence plus importante chez les populations matrilinéaires et matrilocales que les populations patrilinéaires et patrilocales. Une telle différence de flexibilité, plus importante chez les populations matrilocales que les populations patrilocales, a été observée en Thaïlande où le ratio du nombre d'hommes sur le nombre de femmes migrants dans les populations matrilocales est de 1,3 comparé au ratio du nombre de femmes sur le nombre de hommes migrants dans les populations patrilocales qui est de 15 (Hamilton et al., 2005). Les auteurs cependant ne se réfèrent pas au « puzzle matrilinéaire » pour expliquer cette asymétrie mais à une régulation plus stricte des migrations dans les populations patrilocales.

Par ailleurs, les hommes dans les populations matrilinéaires et matrilocales peuvent aussi se marier localement (e.g. au sein de leur village natal) afin de rester à proximité de leur groupe de filiation et pouvoir exercer leur fonction d'autorité (Kloos, 1963; Walker et al., 2013). Le « puzzle matrilinéaire » peut ainsi expliquer la forte association de l'endogamie de village avec le système de parenté matrilinéaire et matrilocale (Murdock, 1949; Johnston et al., 1969). Parmi les 25 sociétés matrilinéaires et matrilocales présentes dans « Social Structure » de Murdock, seulement dans trois sociétés les hommes se marient fréquemment en dehors de leur village natal. Ainsi, nous nous attendons à avoir un taux d'endogamie de village plus élevé chez les populations matrilinéaires et matrilocales et les populations patrilinéaires et patrilocales.

Enfin, dans les populations matrilinéaires, le mari partage l'autorité avec son beau-frère. Ce partage de l'autorité peut engendrer des frictions et nous nous attendons à avoir un taux de divorce plus important dans les populations matrilinéaires que les populations patrilinéaires (Gluckman, 1950; van den Berghe et Barash, 1977; Sear, 2008).

Il nous apparaît donc essentiel de ne pas considérer les systèmes matrilocaux et patrilocaux comme des systèmes miroirs l'un de l'autre mais de passer par une exploration ethno-démographique quantitative de ces systèmes afin de mettre en lumière ces éventuelles asymétries et ainsi de mieux comprendre l'impact des structures sociales sur les structures génétiques.

Enfin, il apparait également important d'étendre ce champ d'études à la diversité autosomale et notamment de caractériser l'impact du système de parenté sur la consanguinité. Des études ont montré qu'il existait des populations dans le monde où le taux de consanguinité était élevé (McQuillan et al., 2008; Leutenegger et al., 2011; Pemberton et al., 2012a; Gazal et al., 2015). Cette plus forte consanguinité a été expliquée par l'existence de préférence pour des mariages entre individus apparentés dans ces populations. Ainsi, les taux de consanguinité sont élevés au Moyen-Orient ou en Asie du Sud, où les mariages entre cousins sont fréquents (Bittles et Black, 2010; Bittles et Hamamy, 2010). Cependant, les études s'intéressant spécifiquement à la question de l'influence du système de parenté sur la consanguinité restent rares (Watkins, 2004; Billari et al., 2007; Guillot et al., 2015). Il nous semble donc pertinent de développer le champ disciplinaire dans cette direction. Notamment, la préférence de mariages entre cousins de premier degré est associée à la règle de filiation unilinéaire (patrilinéaire ou matrilinéaire). En effet, dans les populations mondiales présentes dans le SCCS (Murdock et White, 1969), 41% des populations unilinéaires ont une préférence pour les mariages avec une ou plusieurs des cousines, comparé à 20% pour les populations cognatiques. Ainsi, nous pouvons nous attendre à avoir des taux de consanguinité plus élevés chez les populations unilinéaires que chez les populations cognatiques.

3. Objectif de la thèse

Dans ce contexte, l'objectif de cette thèse est de revisiter la question de l'impact du système de parenté sur la diversité génétique uniparentale et l'étendre à la diversité autosomale par une approche globale des systèmes de parenté prenant en compte ses différentes composantes. Par ailleurs, nous nous ne limiterons pas à une catégorisation des populations selon les règles mais nous nous intéresserons aux comportements ethno-démographiques des individus.

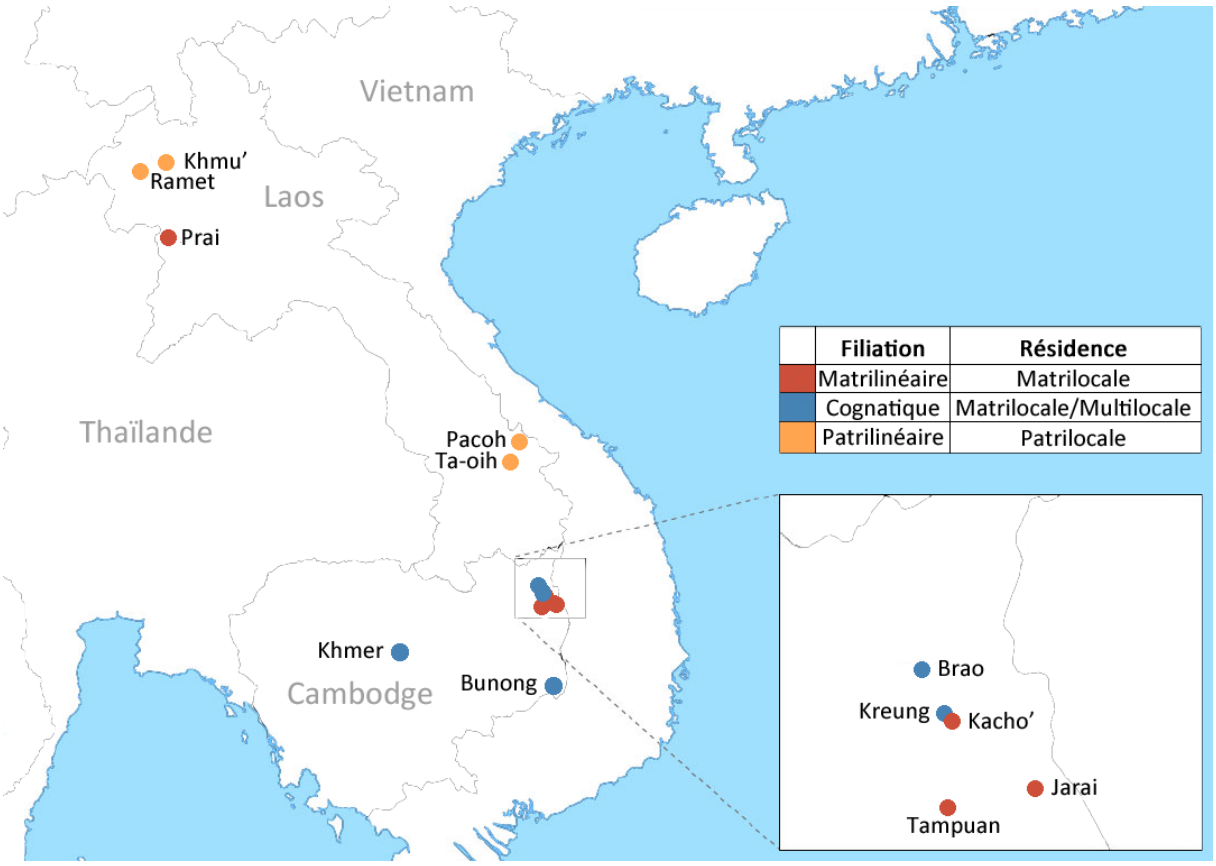


Figure1 : Carte de la région d'étude. Les populations sont regroupées par règle de filiation et résidence.

Nous proposons d'explorer la variabilité ethno-démographique et génétique de 12 populations d'Asie du Sud-Est (figure 1) présentant des règles de filiation, de résidence et d'alliance variées. Cette région du monde est particulièrement pertinente pour cette étude car des populations présentant une grande variabilité de systèmes de parenté mais des modes de vie similaires dans un environnement relativement homogène y résident (table 1). En particulier, nous avons pu étudier des populations matrilinéaires, une règle de filiation très peu étudiée en génétique des populations humaines de par leur rareté à l'échelle mondiale : 12 à 17% des sociétés sont matrilinéaires (Levinson et Malone, 1980; Godelier, 2004). Parmi elles, 58% sont dites harmoniques, c'est-à-dire présentant une résidence matrilocale (Mathieu, 2007). Nous avons

ainsi étudié 4 populations matrilinéaires et matrilocales : les Tampuan, les Jarai, les Prai et les Kacho'. L'échantillonnage inclut également 4 populations patrilinéaires (résidence patrilocale) : les Khmu', les Ramet, les Ta-oih et les Pacoh. La comparaison de ces systèmes matrilinéaires et patrilinéaires va donc nous permettre d'évaluer les potentielles asymétries entre ces systèmes et notamment de tester l'hypothèse du « puzzle matrilinéaire » évoquée ci-dessus, qui prédit au niveau ethno-démographique une plus grande flexibilité de résidence ainsi qu'une endogamie de village plus élevée chez les matrilinéaires, et donc d'éventuelles répercussions sur les patrons génétiques uniparentaux et autosomaux.

Système de parenté	Ethnie	Résidence	Filiation	Alliance	Références
M	Tampuan	Matrilocale	Matrilinéaire	Préférence pour la cousine croisée matrilatérale	(Bourdier, 2006)
	Jarai	Matrilocale	Matrilinéaire	Préférence pour les cousines croisées bilatérales	(LeBar et al., 1964; Dournes, 1972)
	Prai	Matrilocale	Matrilinéaire	Préférence pour les cousines de premier degré	(LeBar et al., 1964; Dessaint, 1981)
	Kacho'	Matrilocale	Matrilinéaire		
C	Bunong	Matrilocale	Cognatique		(UNDP Cambodia, 2010a)
	Khmer	Matrilocale	Cognatique		(Martel, 1975; Ebihara, 1977; Ledgerwood, 1995)
	Brao	Multilocale à prédominante matrilocale	Cognatique	Pas de règle prescriptive	(LeBar et al., 1964; Matras-Troubetzkoy, 1983; Baird, 2008)
	Kreung	Multilocale à prédominante matrilocale	Cognatique	Pas de règle prescriptive	(UNDP Cambodia, 2010b)
P	Khmu'	Patrilocale	Patrilinéaire	Préférence pour la cousine croisée matrilatérale	(LeBar et al., 1964; Evrard, 2006)
	Ramet	Patrilocale	Patrilinéaire	Préférence pour la cousine croisée matrilatérale	(Izikowitz, 1951; LeBar et al., 1964)
	Ta-oih	Patrilocale	Patrilinéaire	Pas de règle prescriptive	(Schmutz, 2013)
	Pacoh	Patrilocale	Patrilinéaire		

Table 1 : Présentation des populations étudiées. Nous avons noté la règle de résidence, de filiation et éventuellement d'alliance décrit dans la littérature ethnographique.

L'échantillonnage inclut enfin des populations cognatiques de résidence matrilocale prédominante : les Bunong, les Khmer, les Brao et les Kreung. Ces systèmes devraient nous permettre, par comparaison avec les systèmes matrilinéaires matrilocaux, de départager l'impact de la résidence et de la filiation sur la diversité génétique uniparentale. En effet, dans

le cas d'une population matrilinéaire et matrilocale, nous nous attendons à avoir un effet synergique de la résidence et de la filiation, ce qui pourrait impacter les patrons génétiques plus fortement que dans le cas d'une population cognatique et matrilocale dans laquelle, seul la résidence impacterait les patrons génétiques.

De plus, l'inclusion de populations cognatiques nous permet de comparer les patrons de consanguinité entre populations unilinéaires et cognatiques. Comme nous l'avons expliqué précédemment, nous nous attendons à observer une préférence plus marquée pour les mariages entre cousins chez les populations unilinéaires par rapport aux populations cognatiques, ce qui aurait alors pour conséquence une consanguinité plus élevée chez les populations unilinéaires.

Les populations étudiées ont été choisies sur la base de l'existence d'une riche littérature ethnographique sur la majorité d'entre elles, ce qui nous a permis d'affecter des règles de filiation, résidence et alliance aux populations et qui permettra de discuter nos résultats à la lumière de nos observations ethnographiques quantitatives (Izikowitz, 1951; LeBar et al., 1964; Dournes, 1972; Martel, 1975; Ebihara, 1977; Dessaint, 1981; Matras-Troubetzkoy, 1983; Ledgerwood, 1995; Bourdier, 2006; Evrard, 2006; Baird, 2008; UNDP Cambodia, 2010a; b; Schmutz, 2013). Par ailleurs, ces populations ont été très peu étudiées génétiquement (Zhang et al., 2013; Kutanan et al., 2017). L'échantillonnage de ces populations a été effectué dans le cadre de l'ANR SoGen pilotée par Raphaëlle Chaix dans 61 villages lors de trois missions interdisciplinaires effectuées entre 2011 et 2012, et auxquelles je n'ai pas participé. Cependant, j'ai eu l'opportunité en décembre 2015 de rencontrer les populations à l'occasion d'une mission de terrain de restitution des résultats du projet SoGen aux individus de 5 populations cambodgiennes ayant participé à l'étude.

Le premier chapitre sera consacré à la comparaison des variables ethno-démographiques entre ces populations. Dans les chapitres suivants, nous analyserons les données génétiques en lumière des observations ethno-démographiques du premier chapitre. Ainsi, dans le second chapitre, nous testerons si les différences observées entre systèmes de parenté pour les variables ethno-démographiques sont reflétées sur la diversité uniparentale. Dans le troisième et dernier chapitre, nous nous focaliserons sur l'influence du système de parenté sur la diversité génétique autosomale.

Chapitre 1 - Des systèmes de parenté aux variables ethno-démographiques

1. Introduction

Afin de caractériser l'impact des systèmes de parenté sur la diversité génétique, il faut dans un premier temps étudier l'impact de ces systèmes de parenté sur les variables ethno-démographiques d'importance pour l'évolution de la diversité génétique.

Ainsi, nous allons dans ce premier chapitre, nous demander si nous pouvons retrouver la catégorisation ethnographique des systèmes de parenté sur la base de variables ethno-démographiques quantitative, et si c'est le cas, identifier les variables permettant cette séparation entre systèmes de parenté. Nous allons dans un second temps vérifier nos hypothèses ethnologiques proposées en introduction sur les différences entre systèmes de parenté en termes de variables ethno-démographiques.

Du fait de l'existence du « puzzle matrilinéaire » chez les populations matrilinéaires et matrilocales, nous nous attendons à observer une flexibilité de la règle de résidence, un taux d'endogamie de village ainsi qu'un taux de divorce plus important chez les populations matrilinéaires et matrilocales par rapport aux populations patrilinéaires et patrilocales. Par ailleurs, si les populations d'Asie du Sud-Est que nous étudions suivent le patron mondial, nous nous attendons à avoir une préférence de mariages entre cousins de premier degré plus élevée chez les populations unilinéaires que les populations cognatiques.

La dynamique de filiation et le taux de polygynie sont des aspects du système de parenté pouvant avoir une influence sur la diversité génétique. Nous n'avons en revanche pas d'a priori sur les possibles différences entre systèmes de parenté.

Ainsi, nous allons tester :

- S'il existe une flexibilité de la règle de résidence, un taux d'endogamie de village et un taux de divorce plus importants chez les populations matrilinéaires du fait du « puzzle matrilinéaire ».

- S'il existe une proportion de mariages entre cousins de premier degré plus importante dans les populations unilinéaires que dans les populations cognatiques.
- S'il existe des dynamiques de groupe de filiation différentes entre les systèmes matrilinéaires et patrilinéaires.
- S'il existe un taux de polygynie différent entre les systèmes de parenté.

2. Matériels et méthodes

2.1 Echantillonnage des populations

Douze populations ont été échantillonnées au Laos et au Cambodge présentant des règles de filiation variées (matrilinéaire, patrilinéaire et cognatique) combinées à des règles de résidence différentes (matrilocale, patrilocale, multilocale à dominante matrilocale). Les Tampuan, les Jarai, les Prai et les Kacho' sont des populations matrilinéaires et matrilocales (LeBar et al., 1964; Dournes, 1972; Dessaint, 1981; Bourdier, 2006). Les Khmer et les Bunong sont des populations cognatiques et matrilocales (Martel, 1975; Ebihara, 1977; Ledgerwood, 1995; UNDP Cambodia, 2010a). Les Brao et les Kreung sont des populations cognatiques et multilocales à dominante matrilocale (LeBar et al., 1964; Matras-Troubetzkoy, 1983; Baird, 2008; UNDP Cambodia, 2010b). Enfin les Khmu', les Ramet, les Ta-oih et les Pacoh sont des populations patrilinéaires et patrilocales (Izikowitz, 1951; LeBar et al., 1964; Evrard, 2006; Schmutz, 2013).

Nous avons regroupé les populations en trois catégories selon leur système de parenté (voir table 1 de l'introduction) : les populations avec une filiation matrilinéaire et résidence matrilocale (M), les populations avec une filiation cognatique et une résidence matrilocale ou multilocale à dominance matrilocale (C) et les populations avec une filiation patrilinéaire et une résidence patrilocale (P).

Il est intéressant de noter que, lors des enquêtes de terrain, les Kacho' et les Prai se sont démarqués des autres populations. Les membres de l'équipe se sont rendu compte sur le terrain que les Kacho' censés être matrilinéaires présentaient un système de filiation en transition entre la matrilinéarité et la patrilinéarité : certaines familles respectent la filiation matrilinéaire alors que chez d'autres la filiation passe du père aux enfants comme dans un système patrilinéaire. Les Prai quant à eux constitue une population montagnarde dont les villages peuvent être très isolés et difficiles d'accès. Leur isolation limite les échanges avec les populations avoisinantes mais aussi entre les villages Prai.

2.2 Exploitation des questionnaires

Lors des trois missions de terrain effectuées entre 2011 et 2012, les membres de l'équipe ont effectué 532 entretiens dans 61 villages avec des individus mariés afin de collecter des informations sur eux et les membres de leur famille. Un questionnaire ethno-démographique (annexe 1) a été utilisé pour la récolte des informations. Les entretiens ont été effectués en

présence du conjoint de l'informateur lorsque cela était possible. Les questions étaient posées par l'intermédiaire d'un interprète. Je n'ai pas participé à la phase de récolte de données présentée ici.

L'informateur ou le couple a tout d'abord fourni des informations sur eux-mêmes, inscrit sur la première page du questionnaire (annexe 1A) : village de naissance, village de résidence (qui est le village dans lequel l'entretien a été effectué), langue maternelle, âge, et, dans les populations unilinéaires, groupes de filiation (clan et éventuellement lignage). L'histoire reproductive de l'informateur a ensuite été renseignée (annexe 1B) : nombre de conjoints qu'il a eu, nombre de divorces, nombre de remariages en cas de veuvage, si l'informateur est dans une relation polygame, nombre total d'enfants, nombre d'enfants décédés, nombre de filles et de fils mariés, divorcés, remariés en cas de veuvage, nombre d'enfants mariés décédés, et enfin nombre d'enfants adoptés et donnés en adoption. Pour certains entretiens, l'histoire reproductive d'un des parents de l'informateur a aussi été renseigné (annexe 1D)

L'informateur a renseigné sur la raison de son mariage (mariage arrangé par la famille ou non), ainsi que sur le lien généalogique avec son conjoint s'ils étaient apparentés. Pour les entretiens au Cambodge, l'informateur indiquait si le mariage a été décidé par le régime Khmer Rouge. En effet pendant le régime Khmer Rouge, les populations ont été déplacées et il y a eu de nombreux mariages forcés par le régime au sein de ces populations. Ces couples ne sont pas représentatifs du système de parenté de la population, et, dans le cas d'une réponse affirmative, le couple est retiré des analyses.

Ensuite, les informateurs ont fourni des informations sur leurs ascendants : leurs parents et leurs grands-parents. Les informations collectées sont : village de naissance, village de résidence, langue maternelle et, dans les populations unilinéaires, groupes de filiation (clan et éventuellement lignage). L'information sur le village de résidence de ces individus pouvait parfois ne pas être renseignée ou difficile à renseigner en absence de liste administrative avec une appellation et une orthographe normalisées des villages. Dans ce cas-là, nous utilisons l'information sur le village de naissance de leurs enfants comme substitut. Par exemple, le lieu de naissance de l'informateur a été utilisé comme substitut pour le village de résidence de ses parents, le lieu de naissance du père de l'informateur pour le village de résidence des grands-parents paternels d'Ego.

L'informateur a ensuite fourni des informations sur ses affins, c'est-à-dire les conjoints de ses frères et sœurs, et ceux de ses enfants (annexe 1C et 1E). Les informations recueillies sont les

mêmes que pour sa famille : village de naissance, village de résidence, langue maternelle et groupes de filiation. Dans ce cas, lorsque l'information sur le lieu de résidence n'était pas disponible, nous n'avions pas de moyen d'inférer l'information manquante.

Ces informations ont été renseignées à l'échelle de l'individu. Afin d'estimer les variables d'intérêt tel que le taux d'endogamie de village qui combine les informations sur les deux conjoints d'un couple (il nous faut savoir si les deux conjoints sont nés dans le même village ou non), nous avons créé un jeu de données dans lequel les informations sont organisées par couple : l'informateur et son conjoint, le père de l'informateur et la mère de l'informateur, etc.

Une difficulté survenue au cours de la création de ce nouveau jeu de données est que le questionnaire a été conçu de telle façon que les informateurs ne nous ont pas renseignés sur le lieu de naissance et les groupes de filiation de leurs frères, sœurs et enfants mais uniquement sur les conjoints de ceux-ci. Pour pouvoir créer les couples des frères, sœurs et enfants de l'informateur, il nous manque donc l'information sur la moitié des individus. Pour résoudre ce problème, nous avons à nouveau inféré les informations pour ces individus à partir des informations de la famille de l'informateur. Pour les frères et sœurs, nous avons supposé que le lieu de naissance et les groupes de filiation étaient identiques à ceux de l'informateur. Pour les enfants, nous avons utilisé le lieu de résidence de l'informateur comme lieu de naissance et pour les groupes de filiation, nous avons utilisé la règle de filiation en vigueur dans la population pour attribuer le clan (celui du père dans les populations patrilinéaires, celui de la mère pour les populations matrilinéaires).

Nous avons réussi par ce biais à obtenir à partir des 532 entretiens des informations ethno-démographiques pour 3530 couples qui seront utilisées pour estimer les différentes variables d'intérêt. Nous avons par ailleurs comparé les résultats obtenus sur ces 3530 couples avec un jeu de données restreint fondé sur les informations obtenues seulement sur les couples des informateurs (532 couples). Les résultats étant dans la plupart des cas très similaires, et les conclusions que nous pouvons en tirer identiques, nous ne présentons que les résultats obtenus sur les 3530 couples.

2.3 Estimation des variables ethno-démographiques d'intérêt pour l'évolution génétique des populations

2.3.1 Variables de résidence

Pour chaque couple, nous possédons leur village de résidence et le village de naissance de chaque conjoint. Nos variables de résidence seront donc toutes définies à l'échelle des villages et nous ne nous intéresserons pas, par manque d'informations, sur le comportement des individus au sein des villages. Par le passé pourtant, les villages étaient structurés géographiquement en maison longue (*long house*) regroupant plusieurs familles notamment par lignage ou clan dans les populations unilinéaires. Cependant, les maisons actuelles dans les villages sont du style « Khmer » avec une famille nucléaire par maison, rendant l'analyse de la résidence post-maritale intra-village difficile.

En utilisant la typologie proposé par Casselberry et Valavanes, nous pouvons identifier 4 types de résidence à l'échelle des villages (1976). Un couple peut être :

- Natolocal, si les conjoints sont nés dans le même village et vivent dans leur village de naissance (cas A Figure 1-1).
- Patrilocal, si les conjoints sont nés dans des villages différents et le couple vit dans le village de naissance du mari (cas B Figure 1-1).
- Matrilocal, si les conjoints sont nés dans des villages différents et le couple vit dans le village de naissance de la femme (cas C Figure 1-1).
- Néolocal, si le couple vit dans un village qui est ni le village de naissance du mari ni celui de la femme. Dans ce dernier cas, les conjoints peuvent être nés dans le même village ou dans des villages différents (cas D Figure 1-1 représente le cas où les conjoints sont nés dans 2 villages différents).

Ces 4 types de résidence sont mutuellement exclusifs, un couple se retrouve nécessairement dans un seul de ces 4 types de résidence. Une fois les couples assignés à un type de résidence, nous avons estimé la proportion de couples pour chaque type de résidence dans les différentes populations.

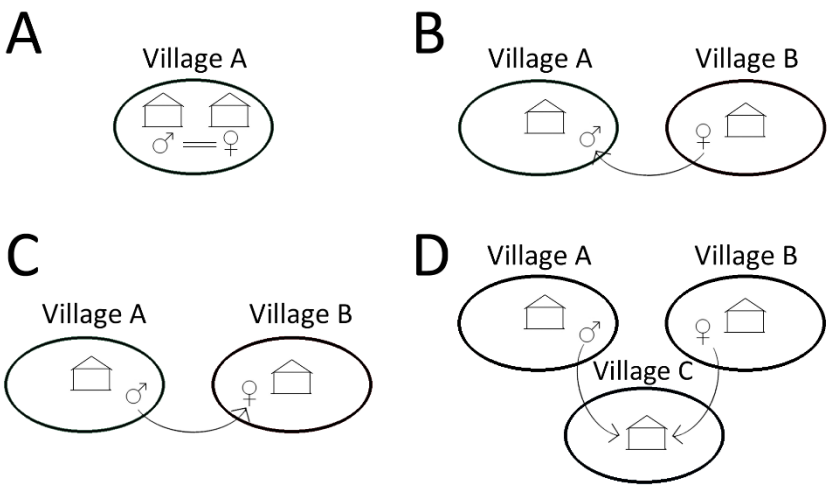


Figure 1-1 : Représentation des 4 types de résidence à l'échelle du village : (A) Résidence natolocale, (B) Résidence patrilocale, (C) Résidence matrilocale, (D) Résidence néolocale.

Nous avons ensuite utilisé ces proportions de résidence pour estimer la flexibilité de la règle de résidence ainsi que les taux de migration d'hommes et de femmes entre villages. Nous avons défini deux estimateurs de la flexibilité de la règle de résidence : le taux de flexibilité restreinte et le taux de flexibilité élargie de la règle de résidence. Le taux de flexibilité restreinte est défini comme le ratio de la proportion de couples avec une résidence opposée à la règle de résidence de la population et de la somme des proportions de couples matrilocaux et patrilocaux. Ainsi, pour les populations patrilocales, le taux de flexibilité restreinte de la règle de résidence s'écrit sous la forme :

$$\frac{Proportion\ de\ couples\ matrilocaux}{Proportion\ de\ couples\ patrilocaux + Proportion\ de\ couples\ matrilocaux}$$

Pour les populations matrilocales, elle s'écrit sous la forme :

$$\frac{Proportion\ de\ couples\ patrilocaux}{Proportion\ de\ couples\ patrilocaux + Proportion\ de\ couples\ matrilocaux}$$

La flexibilité élargie de la résidence incorpore en plus la proportion de couples néolocaux. La néolocalité étant une forme de résidence dans laquelle les deux conjoints s'installent dans un village qui n'est pas leur village natal, nous pouvons considérer cela comme une forme de flexibilité. Ainsi, pour les populations patrilocales, la flexibilité élargie de la règle de résidence s'écrit sous la forme :

$$\frac{Prop\ couples\ matrilocaux + Prop\ couples\ néolocaux}{Prop\ couples\ patrilocaux + Prop\ couples\ matrilocaux + Prop\ couples\ néolocaux}$$

Pour les populations matrilocales, elle s'écrit sous la forme :

$$\frac{Prop\ couples\ patrilocaux\ + Prop\ couples\ néolocaux}{Prop\ couples\ patrilocaux + Prop\ couples\ matrilocaux + Prop\ couples\ néolocaux}$$

Les taux de migration d'hommes et de femmes ont été calculés respectivement comme la somme de la proportion de couples matrilocaux et néolocaux et la somme de la proportion de couples patrilocaux et néolocaux.

Le taux d'endogamie de village a été pour sa part estimé comme la proportion de couples pour lesquels les deux conjoints sont nés dans le même village.

2.3.2 Variables de filiation

Pour caractériser la dynamique des groupes de filiation, nous nous sommes intéressés à trois estimateurs ethno-démographiques : le taux de connaissance des clans des ascendants, le nombre de clans et le taux d'exogamie de clan.

Le taux de connaissance du clan des ascendants a été uniquement calculé pour les conjoints présents lors de l'entretien. Nous avons pris avantage de notre méthode de collecte de données qui demandait à un informateur de fournir les informations sur ses parents et grands-parents. Nous avons estimé à quel point les informateurs connaissaient le clan de leur parents et grands-parents pour la branche paternelle et maternelle. La branche paternelle comprend le père et le grand-père paternel, et la branche maternelle, la mère et de la grand-mère maternelle. Nous avons considéré qu'un individu connaissait le clan de sa branche paternelle ou maternelle s'il nous avait informés du clan de ses deux ascendants (parent et grand-parent).

Nous nous sommes intéressés à cette variable comme un estimateur de l'importance des groupes de filiation comme unité structurante dans les populations. Le fait que les individus connaissent mal les groupes de filiation de leurs ascendants pourrait être un signe que les individus ne portent que peu d'importance à l'appartenance à ces groupes de filiation et donc que les groupes de filiation soient perméables et ne structurent pas la population. Nous nous attendons à avoir une meilleure connaissance des clans paternels dans les populations

patrilinéaires et une meilleure connaissance des clans maternels dans les populations matrilinéaires car ce sont les branches des ascendants transmettant leur clan. Pour comparer la connaissance des clans des ascendants entre les populations patrilinéaires et matrilinéaires, il faut alors comparer le taux de connaissance des clans de la branche paternelle des premiers au taux de connaissance des clans de la branche maternelle des seconds. Nous avons appelé ce taux, le taux de connaissance des clans de la branche principale.

La seconde variable que nous avons estimée est le nombre de clans dans les populations. Nous avons simplement dénombré le nombre de clans différents présents dans les questionnaires. Nous avons supposé qu'un grand nombre de clans présents dans une population témoigne de l'existence d'une dynamique de fission dans cette population particulièrement active. Il est très difficile de maintenir un grand nombre de clans dans une population sans un apport constant de nouveaux clans (voir annexe 2). Les fissions de clan en deux ou plusieurs nouveaux clans permettraient l'existence d'un grand nombre de clans. Si ces fissions de clans se font de façon linéaire, nous nous attendons à avoir un impact sur la diversité uniparentale (Chaix et al., 2007).

Enfin, nous avons aussi estimé le taux d'exogamie de clan. Un couple est considéré comme exogame de clan si les deux conjoints appartiennent à des clans différents. Les Khmu' et les Ramet sont organisés non seulement en clans mais en lignages au sein des clans. Ces lignages sont propres à chaque village et serait l'unité d'échange exogame dans ces populations (Evrard, 2006; Sprenger, 2008). Au contraire, les clans sont l'unité exogame chez les populations M (Dournes, 1972). Ainsi, nous nous attendons à voir un taux d'exogamie de clan plus faible chez les Khmu et les Ramet par rapport aux populations M.

2.3.3 Variables d'alliance

Comme nous l'avons présenté dans la section 2.2, nous possédons très peu d'informations concernant les choix matrimoniaux effectués par les informateurs. Nous savons si l'union a été arrangée par la famille et s'il existe un lien de parenté entres les conjoints. L'obtention de l'information sur le lien de parenté a été compliquée et très souvent incomplète. Retracer les liens généalogiques unissant les deux conjoints est un travail chronophage qui n'a pas pu être accompli systématiquement lors des missions de terrain. Nous n'avons donc pas pu estimer un taux de mariages entre cousins de premier degré à partir de notre jeu de données ethno-démographiques.

En revanche, grâce aux informations collectées sur l'histoire reproductive de l'informateur, et de l'un de ces parents, nous avons pu estimer un taux de divorce, et un taux de polygynie dans les populations. Le taux de divorce a été estimé sur tous les individus pour lesquels nous possédions l'information sans séparation par sexe. Nos données comprennent 1035 individus pour lesquels nous possédons l'histoire reproductive, dont 778 hommes sur lesquels le taux de polygynie a été estimé. Nous n'avons pas observé de cas de polyandrie dans nos données.

2.4 Comparaison des systèmes de parenté

2.4.1 Analyse en composantes principales à l'échelle des populations

Afin de vérifier s'il est possible de distinguer les populations selon leur système de parenté sur la base des variables ethno-démographiques, nous avons tout d'abord effectué une analyse en composantes principales (ACP) sur les variables ethno-démographiques d'intérêt. Il s'agit d'une projection des distances entre les populations pour ces variables sur un plan. La projection maximise la distance entre les points. Ainsi, les populations proches sur cette projection seraient plus proches en termes de variables ethno-démographiques. Nous nous attendons à ce que les populations avec le même système de parenté aient les mêmes caractéristiques ethno-démographiques et donc soient proches sur la projection. Nous pouvons aussi analyser la projection des variables sur le plan, une analyse appelée '*variables factor map*'. Celle-ci permet de déterminer les variables corrélées aux composantes principales (les axes du plan) et donc celle qui permettent de distinguer les systèmes de parenté. Les variables dont le vecteur est à un faible angle par rapport à un axe et avec une longueur proche de 1 sont les variables fortement corrélées aux composantes principales. Cette analyse a été effectuée en utilisant la fonction 'PCA' de la librairie 'FactoMineR' v1.31.3 dans le logiciel R v3.2.2 (R Development Core Team, 2008).

Une première ACP a été effectuée avec les 12 populations sur les variables de résidence et le taux de divorce uniquement. Les variables de résidence utilisées sont : les proportions de couples natolocaux, néolocaux, patrilocaux et matrilocaux, le taux de migration d'hommes et de femmes et le taux d'endogamie de village. Nous ne pouvons pas inclure les variables de filiation sur une ACP avec les 12 populations, car les populations cognatiques n'ont pas d'estimations pour ces variables. Le taux de polygynie a également été exclu car il s'est avéré très peu informatif. Nous avons effectué une seconde ACP uniquement sur les 8 populations unilinéaires intégrant en plus des variables précédentes, les variables de filiation suivantes : le

taux de connaissance du clan de la branche paternelle, de la branche maternelle, le nombre de clan et le taux d'exogamie de clan.

2.4.2 Matrice de corrélation

Nous avons établi la matrice de corrélation entre variables pour voir s'il existait des groupes de variables corrélées. Les mêmes variables que pour les ACP ont été utilisées pour les matrices de corrélation sur les 12 populations et sur les 8 populations unilinéaires. Une des questions sous-jacentes est de tester si les variables de résidence et de filiation sont corrélées les unes aux autres. Comme nous avons des systèmes de parenté harmoniques, nous pouvons nous attendre à avoir une corrélation entre les variables de résidence (patrilocalité, matrilocalité) et les variables de filiation (connaissance des clans paternels, connaissance des clans maternels). Le coefficient de Spearman a été utilisé pour estimer cette corrélation. La matrice a été produite par la fonction 'corrplot' de la librairie 'corrplot' v0.77 dans le logiciel R.

2.4.3 Comparaison des variables ethno-démographiques d'intérêt entre systèmes de parenté

Enfin, nous avons testé s'il existait des différences entre systèmes de parenté pour les variables ethno-démographiques d'intérêt pour l'évolution de la diversité génétique. Pour cela nous avons analysé l'influence du système de parenté sur chacune de ces variables par un modèle linéaire mixte généralisé (GLMM ou *generalized linear mixed model*). La fonction 'glmer' de la librairie 'lme4' v1.1-9 dans le logiciel R a été utilisée.

Les variables d'intérêt étudiées par GLMM sont :

- pour la résidence : natolocalité, néolocalité, patrilocalité, matrilocalité, migration d'hommes, migration de femmes, endogamie de village et flexibilité restreinte et élargie de la règle de résidence.

- pour la filiation : connaissance du clan de la branche maternelle, de la branche paternelle, de la branche principale (i.e. maternelle chez les populations matrilinéaires et paternelle chez les populations patrilinéaires) et exogamie de clan.

- pour l'alliance : divorce et polygynie.

Chaque variable d'intérêt a été codée de façon binaire (trait présent, trait absent) pour chaque couple. Le modèle inclut en effet fixe deux variables catégorielles : le système de parenté (trois catégories) et la population (12 catégories). Le village d'échantillonnage et la famille ont été inclus comme effet aléatoire dans le modèle. Les trois variables, population, village et famille ont été ajoutées au modèle afin de contrôler les biais d'échantillonnage possibles du fait de la structuration possible de nos données à ces trois différentes échelles. La significativité de l'influence du système de parenté sur chacune des variables a été évaluée en comparant les moyennes par système de parenté après avoir contrôlé pour les facteurs confondants (population, village et famille) avec la fonction 'lsmeans' dans la librairie 'lsmeans' v2.27-2 dans le logiciel R. Cette méthode permet de prendre en compte la structure nichée de nos données : les populations sont nichées au sein des systèmes de parenté.

Le nombre de clans a été comparé entre systèmes parenté par un test de Mann-Whitney.

3. Résultats

3.1 Analyse en composantes principales et matrice de corrélation

L'ACP effectuée sur les 12 populations avec les variables de résidence (Figure 1-2), révèle une grande variabilité au sein des populations présentant le même système de parenté, notamment au sein des populations P. Nous pouvons en revanche observer que le premier axe, qui explique 45% de la variance entre les populations, sépare nettement les populations M des populations P. Toutes les populations M ont des valeurs négatives, alors que les populations P ont des valeurs positives mis à part les Khmu qui ont une valeur proche de 0 sur cet axe. Par ailleurs, mis à part les Kacho', les populations M sont proches les unes des autres, montrant que ces populations sont généralement similaires en termes de variables ethno-démographiques.

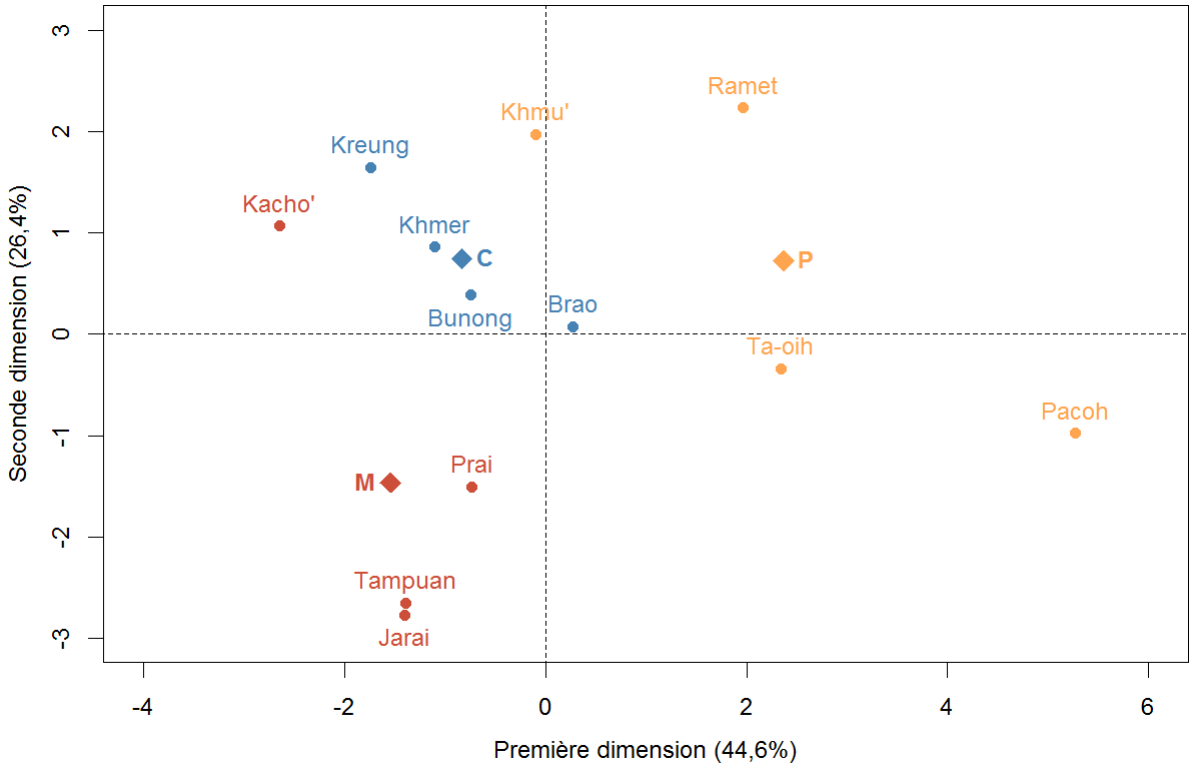


Figure 1-2 : Analyse en composantes principales des variables ethno-démographiques effectuée sur les 12 populations. Les variables de résidence et le divorce ont été pris en compte. Les populations M sont en rouge, les populations C en bleu et les populations P en jaune. Les losanges représentent les barycentres des populations avec le même système de parenté.

Nous avons mentionné dans la partie 2.1 de ce chapitre que les Kacho' sont une population particulière en transition d'un système matrilinéaire à un système patrilinéaire. Il n'est donc pas surprenant qu'elle ne se groupe pas avec les autres populations matrilinéaires. Les populations C se retrouvent entre les populations M et P et sont plutôt regroupées entre elles. Nous voyons donc à la fois, une variabilité importante au sein des systèmes de parenté mais aussi que les

populations avec le même système de parenté possèdent des similarités en termes de variables de résidence.

Lorsque nous incluons les variables de filiation (Figure 1-3), les observations restent similaires à celles de l'ACP sur les variables de résidence. Nous observons, une séparation entre les populations P et M sur le premier axe, qui explique 47% de la variance. Il existe à nouveau une variabilité importante chez les populations P, et un regroupement des populations M à l'exception des Kacho'.

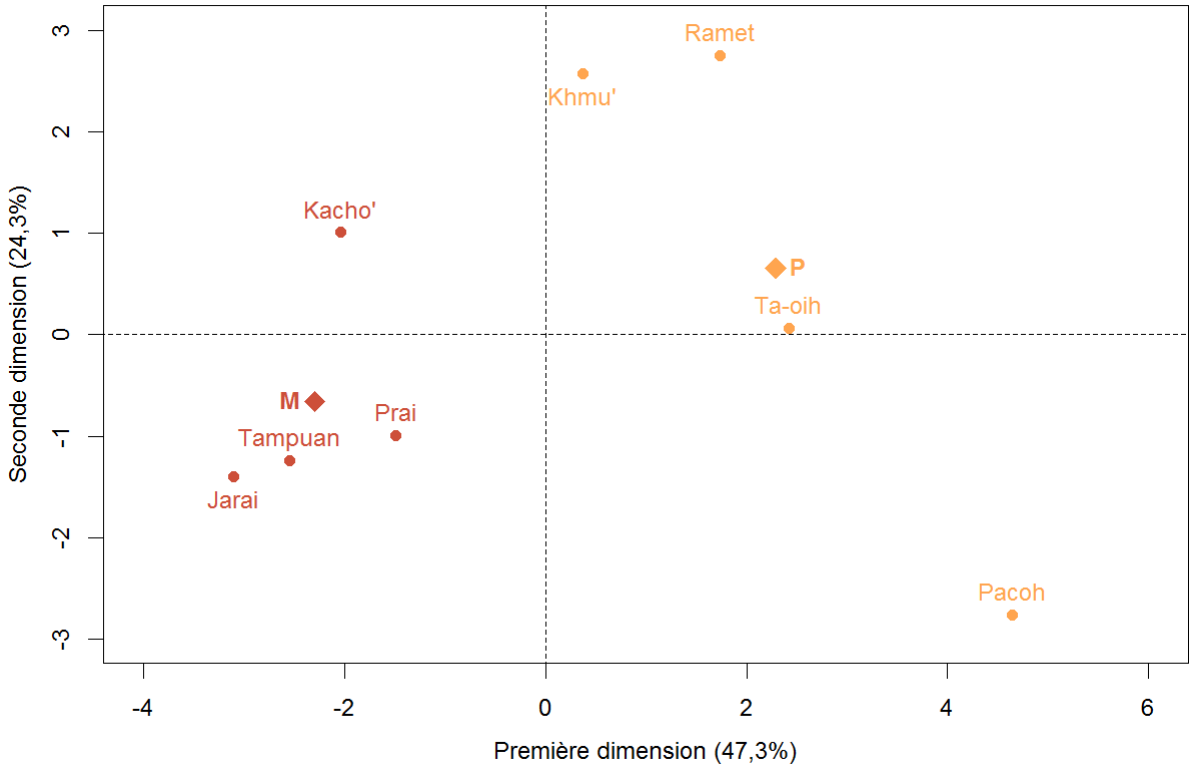


Figure 1-3 : Analyse en composantes principales des variables ethno-démographiques effectuée sur les 8 populations unilinéaires. Les variables de résidence, de filiation et le divorce ont été pris en compte. Les populations M sont en rouge et les populations P en jaune. Les losanges représentent les barycentres des populations avec le même système de parenté.

Nous pouvons donc à partir de ces observations supposer que certaines variables possèdent des valeurs différentes entre les populations avec des systèmes de parenté différents, à la fois en termes de résidence et de filiation. Le système de parenté aurait donc un impact sur les variables d'intérêt pour l'évolution génétique des populations.

En étudiant les '*variables factor maps*', nous pouvons savoir quelles variables peuvent potentiellement avoir des valeurs différentes entre les systèmes de parenté et les départager (Figure 1-4 et Figure 1-6). Sur les deux ACP, les populations M se distinguent des populations

P sur le premier axe. Ainsi, les variables fortement corrélées au premier axe sont celles qui possèdent des valeurs différentes entre les populations M et P. Pour l'ACP incluant les 12 populations (Figure 1-2), les variables corrélées au premier axe sont : la proportion de couples natolocaux, la proportion de couple néolocaux, la proportion de couples patrilocaux, le taux de migration de femmes et le taux d'endogamie de village.

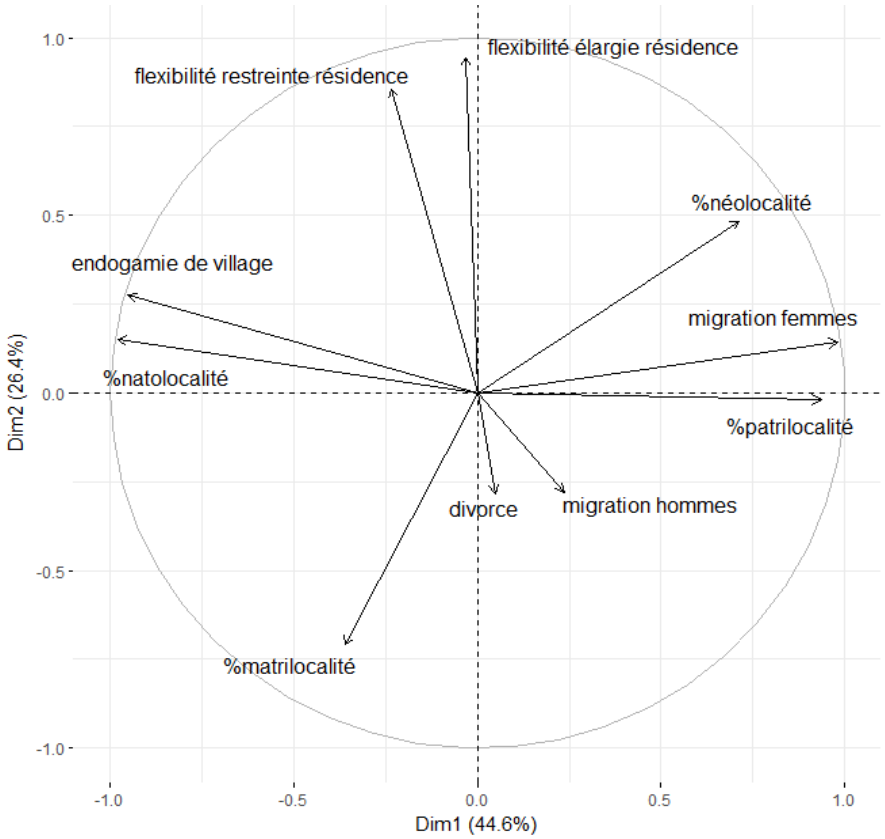


Figure 1-4 : 'Variables factor map' de l'ACP sur les 12 populations. Les variables ethno-démographiques utilisées dans l'analyse sont représentées sur les deux premières dimensions de l'ACP.

Cette observation est confirmée par la matrice de corrélation (Figure 1-5). Sur la matrice de corrélation, nous pouvons voir qu'il existe deux clusters de variables (carré noir). Ces clusters regroupent les variables significativement et positivement corrélées entre elles. De plus, les variables appartenant à un cluster sont négativement corrélées aux variables du second cluster. Le premier cluster regroupe le taux de migration de femmes, la proportion de couples néolocaux, la proportion de couples patrilocaux. Ces variables sont associées aux populations P : leur vecteur sur la *'variables factor map'* (Figure 1-4) va dans le sens des valeurs positives sur le premier axe, et les populations P se trouvent dans les valeurs positives sur le premier axe (Figure 1-2). Comparées aux populations M, les populations P seraient caractérisées par des valeurs élevées pour ces variables. Nous remarquons sur la matrice de corrélation que le taux de

flexibilité élargie de résidence est faiblement mais toujours positivement corrélé aux variables de ce premier cluster (Figure 1-5). Ceci pourrait être le signe d'une association faible de cette variable aux populations P. A l'inverse, le second cluster regroupe le taux d'endogamie de village et la proportion de couples natolocaux qui sont les variables associées aux populations M. Comparé aux populations P, les populations M seraient caractérisées par des valeurs élevées pour ces variables.

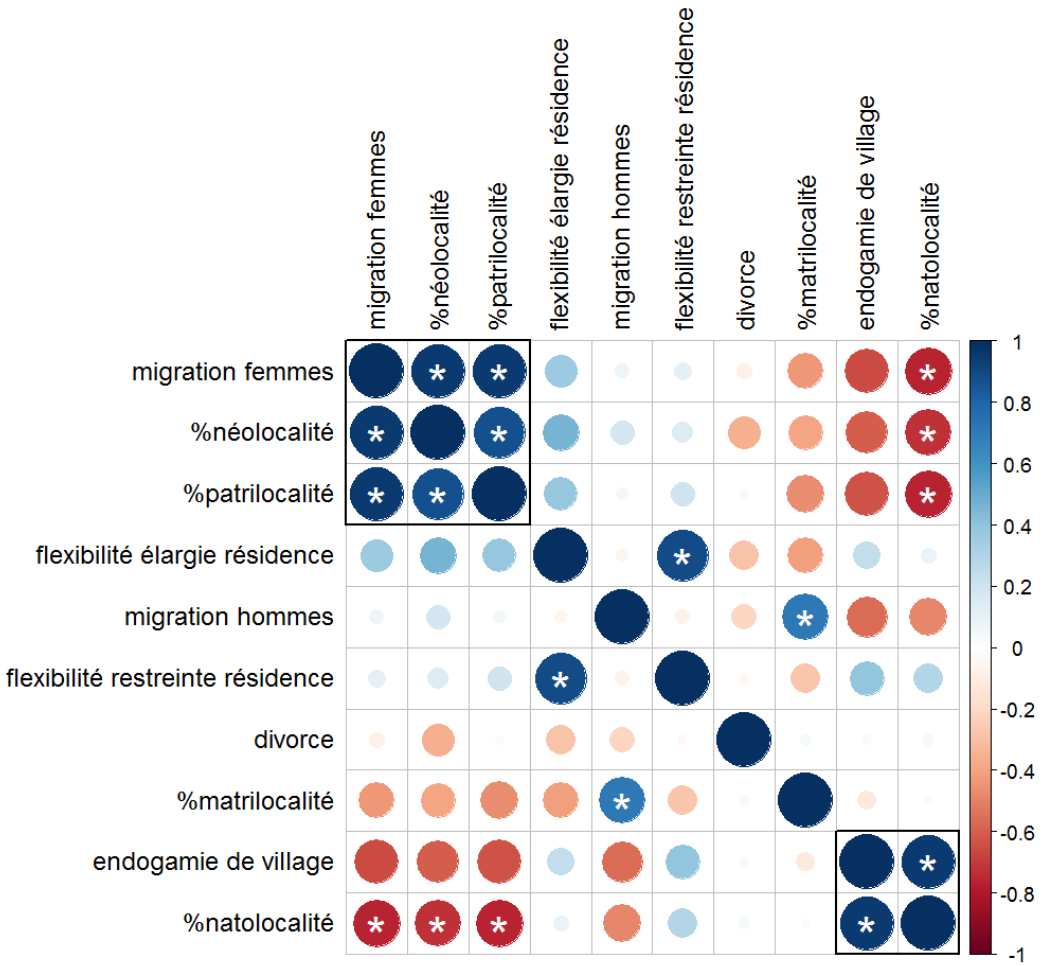


Figure 1-5 : Matrice de corrélation des variables ethno-démographiques sur les 12 populations. Plus un point a une couleur foncée et un rayon important, plus le coefficient de corrélation de Spearman est élevé (en bleu pour les valeurs positives et en rouge pour les valeurs négatives). Les corrélations significatives au seuil de 0,05 après correction de Bonferroni pour tests multiples sont marquées d'un astérisque.

Les mêmes variables ressortent sur la *'variables factor map'* de l'ACP sur les 8 populations qui intègre les variables de filiation (Figure 1-6). La matrice de corrélation (Figure 1-7) nous confirme à nouveau l'existence de 2 clusters, l'un associé au système de parenté patrilinéaire et

patrilocal, et l'autre au système de parenté matrilinéaire et matrilocal. Certaines variables de filiation rejoignent l'un des 2 clusters. Le taux d'exogamie de clan ainsi que le taux de connaissance des clans de la branche maternelle sont associées aux populations M tandis que le taux de connaissance des clans de la branche paternelle est associé aux populations P.

De plus, nous observons plusieurs variables faiblement associées à ces deux clusters. Outre le taux de flexibilité élargie de la règle de résidence que nous avons déjà mentionné, le nombre de clans est aussi faiblement associé au cluster des populations P. La proportion de couples matrilocaux quant à elle est faiblement associée au cluster des populations M.

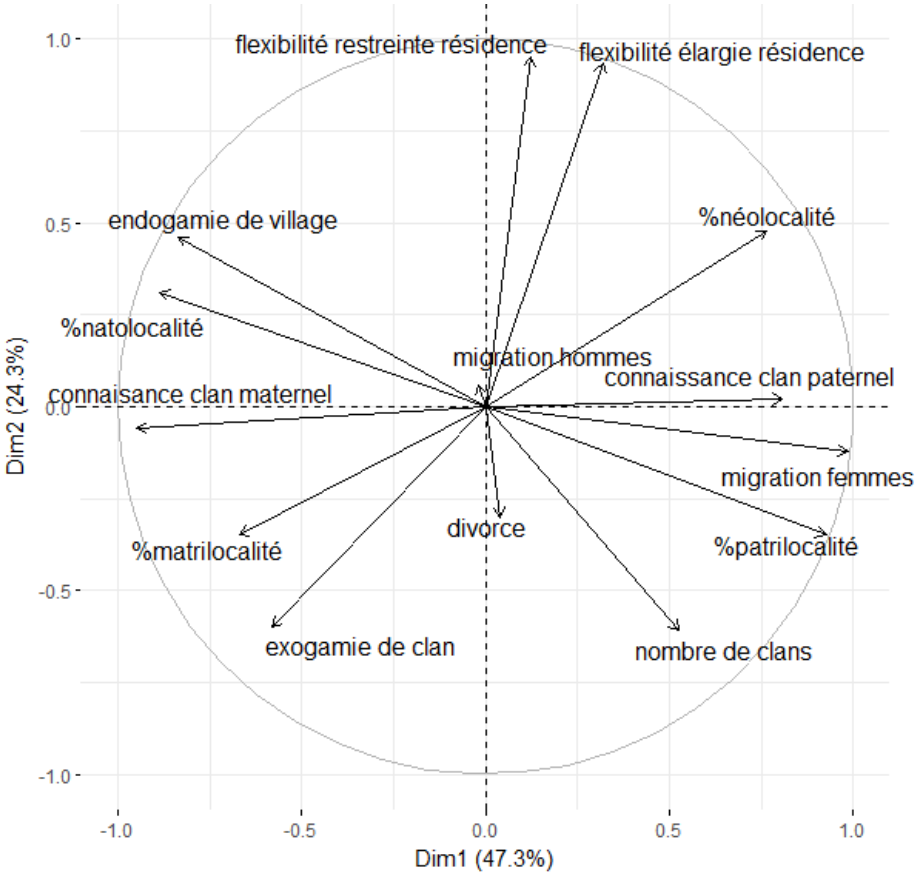


Figure 1-6 : 'Variables factor map' de l'ACP sur les 8 populations. Les variables ethno-démographiques utilisées dans l'analyse sont représentées sur les deux premières dimensions de l'ACP.

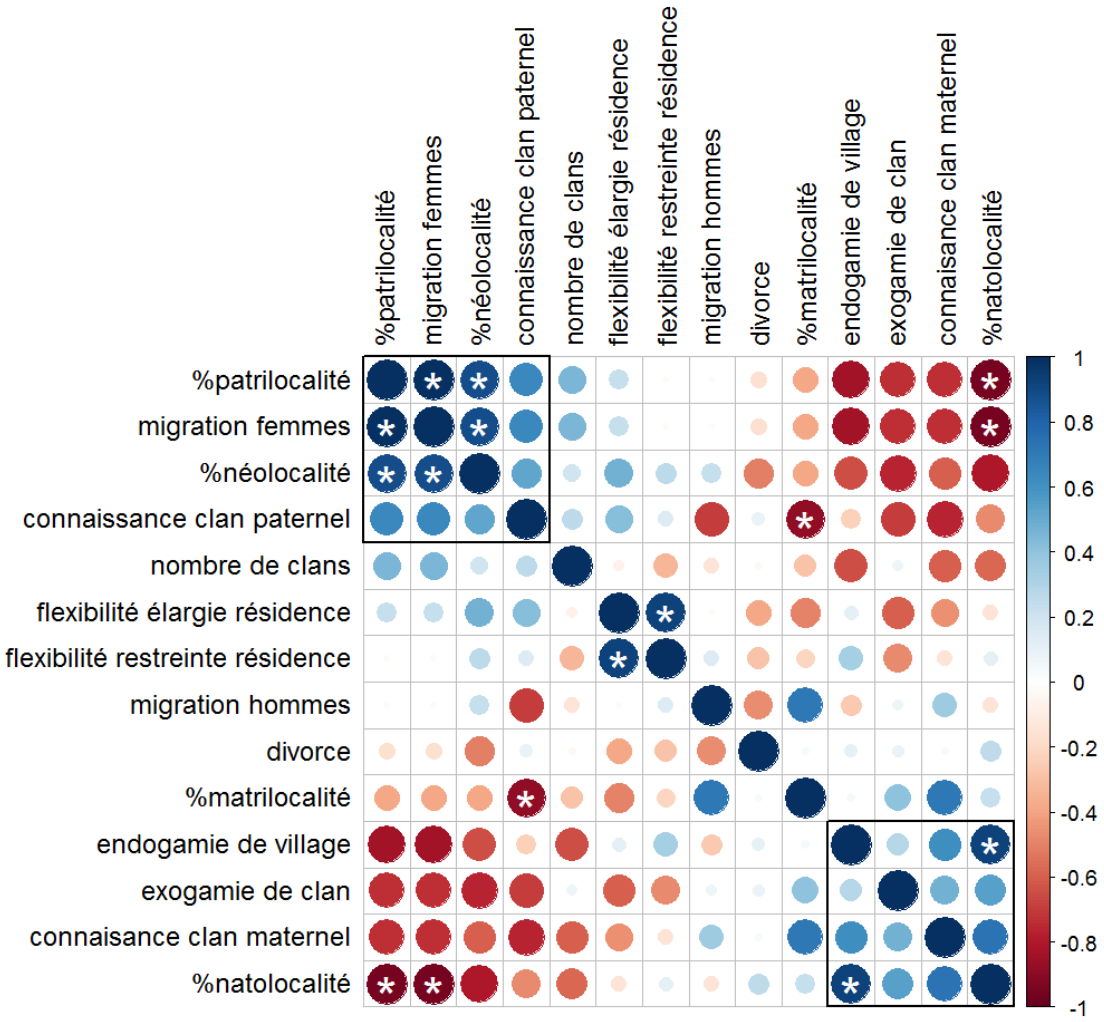


Figure 1-7 : Matrice de corrélation des variables ethno-démographiques sur les 8 populations. Plus un point a une couleur foncée et un rayon important, plus le coefficient de corrélation de Spearman est élevé (en bleu pour les valeurs positives et en rouge pour les valeurs négatives). Les corrélations significatives au seuil de 0,05 après correction de Bonferroni pour tests multiples sont marquées d'un astérisque.

Un des résultats qui nous surprend ici est l'association de la flexibilité élargie de la règle de résidence avec les populations P. Certes, cette association est faible, mais sous l'hypothèse énoncée en introduction en lien avec le « puzzle matrilinéaire », nous nous attendions à avoir une association de la flexibilité (restreinte mais aussi élargie) de la règle de résidence avec les populations M plutôt que les populations P. Nous discuterons ce résultat en détail ultérieurement.

3.2 Comparaison des variables ethno-démographiques entre systèmes de parenté

Maintenant que nous avons montré que les systèmes de parenté peuvent être départagés sur la base des variables ethno-démographiques, nous allons vérifier s'il existe des différences entre les systèmes de parenté pour les variables d'intérêt pour l'évolution de la diversité génétique en testant leur association par des modèles linéaires généralisés mixtes. Les résultats des différentes comparaisons sont résumés dans la Table 1-1.

P-value du GLMM	**M vs C**	**M vs P**	**C vs P**
Natolocalité	0,68	**0,0001**	**0,0021**
Néolocalité	**0,023**	**0,0001**	0,2042
Patrilocalité	**0,007**	**<0,0001**	**<0,0001**
Matrilocalité	0,36	**0,0013**	0,0523
Migration d'hommes	0,99	0,9769	0,9854
Migration de femmes	**0,0004**	**<0,0001**	**<0,0001**
Endogamie de village	0,86	**0,0003**	**0,0012**
Flexibilité restreinte résidence	**0,0012**	0,2416	0,058
Flexibilité élargie résidence	**0,0003**	**0,0048**	0,5806
Exogamie de clan		**<0,0001**	
Connaissance du clan maternel		**<0,0001**	
Connaissance du clan paternel		**<0,0001**	
Connaissance du clan principal		**0,0009**	
Divorce	1,00	0,99	0,99
Polygynie	1,00	1,00	1,00

Table 1-1 : P-values des comparaisons entre systèmes de parenté pour les variables ethno-démographiques d'intérêt. Les p-values pour les comparaisons entre populations M et C sont données dans la première colonne, entre populations M et P dans la seconde et entre populations C et P dans la troisième colonne. La couleur de la p-value indique le système de parenté pour lequel l'estimation de la variable est plus élevée (rouge pour les populations M, bleu pour les populations C et orange pour les populations P). Les p-values en gras sont significatives au seuil de 0,05.

La Figure 1-8 présente les différents taux de résidence des couples. Toutes les populations ont une forte proportion de couples natolocaux (0,74 en moyenne, intervalle de confiance à 95% [0,67 ; 0,82]. Les populations M et C ont un taux de natolocalité (0,81 et 0,79 respectivement) significativement supérieur à celui des populations P (0,63 ; p-value < 0,05). Comme attendu, la proportion de couples patrilocaux est significativement plus élevée chez les populations P (0,20) qui ont une règle de résidence patrilocale que les populations C (0,05 ; p-value < 0,001) qui ont elles même une proportion plus élevée que les populations M (0,02 ; p-values = 0,001).

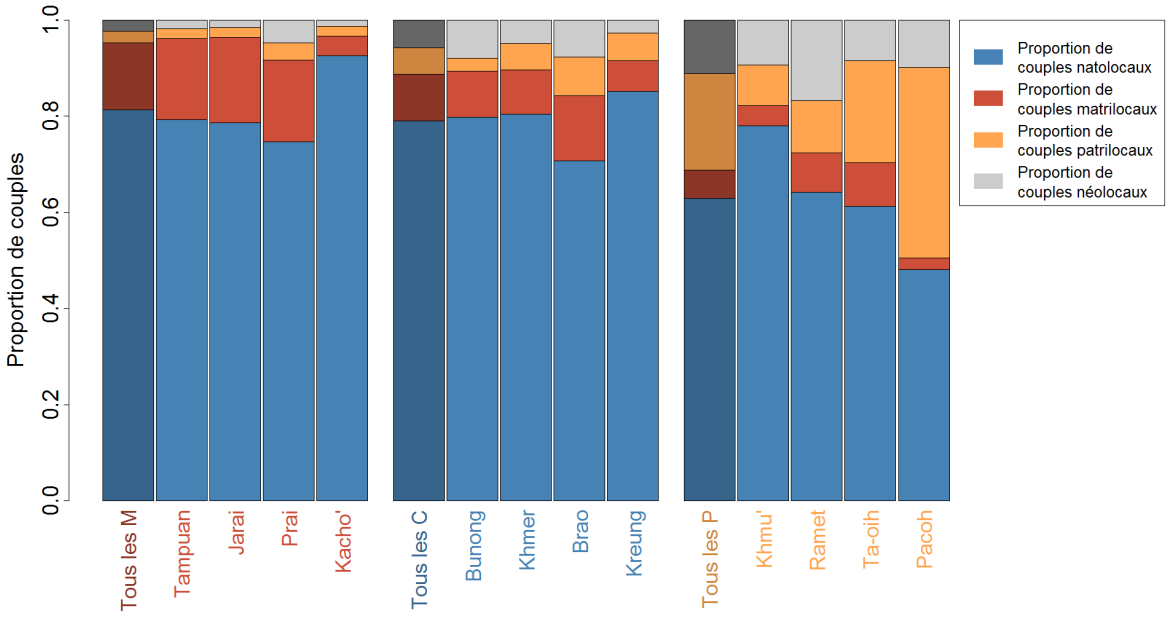


Figure 1-8 : Proportion de couples dans chaque type de résidence à l'échelle du village. La proportion de couples natolocaux est représentée en bleu, la proportion de couples matrilocaux en rouge, la proportion de couples patrilocaux en orange et la proportion de couples néolocaux en gris. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente les valeurs moyennes par système de parenté (M, C, P).

De même, les populations M ont une proportion de couples matrilocaux (0,13) significativement plus élevée que les populations P (0,06 ; p-value < 0,05). En revanche, les populations C n'ont pas une proportion de couples matrilocaux (0,10) significativement différente des populations M (p-value = 0,36) ou P (p-value = 0,052). Enfin, la proportion de couples néolocaux est significativement plus faible chez les populations M (0,02) comparée aux populations C et P (0,06 et 0,011 respectivement ; p-values < 0,05 dans les deux cas). Ces résultats confirment nos observations précédentes basées sur l'ACP et la matrice de corrélation. Les proportions de couples patrilocaux et néolocaux sont plus élevés dans les populations P par rapport aux populations M. Nous observons la même chose pour la proportion de couples natolocaux qui était associée au système de parenté M. Nous notons qu'il existe une différence significative entre les populations M et P pour la proportion de couples matrilocaux (qui était en revanche faiblement associé au cluster des populations M). La proportion de couples matrilocaux intermédiaire entre les populations M et P chez les populations C pourrait expliquer l'exclusion de cette variable dans le cluster des populations M dans la matrice de corrélation.

La Figure 1-9 présente les taux de migration d'hommes et de femmes par population. Pour rappel, le taux de migration d'hommes est la somme du taux de matrilocalité et de néolocalité, le taux de migration de femmes la somme du taux de patrilocalité et de néolocalité. Le taux de migration de femmes est significativement plus faible dans les populations M (0,05) que les populations C (0,11 ; p-value < 0,001), qui à leur tour ont un taux significativement plus faible que les populations P (0,31 ; p-value < 0,001). En revanche, le taux de migration d'hommes est de 0,16, 0,15 et 0,17 respectivement chez les populations M, C et P et n'est pas significativement différent entre les systèmes de parenté (p-values > 0,05).

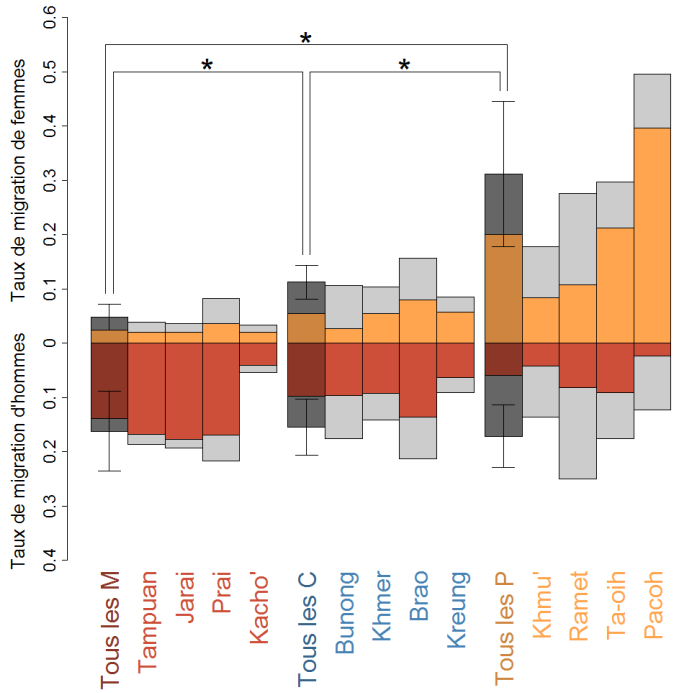


Figure 1-9 : Taux de migration de femmes et d'hommes. Le taux de migration de femmes est la somme de la proportion de couples patrilocaux (orange) et de couples néolocaux (gris). Le taux de migration d'hommes est la somme de la proportion de couples matrilocaux (rouge) et de couples néolocaux (gris). Les taux de migration de femme sont représentés sur le haut de la figure et les taux de migration d'hommes sur le bas de la figure. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire généralisé mixte.

Le taux d'endogamie de village (Figure 1-10) est significativement inférieur chez les populations P (0,68) que chez les populations M ou C (0,83 et 0,82 respectivement ; p-values < 0,05). Il n'y a pas de différence entre les populations M et C (p-values = 0,86).

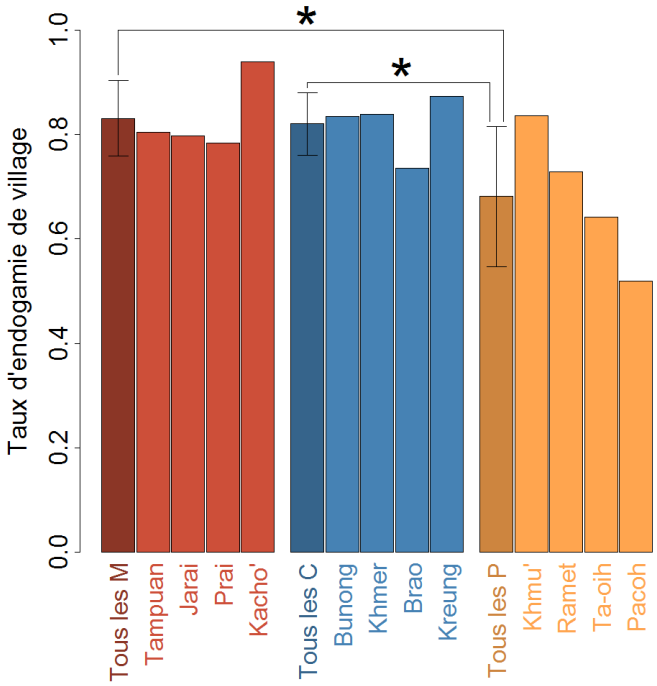


Figure 1-10 : Taux d'endogamie de village. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire généralisé mixte.

Le taux de flexibilité restreint de la règle de résidence est significativement inférieur dans les populations M (0,18) que dans les populations C (0,35 ; p-value = 0,001, Figure 1-11 A). Il n'y a pas de différences significatives entre les populations P (0,28) et M (p-value = 0,24) ou C (p-value = 0,058).

Par ailleurs, les Kacho' ont une flexibilité restreinte de la règle de résidence élevée (0,33) par rapport aux autres populations M (0,13 en moyenne). Cette forte flexibilité pourrait être une conséquence de la transition de cette population d'un système matrilinéaire à un système patrilinéaire. Les Pacoh ont à l'inverse un taux de flexibilité élargie plus faible (0,06) par rapport aux autres populations P (0,36 en moyenne). C'est une différence importante qui explique en partie la variabilité que nous avons observée sur les ACP au sein des populations P.

Le taux de flexibilité élargie de la règle de résidence (Figure 1-11 B) est significativement inférieur dans les populations M (0,28) que dans les populations C (0,54 ; p-value < 0,001) ou P (0,50 ; p-value = 0,005). Il n'y a pas de différences significatives entre les populations C et P (p-value = 0,58).

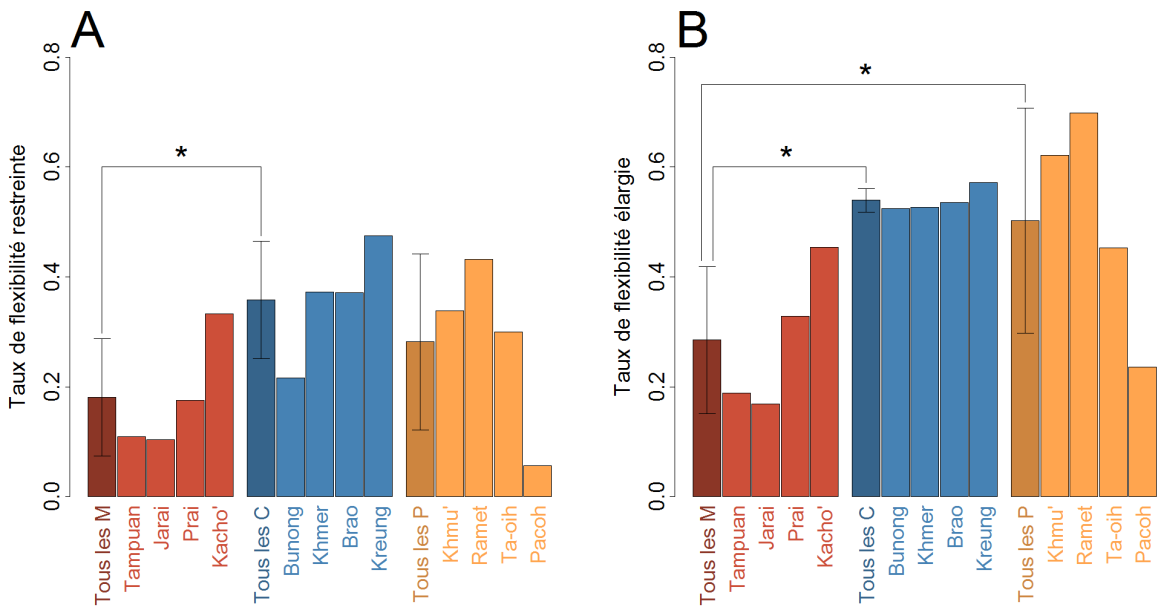


Figure 1-11 : Taux de flexibilité de la règle de résidence. (A) Taux de flexibilité restreinte de la règle de résidence. (B) Taux de flexibilité élargie de la règle de résidence. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire généralisé mixte.

La Figure 1-12 présente les taux de connaissance de clan de la branche paternelle et maternelle dans les populations unilinéaires. Les populations M connaissent mieux les clans de leur branche maternelle que ceux de leur branche paternelle. De même, les populations P connaissent mieux les clans de leur branche paternelle que ceux de leur branche maternelle, à l'exception des Khmu' qui connaissent aussi bien les clans des deux branches. Comme attendu, le taux de connaissance de clan de la branche paternelle est significativement supérieur dans les populations P (0,85) par rapport aux populations M (0,74 ; p-value < 0,001) et le taux de connaissance de clan de la branche maternelle est significativement supérieur dans les populations M (0,93) par rapport aux populations P (0,72 ; p-value < 0,001). Par ailleurs, les populations M connaissent significativement mieux les clans de leur branche maternelle que les populations P ceux de leur branche paternelle (p-value = 0,0039).

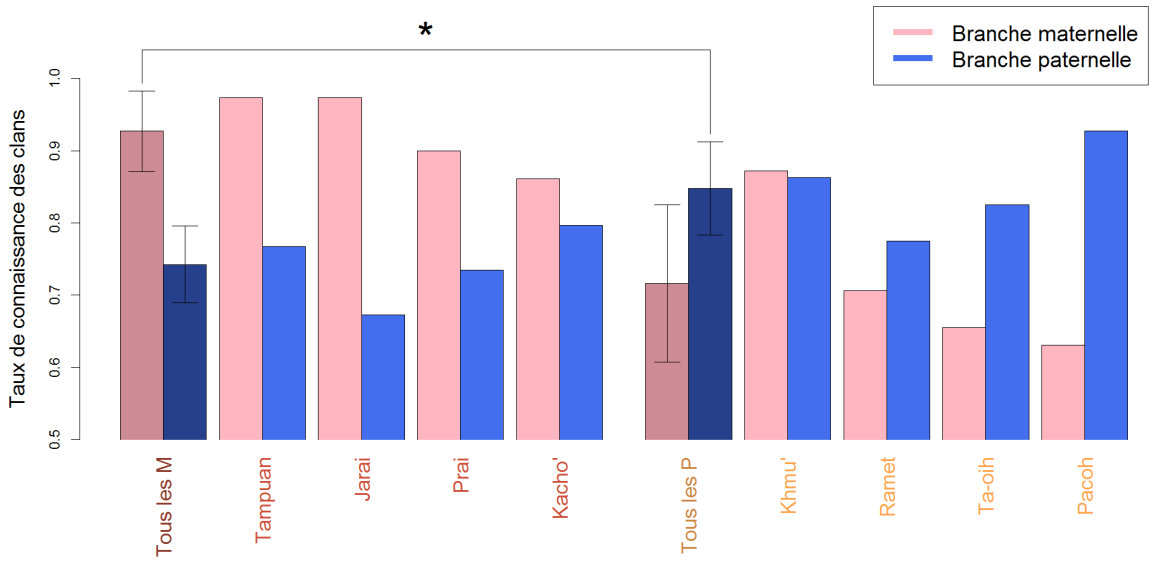


Figure 1-12 : Taux de connaissance des clans des ascendants. Le taux de connaissance de la branche maternelle est représenté en rose et celui de la branche paternelle en bleu. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire généralisé mixte.

Le taux d'exogamie de clan est aussi significativement supérieur dans les populations M (0,94) comparé aux populations P (0,52 ; p-value < 0,001, Figure 1-13).

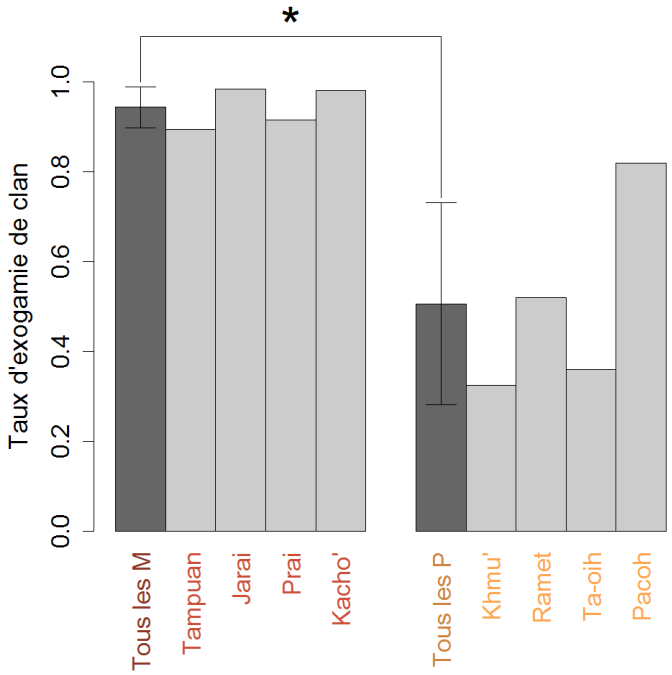


Figure 1-13 : Taux d'exogamie de clan. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire généralisé mixte.

Le nombre de clans n'est pas significativement différent entre les populations M et P (p-value = 0.38, test de Mann-Whitney, Figure 1-14). Cependant, nous notons que les Prai possèdent un grand nombre de clans (48) par rapport aux autres populations M (9, 8 et 17 respectivement chez les Tampuan, Jarai et Kacho'). De même, les Pacoh ont aussi un plus grand nombre de clans (57) comparé aux autres populations P (17, 11 et 29 respectivement chez les Khmu', Ramet et Ta-oih).

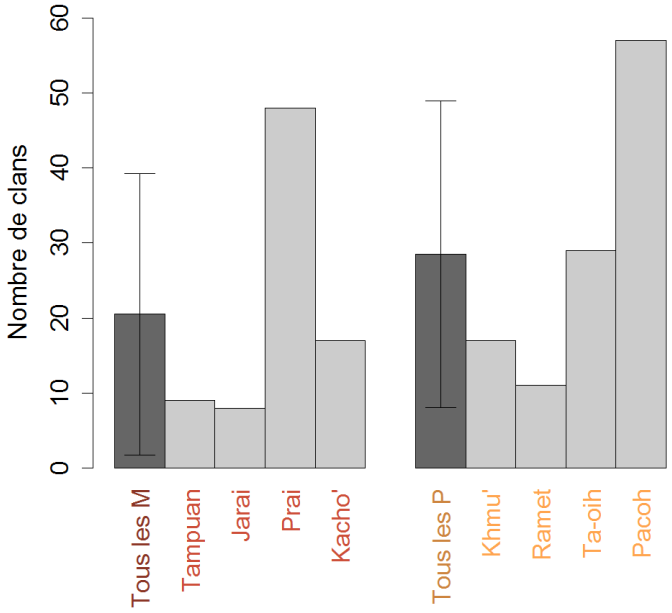


Figure 1-14 : Nombre de clans. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté.

Enfin, la Figure 1-15 présente le taux de divorce (Figure 1-15 A) et de polygynie (Figure 1-15 B) dans les différentes populations. Le taux de divorce varie de 0% chez les Ramet à 13% chez les Ta-oih. Nous observons des individus pratiquants la polygynie uniquement chez 3 des 12 populations : Les Tampuan, les Khmer et les Pacoh. Nous n'avons pas observé de différence significative entre les systèmes de parenté pour ces deux variables ethno-démographiques (p-values > 0,05).

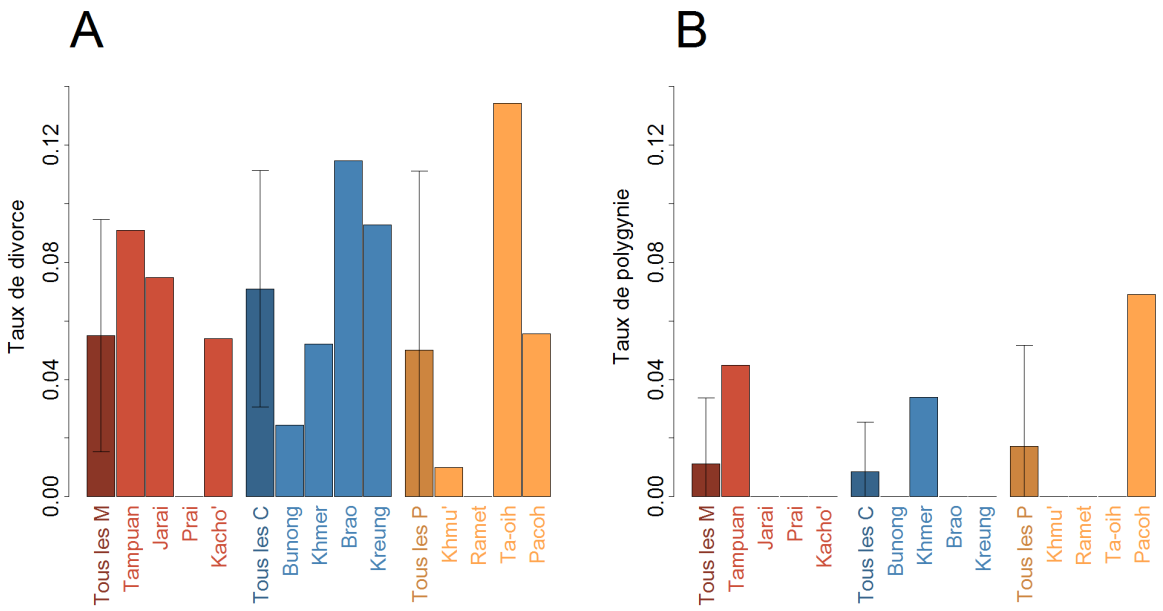


Figure 1-15 : Taux de divorce (A) et taux de polygynie (B) dans les populations. Les populations sont regroupées par système de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté.

4. Discussion

Dans ce chapitre, nous avons tout d'abord montré que nous pouvons distinguer les systèmes de parenté sur la base des variables ethno-démographiques. En particulier, nous pouvons distinguer les populations avec un système de parenté patrilinéaire et patrilocale (P) des populations avec un système de parenté matrilinéaire et matrilocale (M). Les populations possédant le même système de parenté ont des caractéristiques ethno-démographiques similaires. Ces résultats sont importants car ils ouvrent la voie à une possible catégorisation du système de parenté d'une population sur la base de données démographiques, largement disponible par exemple lors des recensements de populations ou par les enquêtes de type *'Household Survey'*. Cependant, il existe une variabilité importante au sein des populations P.

A la lumière des analyses effectuées dans ce chapitre, nous allons maintenant répondre aux questions posées en introduction de chapitre et discuter ces résultats.

Tout d'abord, nous nous sommes demandé s'il existe une flexibilité de la règle de résidence plus importante chez les populations M que P liée au « puzzle matrilinéaire ». Nous avons montré que les populations ont un taux de flexibilité élargie de la règle de résidence différent selon leur système de parenté. Cependant, contrairement à notre supposition dans l'introduction, ce sont les populations P qui sont plus flexibles comparées aux populations M. Ce taux de flexibilité élargie de la règle de résidence provient de deux phénomènes dont les effets sont additifs. Tout d'abord, les populations P ont une flexibilité restreinte de la règle de résidence plus importante que les populations M (0,24 et 0,16 respectivement) malgré le fait que la différence ne soit pas significative. Les travaux ethnographiques documentent l'existence de cette flexibilité restreinte chez certaines populations patrilinéaires et patrilocales étudiées. Olivier Evrard écrit que chez les Khmu : « certains hommes viennent habiter définitivement dans la maison de leur femme. Ils quittent alors leur lignage d'origine et abandonnent leur droit à l'héritage de leurs parents. Il s'agit ici d'une dysharmonie rare mais récurrente, identifiée à quatre reprises (maisons IX, XIII, XIX, XXIII). » (2006, p218). Ainsi, cette flexibilité restreinte de la règle de résidence, qui est contraire à l'attendu du « puzzle matrilinéaire », augmente la proportion de couples matrilocaux chez les populations P.

Ajouté à ce taux de flexibilité restreinte de la règle de résidence plus important, les populations P ont aussi un taux de néolocalité plus élevé que les populations M, particulièrement chez les Ramet. En effet, chez les Ramet, la création de nouveaux villages est très aisée. Il suffit qu'un groupe de plusieurs familles décide de s'installer définitivement sur l'essart qu'ils ont travaillé

ensemble pour qu'un nouveau village émerge. Ce phénomène n'est pas rare, et peut être déclenché par exemple par une mésentente avec certains membres du village d'origine (Izikowitz, 1951).

La combinaison d'un taux de flexibilité restreinte de la règle de résidence élevé ainsi que d'un taux de néolocalité élevé augmentent le taux de flexibilité élargie de la règle de résidence. Ceci entraine alors une augmentation du taux de migration d'hommes dans les populations P qui devient similaire à celui dans les populations M. A l'inverse, le taux de migration de femmes est significativement différent entre les populations M et P comme attendu selon leur règle de résidence.

Toujours en lien avec le « puzzle matrilinéaire », nous avons testé si les populations M ont un taux d'endogamie de village plus important que les populations P. Cette fois-ci, nous avons montré que, comme attendu, les populations M possèdent un taux d'endogamie de village plus important que les populations P. Nous pouvons donc penser que ces populations matrilinéaires et matrilocales d'Asie du Sud-Est pratique l'endogamie de village pour faire face au « puzzle matrilinéaire », plutôt que le mariage patrilocal qui aurait augmenté le taux de flexibilité de la règle de résidence.

Les travaux ethnographiques sur les populations matrilinéaires à l'étude confirment une préférence pour des mariages endogames de village : les Jarai se marient selon la règle du « au plus près, au plus sûr » d'après Jacques Dournes qui estime l'endogamie de village à environ 50% à comparer aux 80% selon nos données (1972). Chez les Tampuan, le conseil des anciens qui devait fournir son aval pour l'intégration d'un nouveau mâle dans le village était d'autant plus réticent que l'homme venait d'un village éloigné, une préférence encore d'actualité aujourd'hui (Bourdier communication personnelle). Ce dernier élément laisse penser que dans le cadre du « puzzle matrilinéaire » c'est autant le futur marié que ses futurs beaux-parents qui sont réticents à un mariage inter- village, d'autant plus que le village est éloigné.

Les populations C qui sont cognatiques et matrilocales ont aussi un taux d'endogamie de village plus élevé que les populations P. Il n'existe pas de groupes de filiation dans les populations cognatiques, cependant, un statut d'autorité peut être acquis par les hommes au sein du village et peut être transmis d'une génération à une autre (Kloos, 1963). Si un homme se marie à une femme d'un autre village, il ne pourra pas hériter du statut de son parent. Ainsi, la préférence de l'endogamie n'est pas seulement présente dans les populations matrilinéaires et matrilocales mais peut aussi être présente dans les populations cognatiques et matrilocales. Il semblerait

donc que les taux d'endogamie de village élevés observés dans ces populations résulteraient moins d'une question d'autorité des hommes au sein des groupes de filiation mais de la possession du pouvoir politique au sein du village par les hommes.

Enfin, la troisième hypothèse liée au « puzzle matrilinéaire » est que les populations M ont un taux de divorce supérieur aux populations P. Nous avons montré que le taux de divorce n'est pas différent entre les populations M et P contrairement à notre attente. Une explication de cette absence de différence entre systèmes de parenté est que nos taux de divorce sont biaisés, car nous avons uniquement récupéré des données auprès d'informateurs actuellement mariés et à un moment donné dans le temps. Nous ne prenons donc pas en compte les individus divorcés et non remariés. De plus, comme nous ne suivons pas les individus sur toute leur vie reproductive, nous ne savons pas si certains d'entre eux divorceront dans le futur. Ainsi, nous avons sans doute sous-estimé les taux de divorce réels dans les populations, et la sous-estimation est d'autant plus importante que le taux de divorce réel est important. Les taux de divorce peuvent donc être en réalité différents entre système de parenté. Par comparaison, Dournes (1972) obtient un taux de divorce de 22% après avoir recensé 2 villages Jarai, à comparer aux 7,4% que nous avons obtenu chez la même population. De plus, il avertit les lecteurs que les individus remariés peuvent cacher leurs essais antérieurs, ce qui complique l'obtention d'une bonne estimation du taux de divorce.

Concernant la préférence pour des mariages entre cousins de premier degré, nos données sur les liens de parenté entre les conjoints n'étaient pas assez complètes pour effectuer des estimations robustes. Nous n'avons donc pas pu vérifier au niveau ethno-démographique s'il existait une préférence plus importante pour les mariages entre cousins de premier degré dans les populations unilinéaires en comparaison aux populations cognatiques. Dans le chapitre 3, nous reviendrons sur cette question en utilisant les données génétiques. Les informations récupérées dans la littérature ethnographique sur les populations d'Asie du Sud-Est étudiées, semblent indiquer que les populations unilinéaires (matrilinéaire et patrilinéaire) auraient une préférence plus marquée pour les mariages entre cousins de premier degré que les populations cognatiques (Izikowitz, 1951; LeBar et al., 1964; Dournes, 1972; Bourdier, 2006; Evrard, 2006; Baird, 2008; Schmutz, 2013). Les travaux ethnographiques mentionnent cette préférence pour cinq des huit populations unilinéaires, tandis qu'aucun ne mentionne cette préférence pour les populations cognatiques (voir table 1 de l'introduction). Ainsi, il se pourrait qu'il existe une différence de proportion de mariages avec les apparentés proches entre les populations

unilinéaires et cognatiques qui entraineraient une consanguinité plus importante chez les populations unilinéaires par rapport aux populations cognatiques.

Nous nous sommes ensuite intéressés à deux aspects du système de parenté d'intérêt pour l'évolution de la diversité génétique : la dynamique des groupes de filiation et la polygynie.

Concernant l'existence de dynamique des groupes de filiation différente entre populations matrilinéaires et patrilinéaires, nous avons montré que l'exogamie de clan est moins respectée et qu'il y a une moins bonne connaissance des clans des ascendants chez les populations P comparées aux populations M. La différence n'est pas surprenante en ce qui concerne l'exogamie de clan car le clan n'est pas l'échelle pertinente d'exogamie chez deux des populations patrilinéaires. En effet, les Khmu' et les Ramet possèdent des lignages au sein des clans qui dans ces populations sont l'échelle pertinente en termes d'exogamie de groupe de filiation. Pour les Ta-oih qui ont aussi un taux d'exogamie de clan relativement faible (0,36), les enquêtes de terrain n'ont pas relevé de lignage mais Schmutz mentionne l'existence de lignage dans cette population (2013). Il se peut donc qu'il existe une structure en lignage qui n'a pas été détectée et qui est l'échelle pertinente d'exogamie chez les Ta-oy, ce qui permet d'expliquer le taux d'exogamie de clan faible chez cette population.

En revanche, la connaissance des clans moins bonne peut être un signe que la structure de filiation est moins rigide chez les populations P par rapport aux populations M. Les travaux ethno-génétiques en cours de Bérénice Alard, et dont je suis co-auteur, confirment également la plus grande perméabilité des lignages et clans dans les populations P que dans les populations M (Alard et al., en préparation). Les travaux ethnographiques rapportent en effet que chez les Khmu', l'accent est mis sur la parenté sociale : ce sont les individus habitant ensemble qui forment l'unité sociale importante. Ainsi : « Au niveau du clan, la filiation est patrilinéaire en théorie, mais en fait elle dépend avant tout du lieu de résidence, de la maison dans laquelle vit l'individu. Ainsi, lorsqu'un homme s'installe définitivement dans la maison des parents de sa femme après le mariage, il adopte leur identité clanique. » (Evrard, 2006).

Toujours sur la dynamique des groupes de filiation, nous avons vu que les Prai et les Pacoh possèdent un grand nombre de clans par rapport aux autres populations unilinéaires. Ce grand nombre peut être le témoin d'une dynamique de fission particulièrement active dans ces populations. En effet, il est difficile de maintenir un tel nombre de clans dans le temps sans constant apport de nouveaux clans (annexe 2). Or une grande proportion des clans sont uniques à un village (56% chez les Prai et 36% chez les Pacoh, voir annexe 3 pour une distribution des

clans par village chez ces deux populations), il est peu probable que ce soient les échanges entre villages qui permettent le maintien de ce grand nombre de clans. Ainsi, la création de clans de novo, par fission de clans déjà existants, est une explication plus probable. Il semble donc qu'il existe une dynamique de fission active chez les Prai et les Pacoh.

La polygynie semble être rare dans toutes les populations. Seul 12 cas de polygynie ont été recensés sur les 766 ménages pour lesquels nous avons cette information.

Nous voulons revenir sur un point important soulevé par nos observations ethno-démographiques : il existe une corrélation entre les variables de résidence et les variables de filiation (Figure 1-16). Au sein des populations M, plus une population a un taux de connaissance des clans de la branche maternelle élevé, plus son taux de flexibilité de la règle de résidence est faible (Figure 1-16 A). De même, au sein des populations P, plus une population a un taux de connaissance des clans de la branche paternelle élevé, plus son taux de flexibilité de la règle de résidence est faible (Figure 1-16 B). Ainsi, plus une population est matrilinéaire, plus elle est matrilocale, et plus une population est patrilinéaire, plus elle est patrilocale. Nous ne pourrons donc pas départager l'effet de la résidence de celui de la filiation sur la diversité génétique. Ainsi, s'il existe par exemple une différence entre les populations M et P au niveau de la diversité mitochondriale, il est tout aussi possible que nous observons l'effet de la matrilocalité que de la matrilinéarité ou un effet combiné de ces deux dernières.

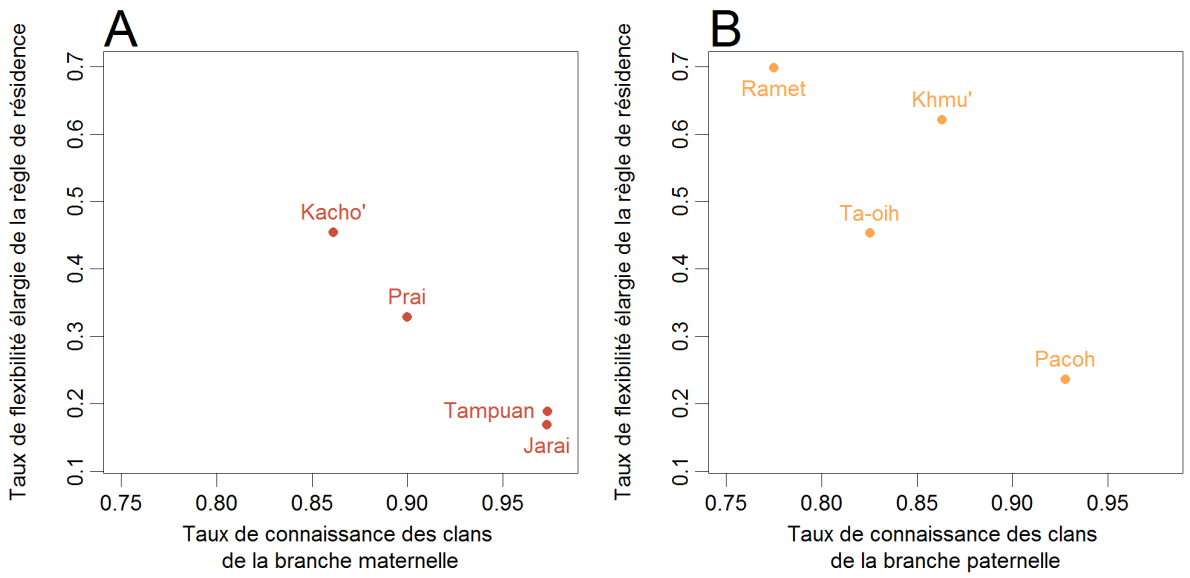


Figure 1-16 : Taux de flexibilité de la règle de résidence en fonction du taux de connaissance des clans de la branche principale pour (A) les populations M, (B) les populations P.

Ajouté à cela, les populations C, censées départager l'effet de la résidence et de la filiation, sont moins matrilocales que les populations M. Ainsi, s'il existe des différences génétiques mitochondriales entre les populations M et C, nous ne serons pas capables de déterminer s'il s'agit d'un effet de la résidence (moins matrilocale chez les populations C) ou de la structuration en groupes de filiation (absente chez les populations C).

Il est à noter que la matrilocalité moindre chez les populations C était attendue du fait que les Brao et les Kreung ont une règle de résidence essentiellement multilocale et donc ont par définition une règle de résidence flexible malgré sa dominante matrilocale (le couple réside généralement avec la famille de la femme juste après le mariage, et le déplacement vers le foyer de la famille du mari et tous les déplacements suivants entre les deux foyers requièrent une cérémonie avec sacrifice (UNDP Cambodia, 2010b)).

5. Conclusion

Dans ce premier chapitre, nous nous sommes intéressés à l'impact des systèmes de parenté sur les variables ethno-démographiques. Les variables ethno-démographiques étant le vecteur canalisant l'influence des systèmes de parenté sur la diversité génétique, nous nous attendons à observer des patrons de diversité génétique en accord avec nos observations ethno-démographiques.

Sous l'hypothèse du « puzzle matrilinéaire », nous nous attendions tout d'abord à retrouver une flexibilité de la règle de résidence plus forte chez les populations M comparées aux populations P. Contrairement à nos attentes, nous avons observé un taux de flexibilité élargie de la règle de résidence plus élevé chez les populations P que les populations M. Par conséquent, le taux de migration d'hommes est similaire entre les systèmes de parenté alors que le taux de migration de femmes est plus important chez les populations M que P. Au niveau génétique, nous nous attendons donc à ne pas voir de différence entre les systèmes de parenté pour le chromosome Y mais une diversité réduite pour l'ADN mitochondrial chez les populations M par rapport aux populations P.

De plus, les groupes de filiation semblent être plus flexibles et perméables chez les populations P que M. Ainsi, les groupes de filiation chez les populations M peuvent sous-structurer les femmes et réduire la diversité de l'ADN mitochondrial alors que chez les populations P, les groupes de filiation ne sous-structurent pas les hommes. Il n'y aurait ainsi que peu ou pas de réduction de la diversité du chromosome Y. Nous nous attendons de plus à un effet synergique de la filiation et de la résidence sur l'ADN mitochondrial, réduisant tous les deux la diversité chez les populations M par rapport aux populations P.

Par ailleurs, il semble y avoir une dynamique de fissions de clans particulièrement active chez les Prai et les Pacoh. Si ces fissions sont linéaires (les individus les plus apparentés se regroupent entre eux), nous pouvons nous attendre à avoir une diversité plus faible pour l'ADN mitochondrial et le chromosome Y respectivement par rapport aux autres populations unilinéaires.

Un second attendu du « puzzle matrilinéaire » est que le taux d'endogamie de village est plus élevé chez les populations M que les populations P. Nous avons effectivement observé un taux d'endogamie de village significativement plus élevé chez les populations M que les populations P. Le taux d'endogamie de village est aussi plus élevé chez les populations C que P. Afin de

rester à proximité de leur groupe de filiation mais aussi pour maintenir leur autorité dans leur village natal, les hommes dans les populations matrilinéaires se marient plus souvent dans leur village natal que dans les populations patrilinéaires. Au niveau génétique, nous nous attendons par conséquent, à avoir une consanguinité plus élevée chez les populations M et C que les populations P.

Enfin, nous avons voulu savoir s'il existait une préférence plus importante pour les mariages entre cousins de premier degré chez les populations unilinéaires par rapport aux populations cognatiques. Nous n'avons pas pu détecter de différences de préférence de mariage entre cousins de premier degré entre populations unilinéaires et cognatiques pour cause de manque de données. Cependant, la littérature ethnographique nous pousse à penser qu'il pourrait exister une préférence plus importante chez les populations unilinéaires comparées aux populations cognatiques et donc une consanguinité plus élevée.

Nous n'avons pas observé de différence de taux de polygynie entre les systèmes de parenté. Nous ne nous attendons donc pas à avoir un effet de la polygynie sur la variance du succès reproducteur des hommes et donc sur la diversité du chromosome Y dans ces populations.

Suite à ces différentes observations, nous allons explorer dans les prochains chapitres l'impact du système de parenté sur la diversité génétique et notamment tester si nous observons des patrons de diversité génétiques uniparentaux (chapitre 2) et autosomaux (chapitre 3) en accord avec ces observations ethno-démographiques.

Chapitre 2 - Des variables ethno-démographiques à la diversité uniparentale

1. Introduction

La quantification des variables ethno-démographiques d'intérêt pour l'évolution génétique des populations nous a permis de mettre en lumière l'existence de différences entre les systèmes de parenté.

Dans ce chapitre, nous allons tester si ces différences se répercutent au niveau de la diversité uniparentale. Comme nous l'avons exposé en introduction, nous nous attendons à ce que la règle de résidence et la règle de filiation aient un effet sur la diversité uniparentale, l'une régulant les flux migratoires des hommes et des femmes entre villages, l'autre les tailles efficaces sexe spécifiques.

Nous avons vu dans le chapitre 1 que les populations P ont une flexibilité élargie de la règle de résidence plus importante que les populations M, avec notamment une part non négligeable des hommes qui s'installent avec leur femme matrilocalement et un grand nombre de couples néolocaux. Cette flexibilité plus importante chez les populations P va dans le sens contraire de l'attendu que nous avions sous l'hypothèse du « puzzle matrilinéaire ». Celle-ci prédit une flexibilité de la règle de résidence plus grande chez les populations M caractérisée par une proportion de femme vivant patrilocalement dans le village de leur mari et permettant à ce dernier de maintenir ainsi une position d'autorité au sein de son clan maternel (Richards, 1950)

Parce que le taux de flexibilité élargie de la règle de résidence est plus important chez les populations P, les migrations d'hommes entre villages ne sont pas aussi réduites qu'attendu sous l'hypothèse de la règle patrilocale. En conséquence, les populations P ont des migrations d'hommes similaires à celles des populations M et C. En revanche, les migrations de femmes sont plus élevées, comme attendu, dans les populations P que dans les populations M et C.

Si ces patrons de migrations sexe-spécifiques perdurent depuis un nombre de générations suffisant, il est attendu qu'ils aient eu un impact sur la diversité uniparentale. Ainsi nous nous attendons à observer une diversité mitochondriale plus faible pour les populations M et C par rapport aux populations P et une diversité du chromosome Y similaire entre les systèmes de parenté.

Par ailleurs, nous avons montré des dynamiques de groupes de filiation variées avec notamment deux des populations unilinéaires possédant un grand nombre de clans : les Prai et les Pacoh. Nous avons émis l'hypothèse que ce grand nombre de clans résultait d'une dynamique de fission particulièrement active. Si ces fissions sont linéaires, elles entrainent une diminution de la diversité génétique dans ces populations (Smouse et al., 1981). Nous nous attendons alors à avoir une diversité de l'ADN mitochondrial plus faible chez les Prai par rapport aux autres populations matrilinéaires et une diversité du chromosome Y plus faible chez les Pacoh par rapport aux autres populations patrilinéaires.

Dans la partie suivante qui se présente sous la forme d'un article qui a été accepté pour publication dans l'American Journal of Physical Anthropology, nous analysons l'influence des systèmes de parenté sur les patrons de diversité uniparentale dans les populations d'Asie du Sud-Est à la lumière des différences ethno-démographiques observées dans le chapitre 1.

2. Article : Residence rule flexibility and descent groups dynamics shape uniparental genetic diversities in South East Asia

ABSTRACT

Objectives

Social organization plays a major role in shaping human population genetic diversity. Previous studies have reported inconsistent relationships between social and genetic structures at local scales in matrilocal and patrilocal human populations. The objective of this study is to understand the reasons for such inconsistencies and further evaluate the influence of social organization on genetic diversity.

Materials and Methods

We explored mitochondrial and Y chromosome diversity patterns in 12 populations (n=619) from mainland South-East Asia exhibiting a wide range of social organizations, along with quantitative ethno-demographic information sampled at the individual level.

Results

We observed a clear signature of matrilocality but not of patrilocality on uniparental genetic markers: mtDNA diversity was lower in matrilocal than in multilocal and patrilocal populations while Y chromosome diversity was similar among these social organizations. The reasons for such asymmetry at the genetic level were understood by quantifying sex-specific migrations rates from our ethno-demographic data: while female migration rates varied between social organizations, male migration rates did not. This unexpected lack of difference in male migrations resulted from a higher flexibility in residence rule in patrilocal than in matrilocal populations. In addition, our data suggested an impact of clan fission process on uniparental genetic patterns.

Conclusions

This higher residence flexibility in the patrilocal populations explains the observed lack of signature of patrilocality on Y chromosome patterns, thus providing an explanation for the apparent discrepancies between social and genetic structures. Altogether, this study highlights the need to quantify the actual residence and descent patterns to fit social to genetic structures.

INTRODUCTION

Larger worldwide genetic differences between human populations have been reported for Y chromosome than for mitochondrial DNA (Lippold et al., 2014; Seielstad, Minch, & Cavalli-Sforza, 1998; Wilkins & Marlowe, 2006). This pattern has been mainly attributed to the high prevalence of patrilocality, with 70% of human populations thought to be patrilocal (Godelier, 2004; Pasternak, Ember, & Ember, 1976). Indeed, patrilocality, i.e. when women move to reside close to their husbands' families, limits male marital migrations and increases those of females. The marital migration flow is reversed in matrilocal populations. We thus expect to see higher within-population and lower among-population mitochondrial diversity in patrilocal than in matrilocal populations, and the opposite patterns for Y-chromosome.

However, at the local scale, no consensus has been reached on the impact of post-marital residence on uniparental genetic diversity with conflicting results being reported. For instance, mtDNA and Y chromosome genetic patterns are correlated with residence rules in Sino-Tibetan matrilocal and patrilocal populations from northern Thailand (Hamilton, Stoneking, & Excoffier, 2005; Oota, Settheetham-Ishida, Tiwawech, Ishida, & Stoneking, 2001), and in matrilocal populations from West Timor (Tumonggor et al., 2014). However, Hmong-Mien patrilocal and matrilocal populations from Thailand have comparable levels of uniparental genetic diversity (Besaggio et al., 2007). Similarly, uniparental genetic diversity is not

correlated to post-marital residence rules in Austro-Asiatic and Dravidian populations from India (Kumar et al., 2006). More recently, a study of two populations from Sumatra have found a signature of matrilocality, but not of patrilocality, on uniparental genetic diversity (Gunnarsdóttir et al., 2011).

These conflicting results may come from the fact that these studies have focused mainly on the post-marital residence rule without taking into account other ethno-demographic behaviors that may shape populations' genetic diversity (reviewed in (Heyer, Chaix, Pavard, & Austerlitz, 2012)), such as the rules of descent (which affiliate individuals to kin groups) (Chaix et al., 2007; Montinaro, Davies, & Capelli, 2016; Sanchez-Faddeev et al., 2013; Ségurel et al., 2008) and the rules of alliance (which regulate mate choice) (Guillot et al., 2015; Kayser et al., 2003; Neel, 1970), as well as other social factors such as the transmission of reproductive success (Austerlitz & Heyer, 1998; Hammer, Mendez, Cox, Woerner, & Wall, 2008). In addition, most of these studies have classified populations according to residence or descent rules which may or may not be strictly followed, without attempting to quantify the actual residence and descent patterns. Such a quantitative and interdisciplinary approach is an essential step towards a better understanding of the influence of social organization on genetic diversity (Destro Bisol, Capocasa, & Anagnostou, 2012; Guillot et al., 2015; Shenk & Mattison, 2011).

Here, we undertake an interdisciplinary approach by collecting quantitative ethno-demographic data at the individual level as well as uniparental genetic data from 12 Southeast Asian populations exhibiting a wide variety of descent (matrilineal, patrilineal, or cognatic) and residence (matrilocal, patrilocal, or multilocal) rules. All these populations live in tropical environments and have rice-based diet. Contrary to most previous studies which focus on a particular component of social organization, this study intends to capture multiple aspects of social organization by analyzing in a quantitative manner post-marital residence as well as descent group structure, and confronting them to uniparental genetic diversity. In particular, we

show that both residence and descent patterns are reflected in genetic diversity. A quantitative and global approach of social organization explains genetic diversity more powerfully than discrete categories based on rules and explains apparent discrepancies between residence rules and uniparental genetic diversity. In addition, our study suggests higher social regulation of sex-specific migrations in matrilocal than in patrilocal populations from South-East Asia, which contrasts with observations in Hill Tribes from Thailand (Hamilton et al., 2005).

MATERIALS AND METHODS

Sampled populations

Twelve populations from Cambodia and Laos were sampled in 61 villages during 3 field missions carried out between 2011 and 2012: the Tampuan, Jarai, Kacho', Bunong, Khmer, Brao and Kreung from Cambodia and the Khmu', Ramet, Ta-oih, Pacoh and Prai from Laos (Fig 1).

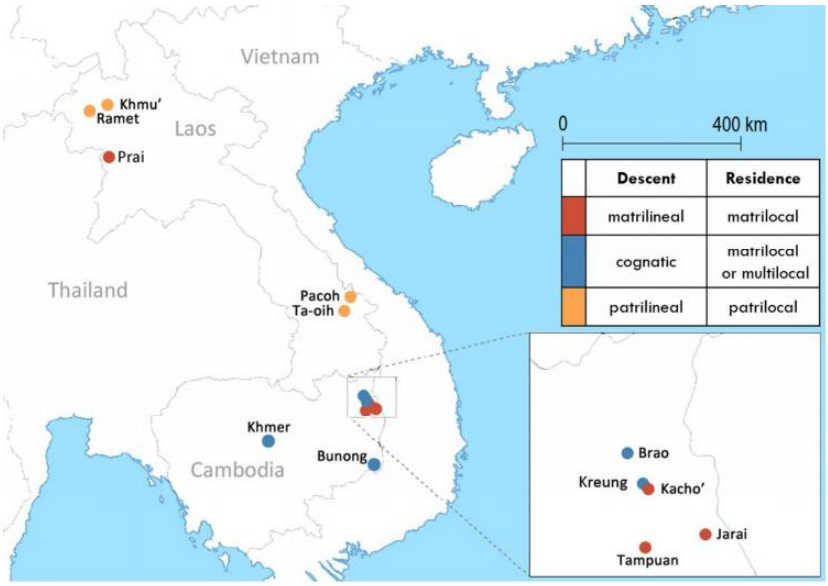


Figure 1: Map of sampled populations. The barycenter of the sampled villages for each population is shown. Populations with matrilineal descent and matrilocal residence are represented in red. Populations with cognatic descent and matrilocal or multilocal residence are represented in blue. Populations with patrilineal descent and patrilocal residence are represented in yellow.

The populations were chosen for their differences in residence and descent rules. Most of them have been the focus of ethnographic works, providing description of their social organization (Table 1). The Tampuan, Jarai, Kacho' and Prai have matrilineal descent and matrilocal residence (Bourdier, 2006; Dessaint, 1981; Dournes, 1972; LeBar, Hickey, & Musgrave, 1964), the Bunong and the Khmer have cognatic descent and matrilocal residence (Ebihara, 1977; Ledgerwood, 1995; Martel, 1975; UNDP Cambodia, 2010b) the Brao and Kreung have cognatic descent and multilocal residence (Baird, 2008; LeBar et al., 1964; Matras-Troubetzkoy, 1983; UNDP Cambodia, 2010a) and the Khmu', Ramet, Ta-oih and Pacoh have patrilineal descent and patrilocal residence (Evrard, 2006; Izikowitz, 1951; LeBar et al., 1964; Lindell, Samuelsson, & Tayanin, 1979; Schmutz, 2013) (see caption of table 1 for definitions of the terms characterizing social organization). Table 1 Description of the studied populations with sampling information.

Population	Descent rule	Residence rule	Abbreviation (Group)	sampled villages	mtDNA samples	Y chromosome samples	ethno-demographic interviews
Tampuan	Matrilineal[1]	Matrilocal[4]	M	8	121	61	65
Jarai	Matrilineal	Matrilocal		6	85	56	56
Prai	Matrilineal	Matrilocal		4	73	38	37
Kacho'	Matrilineal	Matrilocal		3	41	24	27
Bunong	Cognatic[2]	Matrilocal	C	5	49	45	45
Khmer	Cognatic	Matrilocal		5	57	41	44
Brao	Cognatic	Multilocal[5]		6	34	39	39
Kreung	Cognatic	Multilocal		3	34	36	36
Khmu'	Patrilineal[3]	Patrilocal[6]	P	8	43	121	65
Ramet	Patrilineal	Patrilocal		4	21	40	40
Ta-oih	Patrilineal	Patrilocal		4	21	33	34
Pacoh	Patrilineal	Patrilocal		5	40	49	44
Total				61	619	583	532

Table 1: Description of the studied populations with sampling information. [1]matrilineal descent: descent group affiliation is transmitted to the children through the mother. [2]cognatic descent: recognition of descent from both sides of the family in the absence of any specified lines of descent. [3]patrilineal descent: descent group affiliation is transmitted to the children through the father. [4]matrilocal residence: the husband moved to his wife's natal village after marriage. [5]multilocal residence: the couple lives alternatively in the husband's and wife natal villages before settling definitively in one place. [6]patrilocal residence: the wife moves to her husband's natal village after marriage.

Although all unilineal populations under study have harmonious systems (i.e. matrilinearity associated with matrilocality and patrilinearity associated with patrilocality), we expect residence and descent rules to impact genetic diversity in different ways. The residence rule

influences sex-specific migratory processes while the descent rule, subdividing the populations into unilineal subgroups undergoing fusion, fission and extinction processes, would potentially influence sex-specific effective population sizes. Based on their social organization, the populations were divided into 3 categories: populations with matrilineal descent and matrilocal residence (M), populations with patrilineal descent and patrilocal residence (P), and populations with cognatic descent rule and either matrilocal or multilocal residence rule (C). The Jarai speak an Austronesian language while the other populations belong to the Austro-Asiatic linguistic family.

DNA samples

Unrelated individuals at the first cousin level with all four grandparents from the same population were sampled. We collected two saliva samples for each individual (4mL each). Samples were kept in equivalent volume of lysis buffer with 300µL of 10% SDS and 60µL of proteinase K (20mg/mL). DNA was extracted from saliva samples using a standard ethanol precipitation protocol. All participants provided written informed consents and the study was approved by the National Ethic Comities for Health Research in Cambodia and Laos as well as by the Comité Opérationnel pour l'Ethique (CNRS, France).

Mitochondrial genome sequencing

We sequenced 619 individuals for the hypervariable region 1 (HV1) of the mtDNA control region from position 16001 to 16383 using primer L15925 and HH23 on an ABI 3730XL automated DNA analyzer (Applied Biosystems). Sequences were manually aligned using GenalysWin v.3.3.40a. The polyC stretch (sites 16179-16194) was removed from the aligned sequences before analysis. HV1 sequences are provided in table S1.

Y-chromosome STR genotyping

We genotyped 588 individuals using the AmpFlSTR Yfiler PCR Amplification Kit (Applied Biosystems) for 17 Y-chromosome short tandem repeat (STR) loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and Y GATA H4). PCR were performed according to the manufacturer's manual protocol. PCR products were analyzed using an ABI 3130 automated DNA analyzer (Applied Biosystems) and the Genemapper software (Applied Biosystems). Genotypes are provided in table S2.

Sample sizes per village are listed in table S3. Y chromosome STRs and mtDNA HV1 sequences will be released on the appropriate databank upon acceptance of the manuscript.

Genetic data analysis

Within population haplotype diversity for the mtDNA HV1 sequence and Y chromosome STRs data was estimated using Arlequin version 3.5 (Laurent Excoffier & Lischer, 2010). Mann-Whitney-Wilcoxon tests (MWW) were used to assess differences in haplotype diversity between social organizations.

Among villages Φst (L. Excoffier, Smouse, & Quattro, 1992) based on the mtDNA sequence pairwise differences and Rst based on the number of repeat differences between Y haplotypes (Michalakis & Excoffier, 1996) were computed using Arlequin 3.5 (assigning individuals to their village of residence). Villages with less than 5 sampled individuals were removed from the analysis. Significance was assessed using a permutation approach (1000 permutations of individuals among villages for each population).

Ethno-demographic data collection

We interviewed 532 individuals, conjointly with their spouse when possible, and collected ethno-demographic information (descent group affiliation, mother tongue, birth village and residence village) for them, their spouse and their family members (parents, grandparents,

siblings, children and their respective spouses). This procedure allowed us to gather first hand ethno-demographic information for 532 couples (core dataset) and second hand ethno-demographic information for 3530 couples (extended dataset). We interviewed individuals having all four grandparents from the same population, and similarly so for their spouse, consequently all 532 core couples were endogamous at the population level. The DNA samples were taken from the interviewee, his/her spouse and/or one of his/her children. Sample sizes per village are listed in table S3.

Post marital-residence estimators

We classified the couples into 4 categories of post-marital residence (Carrasco, 1963; Casselberry & Valavanes, 1976): i) natolocal when both spouses were born in the same village and resided in it after marriage; ii) patrilocal if the spouses were born in different villages and the wife moved to her husband's village after marriage iii) matrilocal if the spouses were born in different villages and the husband moved to his wife's village after marriage, iv) neolocal if the couple settled in a different village from their respective natal villages. In this last case, spouses can be born in the same village or in different villages.

We estimated male and female post-marital migration rates. The male migration rate corresponds to the sum of the proportions of matrilocal and neolocal couples, and the female migration rate to the sum of the proportions of patrilocal and neolocal couples.

We used logistic regression to assess the influence of social organization (M, P and C) on proportions of natolocal, neolocal, matrilocal and patrilocal couples, as well as male and female migration rates. In all these models, we incorporated together individuals' population, village of residence and family as random effects in order to adjust results for potential sample bias linked to the fact that couples of the same population, village or family might share similar post-marital residence behavior. All following results were found robust to the incorporation of these

variables. All generalized linear mixed models were performed with the 'lme4' v1.1-9 package in R (Bates, Mächler, Bolker, & Walker, 2015) and p-values were obtained using the package 'lmerTest' v2.0-32.

Descent group structure estimators

All unilineal populations are organized into clans. Members of the same clan recognize themselves as having a common paternal or maternal ancestor, in patrilineal and matrilineal populations respectively (Maquet, n.d.). Clan affiliation is inherited through the father in patrilineal populations and through the mother in matrilineal populations.

We estimated the number of clan names in the extended dataset. It was compared between social organizations by MWW test.

We also estimated the proportion of ascendants (i.e. parents and grandparents) for whom the interviewed individuals correctly knew the clan affiliation. A high clan affiliation knowledge rate could indicate that clan affiliation is strictly transmitted from parents to children in a matrilineal or patrilineal way and thus could be used as a proxy for the rigidity of clan structure. Clan affiliation knowledge rates were calculated separately for paternal branches (their fathers and paternal grandfathers) and maternal branches (their mothers and maternal grandmothers). We expect better clan knowledge on maternal branches for matrilineal populations and better clan knowledge on paternal branches for patrilineal populations. The influence of social organization on clan affiliation knowledge rates were assessed by logistic regression incorporating the same random effects as models for residence estimators.

All statistical analysis were performed in R v3.2.2 (R Development Core Team, 2008).

RESULTS

In order to assess the influence of social organization on uniparental genetic diversity patterns, we compared genetic and ethno-demographic estimators collected in three groups of South-East Asian populations: populations with matrilineal descent and matrilocal residence (M), populations with patrilineal descent and patrilocal residence (P), and populations with cognatic descent rule and either matrilocal or multilocal residence rule (C).

Genetic diversity estimators

Within population mitochondrial haplotype diversity ranged from 0.81 to 0.99 (Fig. 2A), with the Prai exhibiting the lowest mitochondrial diversity. M populations had a significantly lower mitochondrial diversity than P populations (p-value = 0.029). C populations had intermediate diversity with no significant differences when compared to either M or P populations (p-value > 0.05).

In terms of among villages genetic differentiation, we observed significant mitochondrial Φst in M but not in P populations with the exception of the Khmu' (Fig. 2C). C populations with matrilocal (but not multilocal) residence also exhibited significant Φst. Thus, we found that all matrilocal populations (with either matrilineal or cognatic descent) had significant levels of mtDNA differentiation among villages.

On the other hand, within population Y haplotype diversity ranged from 0.81 to 1.0, with little variation between populations except for the Pacoh who had by far the lowest diversity (Fig. 2B). We observed no significant difference in Y chromosome diversity between social organizations (p-value > 0.05).

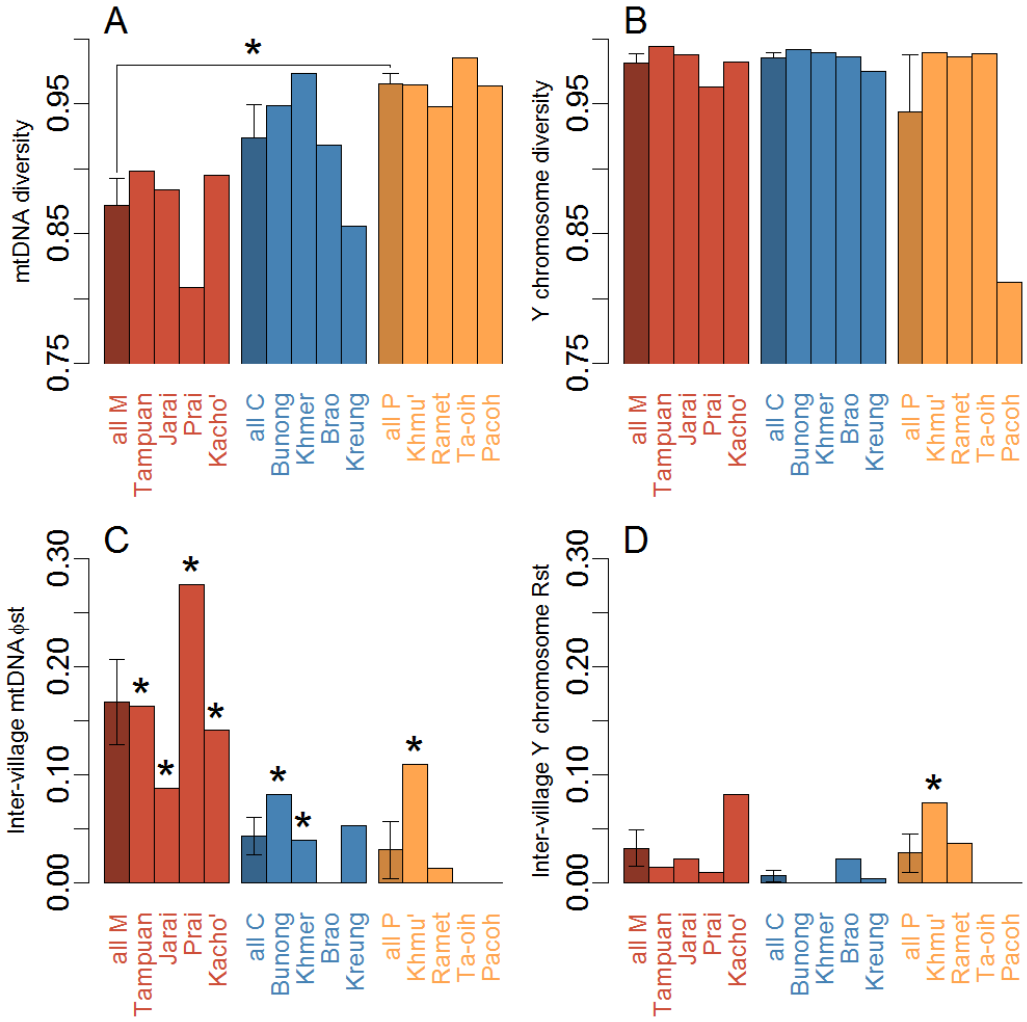


Figure 2: Uniparental genetic estimators. (A) mtDNA haplotype diversity; (B) Y chromosome haplotype diversity; (C) Inter-village mtDNA Φst based on pairwise differences between haplotypes; (D) Inter-village Y chromosome Rst based on repeat number differences. The first (darker) bar in each group represents the mean in each group with standard error. Asterisks indicate statistical significance (p-value < 0.05) assessed by MWW tests (A, B) or permutation tests (C, D)

In addition, very low Y chromosome Rst were observed among villages in all populations (Fig. 2D). The Khmu' were the only population to exhibit significant Rst (p-value = 0.013).

We checked that there was no saturation issue linked to the important number of STRs used here (17 STRs) by replicating our analyses on a restricted number of STRs (the 9 STRs of the minimal haplotype [47]). Similar conclusions were reached.

Influence of residence patterns on genetic diversity

We estimated the proportion of natolocal, matrilocal, patrilocal and neolocal couples in each population for the extended dataset (Fig. S1A) and the core dataset (Fig. S1B). In each population, the core dataset exhibited a lower proportion of natolocal couples than the extended dataset (on average 0.52 vs. 0.74). This lower proportion could be explained by a recent transition towards lower natolocality in the current generation (the extended dataset comprises a larger proportion of elder people than the core data set) or by a memory bias (erroneous assignations of birth places for parents and grandparents by the interviewee). However, the relative proportion of neolocal, patrilocal and matrilocal couples were similar in the core and extended datasets for 10 out of 12 populations (χ^2 test p-value > 0.05), and conclusions drawn from both datasets were similar. Hence, we focused here on the results provided by the extended dataset.

Natolocality was the major type of residence in every population, ranging from 0.47 to 0.92, although it was lower in P populations than in M (p-value =0.011) and C populations (p-value = 0.037).

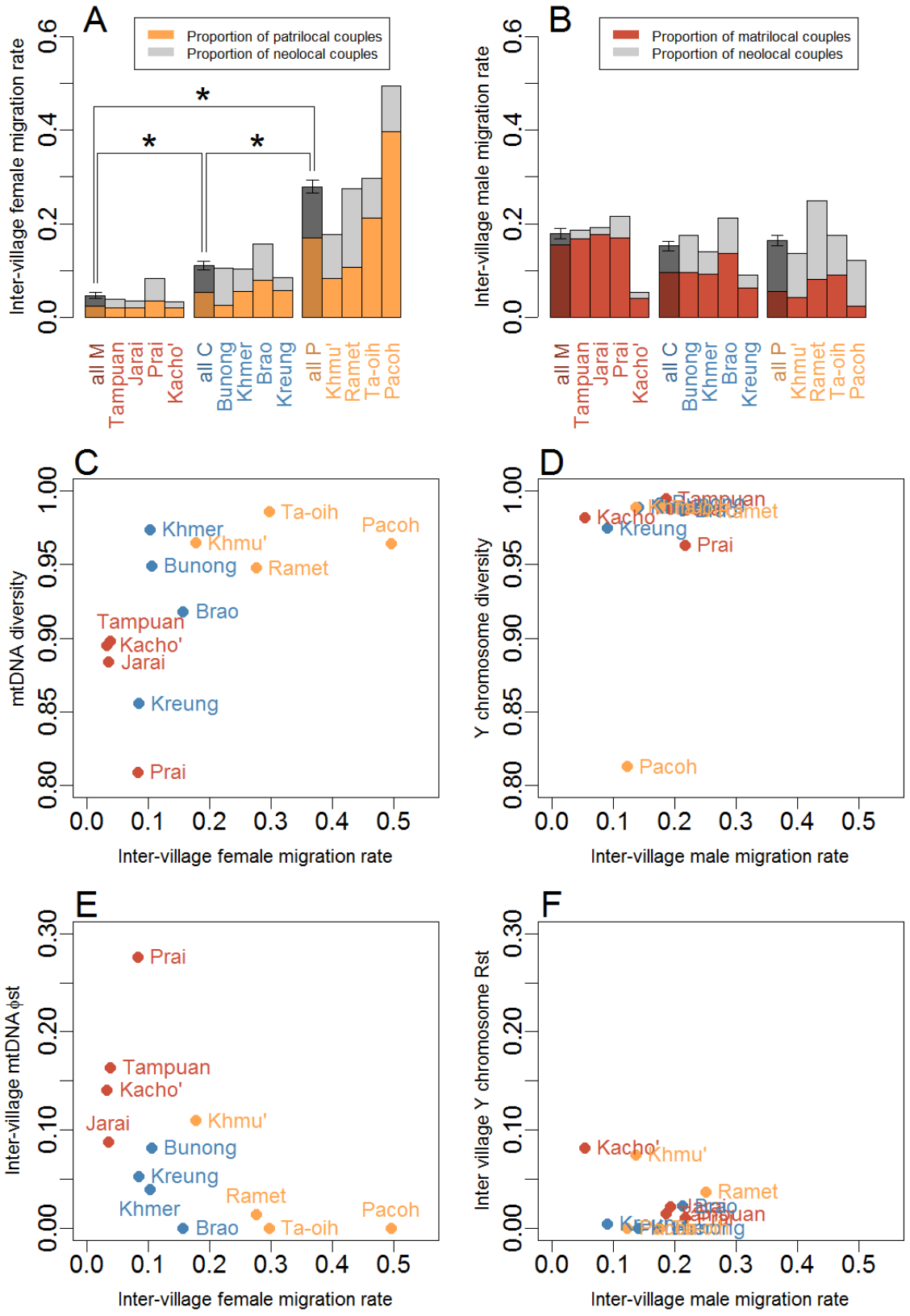


Figure 3: Sex-specific inter-village migration rates estimated from our ethno-demographic data on 3530 couples (extended data set) and correlation with genetic estimators: (A) Female migration rate; (B) male migration rate; (C) female migration rate against mtDNA diversity (ρ = 0.71, p-value = 0.012); (D) male migration rate against Y chromosome diversity (ρ = 0.098, p-value = 0.77); (E) female migration rate against mtDNA Φst (ρ = -0.76, p-value < 0.01); (F) male migration rate against Y chromosome Rst (ρ = 0.13, p-value = 0.68). The first (darker) bar in each group represents the mean in each group with standard error. Asterisks indicate statistical significance (p-value < 0.05) assessed by logistic regression model.

As expected, P populations had a lower proportion of matrilocal couples than M (p-value < 10^{-3}) and C populations (p-value = 0.014) (Fig. 3B). P populations had also a higher proportion of patrilocal couples than M (p-value < 10^{-3}) and C populations (p-value < 10^{-3}), thus confirming the residence rules described in ethnographical studies (Fig. 3A). Proportion of neolocal couples was lower in M populations than in C (p-value < 10^{-3}) and P populations (p-value <10^{-3}).

We estimated male and female inter-villages migration (Fig. 3A, 3B). As expected, female migration rates were significantly different among social organizations, with P populations having the highest rate compared to M and C populations. C populations had also a higher female migration rate than M populations (all 3 comparisons p-value < 10^{-3}). On the other hand, no differences in male migration rates were observed across social organizations (p-value > 0.05 for all 3 comparisons).

Finally, female migration rate between villages correlated positively with mitochondrial diversity (Fig. 3C, Spearman's $\rho = 0.71$, p-value = 0.012) and negatively with among villages Φst (Fig. 3E, Spearman's $\rho = -0.76$, p-value < 0.01). On the other hand, we did not observe significant correlations between male migration rate and Y chromosome diversity (Fig. 3D Spearman's $\rho = 0.098$, p-value = 0.77) or among villages Rst (Fig. 3F, Spearman's $\rho = 0.13$, p-value = 0.68) and these results were robust to the removal of the Pacoh who had a much lower Y chromosome diversity than other populations.

Influence of descent group structure on genetic diversity

We compared the number of clans of each population. For M populations, Prai had a much higher number of clans (47) than the other populations (9, 8, 17 in the Tampuan, Jarai and Kacho' respectively). Pacoh also had a higher number of clans (57) compared to the other P populations (17, 11 and 29 in Khmu', Ramet and Ta-oih respectively). There was no significant difference in the number of clans between M and P populations (p-value = 0.38).

We further explored differences in descent group structure between populations by comparing the clan affiliation knowledge rate of each population on maternal and paternal branches. As expected, all matrilineal populations had a higher clan affiliation knowledge on the maternal branches compared to the paternal branches (Fig. 4A). The opposite was true for patrilineal populations except for the Khmu' who knew both branches similarly well. Moreover, M populations had a significantly higher rate of clan knowledge on their maternal branches than P populations had on their paternal branches (0.94 versus 0.85; p-value < 0.01), suggesting a higher rigidity of the clanic structure in the M than in the P populations.

For both M and P populations, we observed that a higher number of clans, as observed in the Prai and the Pacoh, was associated with lower mitochondrial and Y chromosome diversity respectively (Fig. 4B, 4C). On the other hand, no consistent pattern was observed between the clan knowledge of ascendants and uniparental diversity across M and P populations (Fig. 4D, 4E).

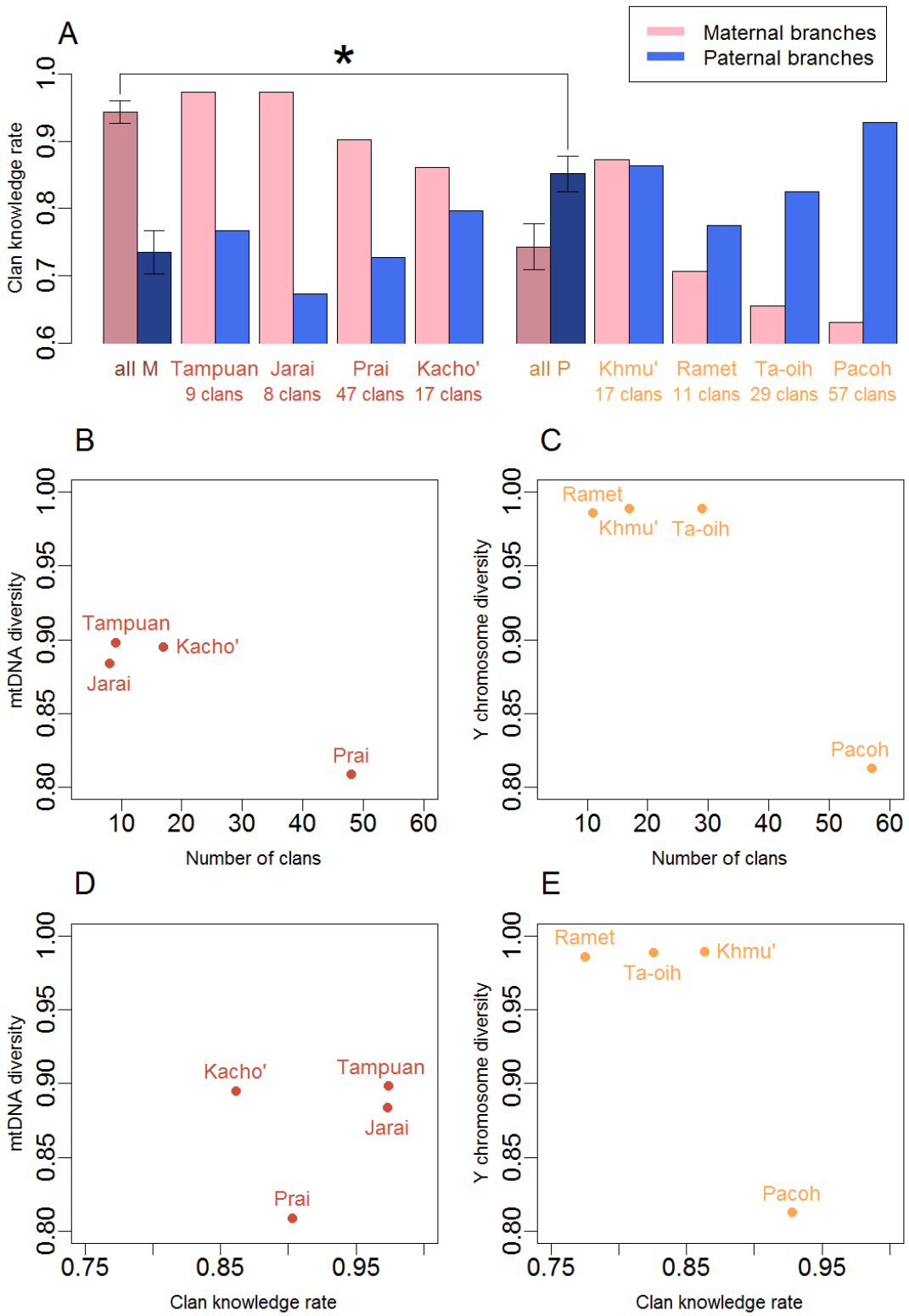


Figure 4: Descent group structure estimators and relationship with genetic diversity. (A) Clan knowledge rate of ascendants. Pink bars represent knowledge rate on the maternal branches (mothers and maternal grandmothers of the interviewee and of his/her spouse), blue bars represent knowledge rate on the paternal branches (fathers and paternal grandfathers of the interviewee and of his/her spouse). The first (darker) bars of each group represent the mean knowledge rate with standard error in each group. Asterisks indicate statistically significant differences between social organizations obtained from the logistic regression model (p-value <0.05); (B) mtDNA diversity against number of clans in M populations; (C) Y chromosome diversity against number of clans in P populations; (D) mtDNA diversity clan affiliation knowledge rate on the maternal branches in M populations; (E) Y chromosome diversity against clan affiliation knowledge rate on the paternal branches in P populations.

DISCUSSION

Our investigation of uniparental genetic patterns in South-East Asian populations with various social organizations showed a clear signature of matrilocality and matrilinearity (M) on mitochondrial genetic patterns, but no signature of patrilocality and patrilineality (P) on Y chromosome genetic patterns. Indeed, while mtDNA diversity was lower and inter-village genetic distances higher in M than in P populations, neither reduction in Y chromosome diversity nor increase in inter-village genetic distances were observed in P populations compared to M populations. These findings are similar to uniparental genetic patterns observed in matrilocal and patrilocal populations from Sumatra (Gunnarsdóttir et al., 2011). Cognatic (C) populations exhibited genetic estimators intermediate between M and P populations.

It is usually assumed that matrilocal populations have higher male migration rates and lower female migration rates compared to patrilocal populations (Besaggio et al., 2007; Gunnarsdóttir et al., 2011; Kumar et al., 2006; Oota et al., 2001) and we usually expect to observe sex-specific genetic signature in accordance with this. For Thailand Hill Tribes, this assumption seems to be accurate (Hamilton et al., 2005; Oota et al., 2001). However, our ethno-demographic data showed that this is not the case for the populations we sampled in South-East Asia. While female migration rate did vary with social organization and was correlated to mitochondrial patterns, male migration rate did not vary with social organization: the same proportion of men migrated away from their birth village in P populations and in M populations. We propose that this lack of difference in male migration rates between social organizations is the main reason for the lack of difference in Y chromosome diversity estimators.

Before going further, we should underscore that populations described as patrilocal by previous ethnographic works had indeed an excess of patrilocal couples (in comparison to matrilocal and neolocal couples), confirming our classification into M, C and P populations. However, our ethno-demographic dataset highlighted a higher residence flexibility in P populations than in

M populations. Indeed, P populations tolerated a higher proportion of matrilocal couples than the proportion of patrilocal couples tolerated by M populations : in P populations the proportion of matrilocal couples divided by the proportion of patrilocal and matrilocal couples was 23.1% versus 14% for the proportion of patrilocal couples divided by the same denominator in M populations, Consequently, male migrations was increased in P populations in comparison to what would be expected under a stricter patrilocal rule.

In addition, a higher proportion of neolocal couples was observed in P populations (0.11) in comparison to M populations (0.024, p-value $< 10^{-3}$). Consequently, P populations had more frequent inter-villages migration, for both males and females, compared to M populations. This higher proportion of neolocal couples in P than in M populations was observed (p-value < 0.01) even when estimated only on the oldest generations of our dataset (i.e. parents and grandparents). This argues for the fact that this behavior existed before recent political and social changes. Previous ethnographic works on Khmu' and Ramet showed that a house or a group of houses may move to a new village, and their members may initiate or integrate a new lineage, through a ritual (Evrard, 2006), suggesting that such flexibility may be a common practice in these patrilineal populations.

Thus, such increased flexibility (tolerance for matrilocality and neolocality) in P populations enhanced male migrations in P populations, making them as frequent as in M populations, despite the patrilocal residence rule. Consequently, this analysis showed that patrilocality and matrilocality are not just the opposite sides of the same coin, but different regulatory processes modulate male and female migration rates in these populations.

Interestingly, our higher residence flexibility for P than for M populations contrasted with the tighter social regulation observed in patrilocal compared to matrilocal Hill Tribe populations from Thailand using an approximate Bayesian computation method based on genetic data (Hamilton et al., 2005). These contrasted observations could correspond to genuine differences

existing between these unilineal populations depending on the socio-cultural and political context. Alternatively, they could result from different study scales (we estimated migration rates between villages within population while Hamilton et al. (Hamilton et al., 2005) focused on migration rates between populations) or to different time scales (our migration rates reflect contemporary rates, while migration rates estimated from genetic data are more likely to reflect long-term rates).

Moreover, our data highlighted that the descent rule, a social rule largely ignored in population genetic studies, could also influence the genetic diversity. Indeed, the Prai had a singularly low mtDNA diversity while having no reduction of female migration rate in comparison to other M populations. Likewise, the Pacoh had a low Y chromosome diversity while having no reduction in male migration rates in comparison to other P populations. These reductions in uniparental genetic diversity could be associated with the larger number of clans observed in these two populations in comparison to other P and M populations. It had already been shown that the descent group structure can influence genetic diversity patterns (Chaix et al., 2007; Montinaro et al., 2016; Sanchez-Faddeev et al., 2013; Ségurel et al., 2008). In Central Asia, Y chromosome diversity is reduced in patrilineal populations in comparison to cognatic populations (both having patrilocal residence) (Chaix et al., 2007). The dynamics of lineal clan fissions in which a clan splits into two clans regrouping closely related individuals, combined with extinctions of patrilineal clans, is likely to be responsible for this reduction in Y chromosome diversity. Theoretical works from Smouse et al. (Smouse, Vitzthum, & Neel, 1981) have indeed showed that lineal fissions combined with clan extinctions may decrease the effective size of the population by a factor up to 4. In the South-East Asian populations studied here, clans in M and P populations such as Jarai and Khmu' have been documented as staying stable over time and thus seems to not undergo lineal fissions (Dournes, 1972; Evrard, 2006). However, in Pacoh and Prai for which clan dynamics have not been documented, the much higher number of clans

suggests a lineal fission process, which may be responsible for the lower uniparental genetic diversity observed in these two populations.

In addition to descent group structures, other factors such as sex differences in variance or heritability in reproductive success could also play a part in shaping the uniparental genetic patterns by influencing the level of genetic drift on uniparental DNA (Heyer et al., 2012). Further work on the reproductive history of the individuals is needed in order to assess the variance in reproductive success among individuals and its heritability across generations. However, polygyny, one of the main factor linked to sex-specific variance in reproductive success was found at low rates in all 12 populations with a mean of 1.4% (SD = 3.0%) and thus it is doubtful that it had a meaningful impact in the studied populations.

To conclude, our study shows that fine quantification of social organization was necessary to better grasp its impact on genetic diversity and solve apparent discrepancies between social rules and genetic patterns as some authors have previously suggested (Destro Bisol et al., 2012; Guillot et al., 2015; Marks, Levy, Martinez-Cadenas, Montinaro, & Capelli, 2012; Shenk & Mattison, 2011; Wilkins, 2006). Moreover, it highlights the need to consider the different components of social organization altogether (residence but also descent and alliance rules), in order to better understand its impact on genetic diversity.

Ethics

All participants provided written informed consents and the study was approved by the National Ethic Comities for Health Research in Cambodia and Laos as well as by the Comité Opérationnel pour l'Ethique (CNRS).

Competing interest

The authors declare no competing interests.

Data deposition

Mitochondrial DNA HV1 sequences are provided in table S1. Y chromosome STR genotypes are provided in table S2. Y chromosome Data will be released on the appropriate databank upon acceptance of the manuscript.

Acknowledgements

We thank all the volunteers for their participation, as well as Yves Buisson from the Institut de la Francophonie pour la Médecine Tropicale (IFMT) in Laos for his help in facilitating field work.

Authors' contributions

R.C., S.P. and B.T. initiated the project; R.C., S.P., R.L, S.L., O.E., F.B. G.D and C.M. contributed to the ethno-demographic and genetic sampling; S.L. and C.M. carried out the lab work; G.L., B.A. and R.L. analyzed the data; G.L., R.C. and S.P. wrote the paper. All authors gave final approval for publication.

Funding

This work was supported by the ANR SoGen (JC09_441218) grant.

References

Austerlitz, F., & Heyer, E. (1998). Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15140–15144. http://doi.org/10.1073/pnas.95.25.15140

Baird, I. G. (2008). *Various Forms of Colonialism : the Social and Spatial Reorganisation of the Brao in Southern Laos and Northeastern Cambodia*. University of British Columbia.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1). http://doi.org/10.18637/jss.v067.i01

Besaggio, D., Fuselli, S., Srikummool, M., Kampuansai, J., Castrì, L., Tyler-Smith, C., … Bertorelle, G. (2007). Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evolutionary Biology*, *7 Suppl 2*(Suppl 2), S12. http://doi.org/10.1186/1471-2148-7-S2-S12

Bourdier, F. (2006). *The Mountain of Precious Stones*. Ratanakiri: Cambodia: Essays in Social Anthropology: Phnom Penh: Center for Khmer Studies.

Carrasco, P. (1963). The Locality Referent in Residence Terms. *American Anthropologist*, *65*(1), 133–134. http://doi.org/10.1525/aa.1963.65.1.02a00100

Casselberry, S., & Valavanes, N. (1976). “matrilocal” Greek peasants and a reconsideration of residence terminology. *American Ethnologist*, *3*(2), 215–226. http://doi.org/10.1525/ae.1976.3.2.02a00030

Chaix, R., Quintana-Murci, L., Hegay, T., Hammer, M. F., Mobasher, Z., Austerlitz, F., & Heyer, E. (2007). From social to genetic structures in central Asia. *Current Biology*, *17*(1), 43–48. http://doi.org/10.1016/j.cub.2006.10.058

Dessaint, W. Y. (1981). The T’in (Mal), dry rice cultivators of Northern Thailand. *Journal of the Siam Society*, *69*, 107–137.

Destro Bisol, G., Capocasa, M., & Anagnostou, P. (2012). When gender matters: New insights into the relationships between social systems and the genetic structure of human populations. *Molecular Ecology*, *21*(20), 4917–4920. http://doi.org/10.1111/mec.12001

Dournes, J. (1972). *Coordonées, structures jörai familiales et sociales* (Vol. 77). Paris: Muséum national d’Histoire naturelle (Travaux et Mémoires de l’Institut d’Ethnologie

n°77).

Ebihara, M. (1977). Residence Patterns in a Khmer Peasant Village. *Annals of the New York Academy of Sciences*, *293*(1), 51–68. http://doi.org/10.1111/j.1749-6632.1977.tb41805.x

Evrard, O. (2006). *Chroniques des cendres*. Paris: IRD Edition.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. http://doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491. http://doi.org/10.1007/s00424-009-0730-7

Godelier, M. (2004). *Métamorphoses de la parenté*. Paris: Fayard.

Guillot, E. G., Hazelton, M. L., Karafet, T. M., Lansing, J. S., Sudoyo, H., & Cox, M. P. (2015). Relaxed Observance of Traditional Marriage Rules Allows Social Connectivity without Loss of Genetic Diversity. *Molecular Biology and Evolution*, *32*(9), 2254–2262. http://doi.org/10.1093/molbev/msv102

Gunnarsdóttir, E. D., Nandineni, M. R., Li, M., Myles, S., Gil, D., Pakendorf, B., & Stoneking, M. (2011). Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. *Nature Communications*, *2*, 228. http://doi.org/10.1038/ncomms1235

Hamilton, G., Stoneking, M., & Excoffier, L. (2005). Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proceedings of the National Academy of Sciences of the United States of America*,

102(21), 7476–7480. http://doi.org/10.1073/pnas.0409253102

Hammer, M. F., Mendez, F. L., Cox, M. P., Woerner, A. E., & Wall, J. D. (2008). Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genetics*, *4*(9), e1000202. http://doi.org/10.1371/journal.pgen.1000202

Heyer, E., Chaix, R., Pavard, S., & Austerlitz, F. (2012). Sex-specific demographic behaviours that shape human genomic variation. *Molecular Ecology*, *21*(3), 597–612. http://doi.org/10.1111/j.1365-294X.2011.05406.x

Izikowitz, K. G. (1951). *Lamet hill peasants in French Indochina*. Göteborg: Etnografiska Museet.

Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., … Stoneking, M. (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *American Journal of Human Genetics*, *72*(2), 281–302. http://doi.org/10.1086/346065

Kumar, V., Langstieh, B. T., Madhavi, K. V., Naidu, V. M., Singh, H. P., Biswas, S., … Reddy, B. M. (2006). Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genetics*, *2*(4), 420–424. http://doi.org/10.1371/journal.pgen.0020053

LeBar, M., Hickey, G. C., & Musgrave, J. K. (1964). *Ethnic groups of mainland Southeast Asia*. Human Relations Area Files.

Ledgerwood, J. L. (1995). Khmer Kinship - the Matriliny Matriarchy Myth. *Journal of Anthropological Research*, *51*(3), 247–261.

Lindell, K., Samuelsson, R., & Tayanin, D. (1979). Kinship and Marriage in Northern Kammu Villages: The Kinship Model. *Sociologus*, *29*(1), 60–84.

Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., … Stoneking, M. (2014). Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genetics*, *5*(1), 13. http://doi.org/10.1186/2041-2223-5-13

Maquet, J. (n.d.). Clan. Retrieved March 3, 2017, from http://www.universalis.fr/encyclopedie/clan/

Marks, S. J., Levy, H., Martinez-Cadenas, C., Montinaro, F., & Capelli, C. (2012). Migration distance rather than migration rate explains genetic diversity in human patrilocal groups. *Molecular Ecology*, *21*(20), 4958–4969. http://doi.org/10.1111/j.1365-294X.2012.05689.x

Martel, G. (1975). *Lovea. Village des environs d'Angkor. Aspects démographiques, économiques et sociologiques du monde rural cambodgien dans la province de Siem-Réap*. Paris: Ecole française d'Extrême-Orient.

Matras-Troubetzkoy, J. (1983). *Un Village En Foret. L'essartage Chez Les Brou Du Cambodge*. (Selaf, Ed.). Paris: Peeters.

Michalakis, Y., & Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, *142*(3), 1061–1064. http://doi.org/10.1520/D0850-11.1

Montinaro, F., Davies, J., & Capelli, C. (2016). Group membership, geography and shared ancestry: Genetic variation in the Basotho of Lesotho. *American Journal of Physical Anthropology*, *160*(1), 156–161. http://doi.org/10.1002/ajpa.22933

Neel, J. V. (1970). Lessons from a "primitive" people. *Science*, *170*(960), 815–822. http://doi.org/10.1126/science.170.3960.815

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., & Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genetics*, *29*(1), 20–1. http://doi.org/10.1038/ng711

Pasternak, B., Ember, C. R., & Ember, M. (1976). On the Conditions Favoring Extended Family Households. *Journal of Anthropological Research*, *32*(2), 109–123. http://doi.org/10.2307/3629658

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing.

Sanchez-Faddeev, H., Pijpe, J., van der Hulle, T., Meij, H. J., van der Gaag, K. J., Slagboom, P. E., … de Knijff, P. (2013). The influence of clan structure on the genetic variation in a single Ghanaian village. *European Journal of Human Genetics*, *21*(10), 1134–1139. http://doi.org/10.1038/ejhg.2013.12

Schmutz, J. (2013). The Ta'Oi Language and People. *Mon-Khmer Studies*, *42*, i–xiii.

Ségurel, L., Martínez-Cruz, B., Quintana-Murci, L., Balaresque, P., Georges, M., Hegay, T., … Vitalis, R. (2008). Sex-specific genetic structure and social organization in Central Asia: Insights from a multi-locus study. *PLoS Genetics*, *4*(9), e1000200. http://doi.org/10.1371/journal.pgen.1000200

Seielstad, M. T., Minch, E., & Cavalli-Sforza, L. L. (1998). Genetic evidence for a higher female migration rate in humans. *Nature Genetics*, *20*(3), 278–280. http://doi.org/10.1038/3088

Shenk, M. K., & Mattison, S. M. (2011). The Rebirth of Kinship. *Human Nature*, *22*(1), 1–15. http://doi.org/10.1007/s12110-011-9105-9

Smouse, P. E., Vitzthum, V. J., & Neel, J. V. (1981). The impact of random and lineal fission

on the genetic divergence of small human groups: A case study among the Yanomama. *Genetics*, *98*(1), 179–197.

Tumonggor, M. K., Karafet, T. M., Downey, S., Lansing, J. S., Norquest, P., Sudoyo, H., … Cox, M. P. (2014). Isolation, contact and social behavior shaped genetic diversity in West Timor. *Journal of Human Genetics*, *59*(9), 494–503. http://doi.org/10.1038/jhg.2014.62

UNDP Cambodia. (2010a). *Kreung Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village*. Phnom Penh.

UNDP Cambodia. (2010b). *Phnong Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village*. Phnom Penh.

Wilkins, J. F. (2006). Unraveling male and female histories from human genetic data. *Current Opinion in Genetics and Development*, *16*(6), 611–617. http://doi.org/10.1016/j.gde.2006.10.004

Wilkins, J. F., & Marlowe, F. (2006). Sex-biased migration in humans: what should we expect from genetic data? *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *28*(3), 290–300. http://doi.org/10.1002/bies.20378

Supplementary figure

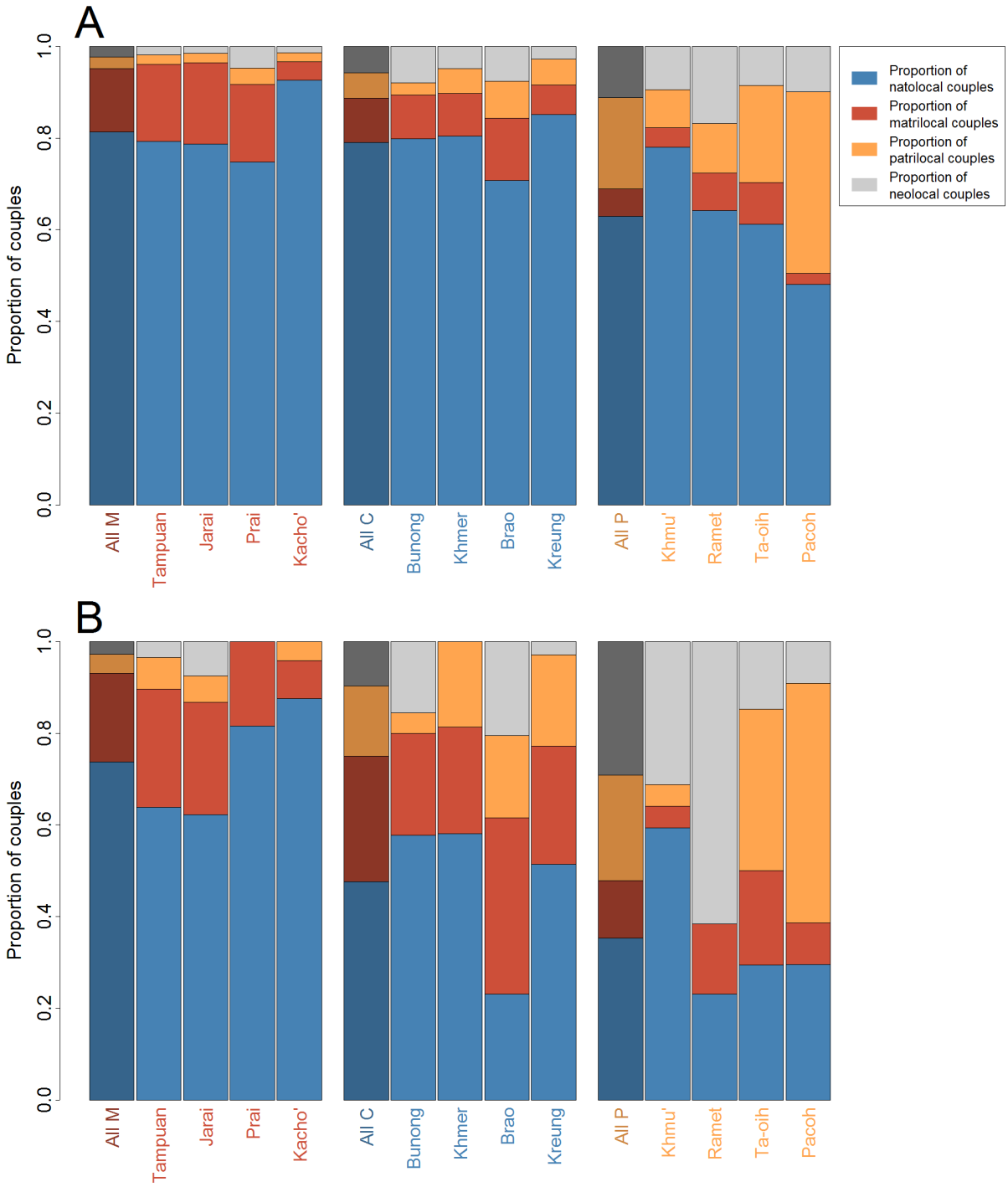


Figure S1 : Proportions of natolocal (blue), matrilocal (red), patrilocal (green) and neolocal (gray) couples for each population for the extended dataset (3530 couples) (A) and the core dataset (532 couples) (B). The first (darker) bar in each group represents the mean in each group.

3. Discussion

Dans cet article, nous avons montré que les systèmes de parenté, en influant sur les migrations matrimoniales sexe-spécifiques et la dynamique des groupes de filiation, impactent la diversité génétique des populations humaines.

3.1 L'influence de la flexibilité de la règle de résidence sur la diversité uniparentale

En effet, la diversité de l'ADN mitochondrial est plus faible dans les populations M que dans les populations P, en accord avec nos données ethno-démographiques, montrant que les femmes migrent moins dans les populations M que dans les populations P. En revanche, les populations étudiées présentent des niveaux de diversité du chromosome Y similaires quel que soit leur système de parenté. Cette observation peut s'expliquer par la flexibilité plus grande des populations P quant au respect de la règle de résidence induisant des taux de migration d'hommes similaires entre les systèmes de parenté.

Nous observons donc une signature de la matrilocalité sur la diversité de l'ADN mitochondrial mais non de la patrilocalité sur le chromosome Y. Cette observation est similaire à celle faite par Gunnarsdottir et al. pour les populations matrilocales et patrilocales de l'île de Sumatra (2011). Les auteurs avaient émis l'hypothèse d'une plus grande flexibilité de la résidence chez la population patrilocale pour expliquer cette observation. Nos résultats ethno-démographiques en Asie du Sud-Est continentale soutiennent cette hypothèse. Pourtant, cette flexibilité plus grande de la règle de résidence chez les populations P était inattendue car elle est contraire à l'attendu du « puzzle matrilinéaire ». Hamilton et al. avait bien mis en évidence à partir de données génétiques, une flexibilité de la règle de résidence plus grande chez les populations matrilocales (Ahka et deux groupes de Lisu) que chez les populations patrilocales (Lahu, Karen Rouge et Karen Blanc) dans les populations de Thaïlande (Hamilton et al., 2005). Les auteurs ont interprété ce résultat comme la résultante d'un pouvoir décisionnel concernant le choix de la résidence post-maritale plus important des hommes dans les populations matrilocales que celui des femmes dans les populations patrilocales. Cette dominance masculine dans le choix de la résidence est une hypothèse proche de celle du « puzzle matrilinéaire » fondé sur la quasi-universalité de l'autorité masculine.

Nous proposons trois hypothèses pour expliquer l'absence de congruence entre notre étude et celle d'Hamilton et al. La première serait que les populations matrilinéaires et matrilocales étudiées dans cette thèse auraient adopté une autre « solution » pour résoudre le « puzzle

matrilinéaire ». Richards (1950) présentait dans son étude différents types de populations matrilinéaires ayant adopté des mécanismes différents pour faire face aux enjeux du « puzzle matrilinéaire », la flexibilité de la règle de résidence n'étant qu'un de ces mécanismes. Ainsi, le chapitre 1 de cette thèse a mis en évidence que les populations matrilinéaires et matrilocales étudiées résolvent en grande partie le « puzzle matrilinéaire » en se mariant au sein de leur village. Cette endogamie de village permet aux hommes de respecter la règle de résidence matrilocale en habitant avec la famille de leur femme tout en restant à proximité de leur groupe de filiation maternel, ce qui leur permet de maintenir leur position d'autorité.

La seconde hypothèse serait que la flexibilité de la règle de résidence resterait un processus possible mais rarement choisi par les hommes pour faire face aux enjeux du « puzzle matrilinéaire ». En effet, Richards (1950), qui a étudié les populations Bantous d'Afrique Centrale , a montré que la flexibilité de la résidence, qui peut exister dans les populations de type B1 ou B2, ne prend pas toujours effet. Ainsi dans les populations de type B1, nommées aussi le groupe Bemba-Bisa-Lamba, un homme pouvait ramener sa femme dans son village après avoir vécu chez ses beaux-parents et fourni sa force de travail pendant un nombre d'années indéterminé. Cependant, il est précisé que si l'homme se trouvait dans une position confortable dans le village de sa femme, il n'exerçait généralement pas ce droit et ne retournait pas dans son village natal. Dans les populations de type B2, ou groupe Yao-Cewa, le droit de ramener sa femme dans son village natal n'était octroyé qu'au chef du groupe de filiation (le plus souvent le frère aîné ayant autorité sur ses sœurs). Ainsi, il semblerait que la possibilité pour un homme de s'installer avec sa femme patrilocalement dans une population matrilocale prenne rarement effet ou se limite à un petit nombre d'hommes ayant des positions d'autorité particulières. En conséquence, il semble que présenter une forte flexibilité de la règle de résidence ne soit pas généralisable à l'ensemble des populations patrilocales.

En outre, Hamilton et al. ont quantifié dans leur étude les flux migratoires sexe spécifiques entre ethnies (et non, comme dans cette thèse, entre villages au sein de la même ethnie) : il est donc possible que la propension d'un homme à s'installer avec sa femme de façon patrilocale dans une population matrilocale soit augmentée en cas de mariages inter-ethniques mais ne représente pas un processus fréquent lors de mariages intra-ethniques, comme ceux étudiés dans notre étude.

3.2 La difficulté de départager les influences de la résidence et de la filiation

Nous avons également montré qu'il est difficile de faire la part entre l'influence de la résidence et de la filiation sur la diversité génétique. En effet, les variables de résidence et de filiation sont corrélées entre elles dans les populations étudiées (voir chapitre 1) : plus une population est matrilinéaire, plus elle est matrilocale et plus une population est patrilinéaire, plus elle est patrilocale. Les taux de migration de femmes et d'hommes peuvent donc ne pas dépendre uniquement de la règle de résidence, mais peuvent aussi être affectés par d'autres facteurs comme la filiation. La comparaison entre les populations cognatiques et matrilocales d'une part et les populations matrilinéaires et matrilocales d'autre part aurait pu nous permettre de départager l'effet de la résidence et de la filiation. Ces analyses se sont révélées infructueuses car les populations cognatiques et matrilocales de notre étude ont des taux de migration de femmes et des niveaux de diversité mitochondriale supérieurs à ceux des populations matrilinéaires et matrilocales, sans qu'il soit possible de dire s'il s'agit d'un effet de la résidence ou d'absence de groupes de filiation maternels. Il est donc possible que la signature de la matrilocalité observée sur les niveaux de diversité mitochondriale et l'absence de signature de la patrilocalité observée sur les niveaux de diversité du chromosome Y dans notre étude ne correspondent pas uniquement à des flexibilités variables dans le respect de la règle de résidence mais également à des dynamiques différentes des groupes de filiation entre les populations P et M. En effet, nous avons montré que le taux de connaissance des clans de la branche paternelle par les individus des populations P tend à être moins important que celui des clans de la branche maternelle par les individus des populations M. Ceci pourrait témoigner d'une moins grande rigidité de ces groupes de filiation chez les populations P. Cette moindre rigidité pourrait contribuer à limiter l'impact de la sous structuration en clans sur la réduction de la diversité génétique du chromosome Y.

3.3 L'influence des fissions de clans sur la diversité uniparentale

Enfin, notre étude ethno-démographique (chapitre 1) avait également mis en évidence des dynamiques de clans spécifiques à certaines populations. En particulier, nous avons observé un nombre de clans particulièrement élevé chez les Prai (matrilinéaires) et les Pacoh (patrilinéaires). L'étude de leur diversité uniparentale a souligné la diversité mitochondriale particulièrement faible des Prai par rapport aux autres populations matrilinéaires et la diversité du chromosome Y particulièrement faible des Pacoh par rapport aux autres populations patrilinéaires. Il pourrait s'agir de la signature de l'existence d'une dynamique de

fission linéaire des clans particulièrement active dans ces populations entraînant une diminution de leur taille efficace (Smouse et al., 1981) et donc de leur diversité génétique sexe-spécifique.

Chapitre 3 - Des variables ethno-démographiques à la diversité autosomale

1. Introduction

Dans le chapitre précédent, nous avons évalué l'impact des systèmes de parenté sur la diversité uniparentale. Un de ces objectifs était de comprendre les raisons de l'absence de consensus parmi le grand nombre d'études en génétique des populations sur l'impact des systèmes de parenté sur la diversité uniparentale.

En revanche, peu d'études se sont intéressées à l'impact des systèmes de parenté sur la variabilité autosomale (Watkins, 2004; Guillot et al., 2015). Ainsi, nous voulons étudier dans ce chapitre, l'effet des systèmes de parenté à l'étude ici sur la diversité autosomale, et en particulier sur la consanguinité.

Nous avons montré dans le premier chapitre qu'il existait un taux d'endogamie de village plus élevé chez les populations matrilinéaires et matrilocales (M) par rapport aux populations patrilinéaires et patrilocales (P), probablement en lien avec le « puzzle matrilinéaire » chez les hommes des populations M. Or, une augmentation de l'endogamie de village entraîne une augmentation de la consanguinité (Gilbert et Hammel, 1966; Bittles, 2008). Nous nous attendons donc à observer une consanguinité plus élevée chez les populations M comparé aux populations P en relation avec la différence d'endogamie de village entre ces deux systèmes de parenté.

Par ailleurs, des études génétiques à l'échelle mondiale ont détecté une consanguinité plus élevée dans certaines régions du monde (Asie du Sud, Moyen-Orient, Amérique du Sud) où les mariages entre apparentés sont fréquents (Leutenegger et al., 2011; Pemberton et al., 2012a; Gazal et al., 2015). A la lumière de ces résultats, nous voulons tester si le taux des mariages entre cousins influence la diversité génétique des populations d'Asie du Sud-Est étudiées dans cette thèse. En effet, si notre recueil de données ethno-démographique n'a pas permis d'évaluer

les fréquences de mariages entre cousins au sein des différentes populations étudiées, il semblerait qu'à l'échelle mondiale, les populations unilinéaires aient une préférence pour les mariages entre cousins de premier degré plus importante que les populations cognatiques (41% contre 20%) (Murdock et White, 1969). Par ailleurs, selon la littérature ethnographique, 5 des 8 populations unilinéaires étudiées dans cette thèse (Tampuan, Jarai, Prai, Khmu' et Ramet) auraient une préférence pour les mariages entre cousins (Izikowitz, 1951; LeBar et al., 1964; Dournes, 1972; Bourdier, 2006; Evrard, 2006). Au contraire, aucune mention de préférence pour des mariages entre cousins de premier degré n'a été trouvée concernant les populations cognatiques à l'étude (LeBar et al., 1964; Ebihara, 1977; Baird, 2008; UNDP Cambodia, 2010a; b). Ainsi, il est probable que les populations unilinéaires étudiées ici aient une préférence plus importante pour les mariages entre cousins que les populations cognatiques, et de ce fait présentent une consanguinité plus élevée chez les populations unilinéaires que cognatiques.

Ainsi, dans ce chapitre, nous allons tester :

- L'effet de l'endogamie de village potentiellement liée au « puzzle matrilinéaire » sur la diversité autosomale et en particulier sur la consanguinité. Plus spécifiquement, nous allons tester si les populations M et C ont une consanguinité plus élevée que les P. Ce test fait l'objet d'un article en phase d'écriture finale présenté en annexe 4.
- L'effet de la préférence pour les mariages entre cousins chez les populations unilinéaires sur la diversité autosomale et plus particulièrement sur la consanguinité. Plus spécifiquement, nous allons tester si les unilinéaires ont une consanguinité plus élevée que les cognatiques du fait de leur supposé taux de mariages entre cousins premier degré plus élevé.

Pour ce faire, nous avons estimé la consanguinité génétique dans les populations à l'étude de deux manières différentes : le coefficient de consanguinité du logiciel FEstim (Leutenegger et al., 2003), et la longueur totale des segments d'homozygotie (*run of homozygosity* ou ROH) longs (Pemberton et al., 2012a).

Au préalable, nous avons réalisé un certain nombre d'analyses descriptives de la diversité génétique de ces populations, notamment par l'estimation des différences génétiques entre populations, et celle de leur taille efficace. Ce dernier estimateur est particulièrement important ici dans la mesure où l'histoire démographique des populations peut également façonner les profils de consanguinité génétique des populations (Pemberton et al., 2012a) et le test de nos

deux hypothèses nécessite de comparer des populations ayant des histoires démographiques semblables.

2. Matériels et méthodes

2.1 Description du jeu de données autosomal

Nous avons génotypé 753 individus dont les quatre grands-parents appartenaient à la même population. Notre stratégie d'échantillonnage consistait, lorsque cela était possible, à échantillonner des trios.

529 individus ont été génotypés sur une puce Illumina Omni1 et 226 individus sur une puce Illumina Omni2.5. 701 163 SNPs sont partagés par les deux puces. Après un contrôle qualité effectué par Romain Laurent (annexe 5), nous avons un jeu de donnée comprenant 598 764 SNPs pour 743 individus (Table 3-1).

Population	Nombre d'individus après QC	Nombre d'individus sans fratries	Nombre d'individus sans apparentés plus proches que cousins au second degré
Tampuan	101	96	57
Jarai	92	89	53
Prai	48	44	26
Kacho'	32	30	21
Bunong	88	84	49
Khmer	92	92	57
Brao	49	49	34
Kreung	51	50	34
Khmu'	72	68	44
Ramet	26	26	21
Ta-oih	28	28	21
Pacoh	64	63	40
Total	743	719	457

Table 3-1 : Taille des échantillons génétiques pour les 12 populations étudiées

Nous avons ensuite vérifié la présence de fratries dans notre jeu de donnée en utilisant la méthode PC-Relate (Conomos et al., 2016). En utilisant le coefficient de parenté entre paires d'individus ainsi que k_0, la probabilité d'avoir zéro allèle en commun identique par descendance (IBD), nous pouvons distinguer les relations parents-enfants des relations d'individus d'une même fratrie (Figure 3-1). En effet, ces deux relations ont toutes les deux un coefficient de

parenté de 0,25 mais diffèrent par leur k0 qui est respectivement de 0 et 0,25 pour les relations parents-enfants et les relations entre individus d'une même fratrie. La plupart des relations parents-enfants nous sont connues par avance, car elles proviennent de notre échantillonnage de trios, mais il nous faut néanmoins vérifier que nous n'avons pas échantillonné par erreur des parents et leurs enfants entre les trios échantillonnés (par exemple le père d'un trio peut aussi être le fils d'un individu compris dans un autre trio échantillonné).

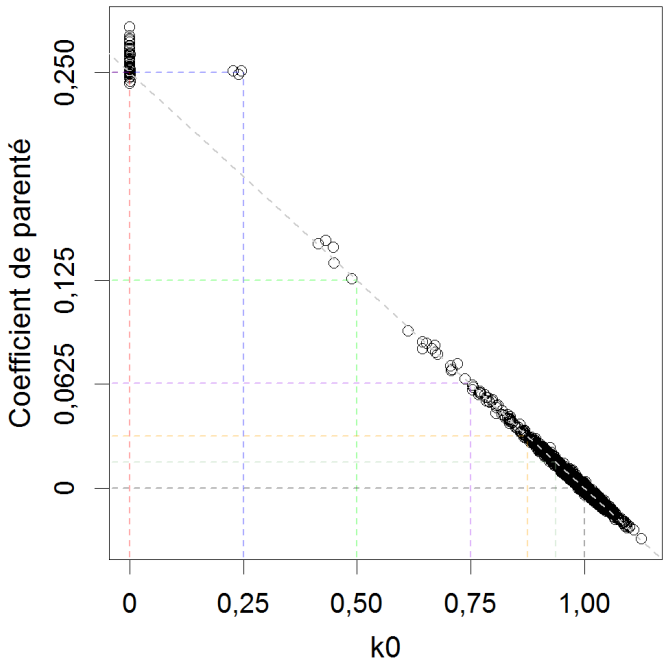


Figure 3-1 : Coefficient de parenté en fonction de k_0, la probabilité d'avoir zéro allèle en commun identique par descendance pour les paires d'individus chez les Jarai. Les relations parent-enfant et entre individus d'une même fratrie ont tous les deux un coefficient de parenté de 0,25, mais se distinguent en termes de k0. Les premiers ont un k_0 de 0 alors que les seconds un k0 de 0,25.

Lorsque deux individus d'une même fratrie étaient présents dans une population, nous avons retiré un des individus en essayant de préserver au maximum les trios puis les couples dans le jeu de données. Nous avons retiré 24 individus au total (Table 3-1). Le jeu de données final comporte 598 764 SNPs pour 719 individus.

Un jeu de données comprenant uniquement les 457 individus n'étant pas plus apparentés que des cousins de second degré a aussi été généré par Romain Laurent (Table 3-1). Pour cela, la première étape a consisté à éliminer tous les enfants du jeu de données. Ensuite, dans chaque population, le coefficient d'apparentement de Rousset (Rousset, 2002) a été estimé entre toutes les paires d'individus. Lorsqu'une paire avait un coefficient supérieur ou égal à 1/64, ce qui correspond à la valeur attendue entre cousins de second degré, un des individus a été retiré en

essayant de préserver au maximum les couples. Cette étape a été répétée jusqu'à ce qu'il n'y ait plus de nouvelles paires d'individus avec un coefficient supérieur ou égal à 1/64.

Les deux jeux de données génétiques ainsi générés (sans fratries, sans individus plus apparentés que des cousins de second degré) vont être utilisés dans les différentes analyses qui suivent. Les mesures de distance génétique (Fst et *allele sharing distance*) seront estimées sur le jeu de données sans individus plus apparentés que des cousins de second degré. La taille efficace et l'hétérozygotie seront aussi estimées sur ce même jeu de données. Enfin, pour les deux estimateurs de consanguinité (coefficient de consanguinité de FEstim et longueur totale des ROHs longs) nous utiliserons le jeu de données sans fratries.

2.2 Estimation de la différenciation génétique entre populations

L'indice de fixation (Fst) a été estimé entre les paires de populations en utilisant la méthode de Weir & Cockerham (Weir et Cockerham, 1984). Les SNPs ayant une fréquence de l'allèle mineur inférieure à 0,05 ont été exclus. Afin de visualiser la matrice des valeurs de Fst entre paires de populations, nous avons effectué une analyse de positionnement multidimensionnel (multidimensional scaling ou MDS) avec la fonction 'cmdscale' dans le logiciel R v3.2.2 (R Development Core Team, 2008). Dans le but d'évaluer dans quelle mesure les deux premiers axes de la MDS représentent bien la matrice des Fst, nous avons estimé le coefficient de corrélation de Spearman, ρ, entre les distances euclidiennes sur la MDS et les valeurs de Fst correspondantes entre toutes les paires de populations (Verdu et al., 2014).

2.3 Estimation de *l'allele sharing distance* entre individus

Nous avons estimé *l'allele sharing distance* (ASD) entre toutes les paires d'individus. Les enfants des trios ont été retirés du jeu de données pour cette analyse. Cette distance est définie entre les individus i et j comme :

$$D_{ij} = \frac{1}{L}\sum_{l=1}^{L} d_{ij(l)}$$

avec $d_{ij(l)} = \begin{cases} 0, si\ les\ individus\ i\ et\ j\ ont\ deux\ allèles\ en\ commun\ au\ locus\ l \\ 1, si\ les\ individus\ i\ et\ j\ ont\ un\ allèle\ en\ commun\ \ au\ locus\ l \\ 2, si\ les\ individus\ i\ et\ j\ n'ont\ pas\ d'allèleen\ commun\ \ au\ locus\ l \end{cases}$,

et *L* le nombre de SNP utilisés dans l'estimation. L'estimation a été effectuée sur les 365 325 SNPs pour lesquels aucun individu n'avait de génotypes manquants.

Nous avons utilisé ces distances dans une MDS par le biais de la fonction 'cmdscale' dans le logiciel R. La distance moyenne de lien, L_0, entre chaque paire de populations a été estimée (Timm, 2002; Kopelman et al., 2009). Cette distance correspond à la moyenne des distances entre toutes les paires formées par un individu dans la première population et un individu dans la seconde population. Nous avons estimé une p-value pour cette distance en comparant la valeur obtenue à une distribution nulle de L_0 obtenue en permutant les individus entre les deux populations 10 000 fois. La p-value correspond à la proportion de valeurs dans la distribution nulle supérieures ou égales à la valeur observée.

2.4 Estimation de la taille efficace des populations

La taille efficace d'une population a une influence sur la force de la dérive dans cette population. Or, la dérive a une influence sur la consanguinité. Comme nous l'avons expliqué en introduction, à chaque génération, la dérive va réduire la diversité dans la population. Les individus vont donc être de plus en plus génétiquement similaires. Ainsi, même si les unions se font de manière aléatoire, la consanguinité va augmenter dans la population juste par l'effet de la dérive : c'est ce que Jacquard a nommé la consanguinité de dérive (Jacquard, 1968). Ainsi, dans le but d'isoler l'effet du système de parenté sur la consanguinité, nous voulons vérifier que nos populations ont des tailles efficaces similaires.

Ce travail d'estimation de la taille efficace a été effectué par Gaspard Baudrin, un étudiant inscrit en licence 3 à l'Université Paris Descartes ayant effectué un stage au sein de notre UMR 7206. Je décrirai succinctement la partie de son travail que j'ai utilisée pour comparer la taille efficace des populations d'Asie du Sud-Est.

Les tailles efficaces des populations ont été calculées en utilisant le logiciel 'rhomap', qui utilise la méthode dite de McVean (Auton et McVean, 2007). Cette méthode estime le taux de recombinaison au sein de la population sur chaque chromosome en utilisant une méthode fondée sur une chaîne de Markov à inférence bayésienne (Monte-Carlo Markov Chain ou MCMC). Cette méthode d'approximation de la vraisemblance permet d'obtenir la probabilité d'observer les données réelles sous un modèle donné dont nous connaissons les paramètres. Ainsi, nous pouvons inférer le taux de recombinaison sachant le taux de mutation.

La taille efficace est alors déduite utilisant la formule suivante :

$$\rho = 4Ne\, r$$

Avec ρ, le taux de recombinaison cumulé dans la population, calculé par 'rhomap', et r, le taux de recombinaison par génération extrait de la carte génétique Decode (Kong et al., 2002).

Cette approche fournit ainsi une estimation de la taille efficace de la population pour chaque chromosome.

Nous avons ensuite comparé la moyenne de la taille efficace par chromosome entre populations en utilisant le test de comparaison de moyenne de Welch et nous appliquons une correction de Bonferroni pour tests multiples.

2.5 Estimation de l'hétérozygotie haplotypique

L'hétérozygotie haplotypique a été estimée selon la méthode proposée par Li et al. (Li et al., 2008). Dans un premier temps, nous avons estimé le taux de recombinaison moyen entre les SNPs de notre jeu de données en utilisant les taux de recombinaison du jeu de données HapMap (Build GR37). Le taux de recombinaison entre deux SNPs présents dans notre jeu de données a été calculé comme le taux de recombinaison moyen entre tous les SNPs présents dans le même intervalle sur le jeu de donnée HapMap. Ces taux de recombinaison permettent de définir des blocs de 5 à 15 SNPs consécutifs pour lesquels le taux de recombinaison entre deux SNPs consécutifs est inférieur à 0.5 cM/Mb. Nous identifions 12 865 blocs remplissant les caractéristiques ci-dessus au total sur les 22 chromosomes avec une moyenne de 584,8 blocs (déviation standard de 319,9) par chromosome. Nous calculons ensuite pour chaque population, les fréquences haplotypiques de ces 12 865 blocs avec la fonction '–hap-freq' disponible sous PLINK (Purcell et al., 2007). Enfin, nous utilisons la formule de Nei pour estimer l'hétérozygotie haplotypique sur chaque chromosome (Nei et Tajima, 1981).

$$H = \frac{2N}{2N-1} \times \frac{1}{L} \sum_{l=1}^{L} (1 - \sum_{i} {p_{i(l)}}^2)$$

Avec N le nombre d'individus de la population, L le nombre de blocs sur le chromosome, et $p_{i(l)}$ la fréquence du i^{ième} haplotype pour le l^{ième} bloc.

La valeur finale de l'hétérozygotie est la moyenne sur les 22 chromosomes. Elle est comparée entre populations par un test de comparaison de moyenne de Welch et nous appliquons une correction de Bonferroni pour tests multiples.

2.6 Estimation du coefficient de consanguinité par FEstim

Nous avons estimé le coefficient de consanguinité à l'aide du logiciel FEstim (Leutenegger et al., 2003) intégré dans la pipeline FSuite (Steven et al., 2014). FEstim utilise une méthode fondée sur une chaîne de Markov cachée et son estimateur est fondé sur un maximum de vraisemblance. FEstim estime le coefficient de consanguinité sur des sous-cartes. Chaque sous-carte est un sous-échantillonnage de SNPs distant les uns des autres, ce qui permet de s'affranchir des problèmes de déséquilibre de liaison qui peuvent exister entre les marqueurs. L'estimateur donné est la valeur médiane des estimateurs sur les 100 différentes sous-cartes. Nous avons utilisé la fonction '–hotspots' qui prend en compte les points chauds de recombinaison lors de la création des sous-cartes. Les fréquences alléliques ont été estimées pour chaque population sur le jeu de données sans apparentés plus proches que cousins au second degré.

Les coefficients de consanguinité ont été comparés entre systèmes de parenté en utilisant un modèle linéaire mixte avec la fonction 'lmer' de la librairie 'lme4' v1.1-9 dans le logiciel R. Tout comme les modèles linéaires généralisés mixtes utilisés dans le chapitre 1 (partie 2.4.3), le modèle linéaire mixte inclut en effet fixe deux variables catégorielles : le système de parenté et la population. Le village d'échantillonnage et la famille ont été à nouveau inclus comme effet aléatoire dans le modèle. Les variables population, village et famille ont été ajoutées au modèle pour prendre en compte la structuration possible de nos données à ces trois différentes échelles.

Nous avons effectué une transformation logarithmique des coefficients de consanguinité pour obtenir une distribution normale des valeurs afin de respecter les hypothèses du modèle linéaire mixte. Nous avons aussi redistribué les valeurs nulles du coefficient de consanguinité de façon uniforme entre 0 et la valeur minimale non nulle avant la transformation logarithmique.

La significativité de la différence entre systèmes de parenté a été évaluée en comparant les moyennes par système de parenté après avoir contrôlé pour les facteurs confondants (population, village et famille) avec la fonction 'lsmeans' de la librairie 'lsmeans' v2.27-2 dans le logiciel

R. Cette méthode est indispensable pour prendre en compte la structure nichée de nos données : les populations sont nichées au sein des systèmes de parenté.

2.7 Détection des segments d'homozygotie

Les segments d'homozygotie (ou runs of homozygosity, ROHs) sont les segments du génome où les allèles présents sur les deux chromosomes de l'individu sont homozygotes, c'est-à-dire identiques. Leur longueur peut varier de quelques centaines de kilobases (kb) à plusieurs dizaines de megabases (Mb) (Broman et Weber, 1999; Gibson et al., 2006; McQuillan et al., 2008; Kirin et al., 2010).

Les ROHs courts, moyens et longs selon la définition de Pemberton (Pemberton et al., 2012a) seraient générés par différents mécanismes évolutifs (Kirin et al., 2010; Pemberton et al., 2012a). La longueur totale des ROHs courts ou de taille intermédiaire augmente à mesure qu'on s'éloigne d'Afrique. Ces ROHs seraient donc issus des événements du passé lointain lié à l'histoire du peuplement de la planète et seraient une mesure de la consanguinité de dérive liée à l'histoire démographique des populations. En revanche, la longueur totale des ROHs longs est plus importante dans des populations connues pour avoir un taux élevé de mariages entre apparentés proches, comme les populations du Moyen-Orient ou d'Asie du Sud (McQuillan et al., 2008; Kirin et al., 2010; Pemberton et al., 2012a). La longueur totale des ROHs longs seraient donc liées à de la consanguinité récente due à des mariages entre apparentés proches (Auton et al., 2009; Ben Halim et al., 2015).

Nous avons détecté les ROHs selon la méthode proposée par Pemberton et al. (Pemberton et al., 2012a) séparément dans chaque population. Les ROHs ainsi identifiés ont été répartis en 3 classes (court, moyen et long) par la fonction 'mclust' de la librairie 'mclust' v5.0.2 dans le logiciel R. La fonction 'mclust' décompose la distribution des longueurs de ROH en 3 classes de taille caractérisées chacune par une distribution gaussienne. Les termes « ROH court », « ROH moyen » et « ROH long » désigneront dans la suite de ce chapitre, les ROHs appartenant respectivement aux classes de taille courte, moyenne et longue. La longueur totale des ROHs dans chaque classe de taille a été calculée pour chaque individu et comparée entre systèmes de parenté par un modèle linéaire mixte. Le modèle est identique à celui pour les coefficients de consanguinité de FEstim. Comme pour les coefficients de consanguinité de FEstim, la significativité de la différence entre systèmes de parenté a été évaluée avec la fonction 'lsmeans' de la librairie 'lsmeans' v2.27-2 dans le logiciel R.

3. Résultats

3.1 Différenciation génétique entre populations

Nous observons sur le premier axe de la MDS des Fst, une séparation entre les Prai et les autres populations (Figure 3-2). Le Fst moyen entre les Prai et les autres populations est de 0,041 (écart-type = 0,0040). Les regroupements de populations sur la MDS reflètent la proximité géographique des populations. En effet, les Pacoh et les Ta-oih, les Khmu et les Ramet ainsi que les Tampuan, les Jarai et les Kacho résident dans les mêmes aires géographiques (voir carte en introduction). Cependant, les Brao et les Kreung se distinguent sur le second axe de la MDS alors que ce sont des populations résidant dans la même région que les Tampuan, les Jarai et les Kacho'. La distance génétique moyenne séparant les trois dernières populations avec les deux premières (Fst = 0,014 et 0,017 respectivement) est supérieure à la distance moyenne les séparant des Khmer (Fst = 0,008).

Figure 3-2 : Analyse MDS fondée sur la matrice de Fst entre populations. Les populations M sont représentées en rouge, les populations C en bleu et les populations P en orange.

3.2 Distance génétique entre individus

La Figure 3-3 présente les analyses MDS des *allele sharing distances* (ASD) entre les individus des 12 populations. Les individus des différentes populations se retrouvent à des distances relatives similaires à celle des populations sur la MDS des Fst présentée dans la partie précédente. Le premier axe sépare les Prai de toutes les autres populations. Les individus des autres populations sont distribués sur le second axe allant des Ta-oih et Pacoh jusqu'au Brao et Kreung tout comme sur la MDS fondée sur les Fst (Figure 3-2). La distance moyenne de lien entre individus de différentes populations (Table 3-2) montre que les individus d'une même population se séparent des individus des autres populations, à l'exception de trois paires de populations : Les Tampuan et les Kacho' (L_0 = 0,0068, p-value = 0,085), les Khmu' et les Ramet (L_0 = 0,0030, p-value = 0,96) et les Ta-oih et Pacoh (L_0 = 0,0032, p-value = 0,36). Ainsi, les regroupements géographiques que nous avons observés sur la MDS fondée sur les Fst sont confirmés par cette analyse. Les Jarai que nous avions regroupés précédemment avec les Tampuan et Kacho' se distinguent à l'échelle des individus.

	Jarai	Prai	Kacho'	Bunong	Khmer	Brao	Kreung	Khmu'	Ramet	Ta-oih	Pacoh
Tampuan	**0.008**	**0.083**	0.007	**0.010**	**0.012**	**0.028**	**0.038**	**0.021**	**0.020**	**0.021**	**0.022**
Jarai		**0.083**	**0.005**	**0.004**	**0.006**	**0.035**	**0.045**	**0.016**	**0.015**	**0.014**	**0.015**
Prai			**0.083**	**0.082**	**0.079**	**0.088**	**0.093**	**0.071**	**0.071**	**0.086**	**0.085**
Kacho'				**0.007**	**0.009**	**0.031**	**0.041**	**0.018**	**0.017**	**0.018**	**0.019**
Bunong					**0.004**	**0.037**	**0.047**	**0.014**	**0.013**	**0.012**	**0.013**
Khmer						**0.039**	**0.049**	**0.011**	**0.010**	**0.011**	**0.012**
Brao							**0.011**	**0.045**	**0.044**	**0.049**	**0.049**
Kreung								**0.055**	**0.054**	**0.059**	**0.060**
Khmu'									0.003	**0.015**	**0.016**
Ramet										**0.015**	**0.015**
Ta-oih											0.003

Table 3-2 : Distance moyenne de lien entre populations. Les distances en gras sont significatives au seuil de 0,05 après correction de Bonferonni pour tests mulitples.

Sur le troisième axe, les deux paires de populations patrilinéaires qui étaient proches sur les deux premiers axes se séparent avec d'un côté les Khmu' et Ramet et de l'autre les Ta-oih et Pacoh, reflétant leur éloignement géographique. Par ailleurs, les Bunong et les Khmer se distinguent aussi plus nettement que sur les deux premiers axes.

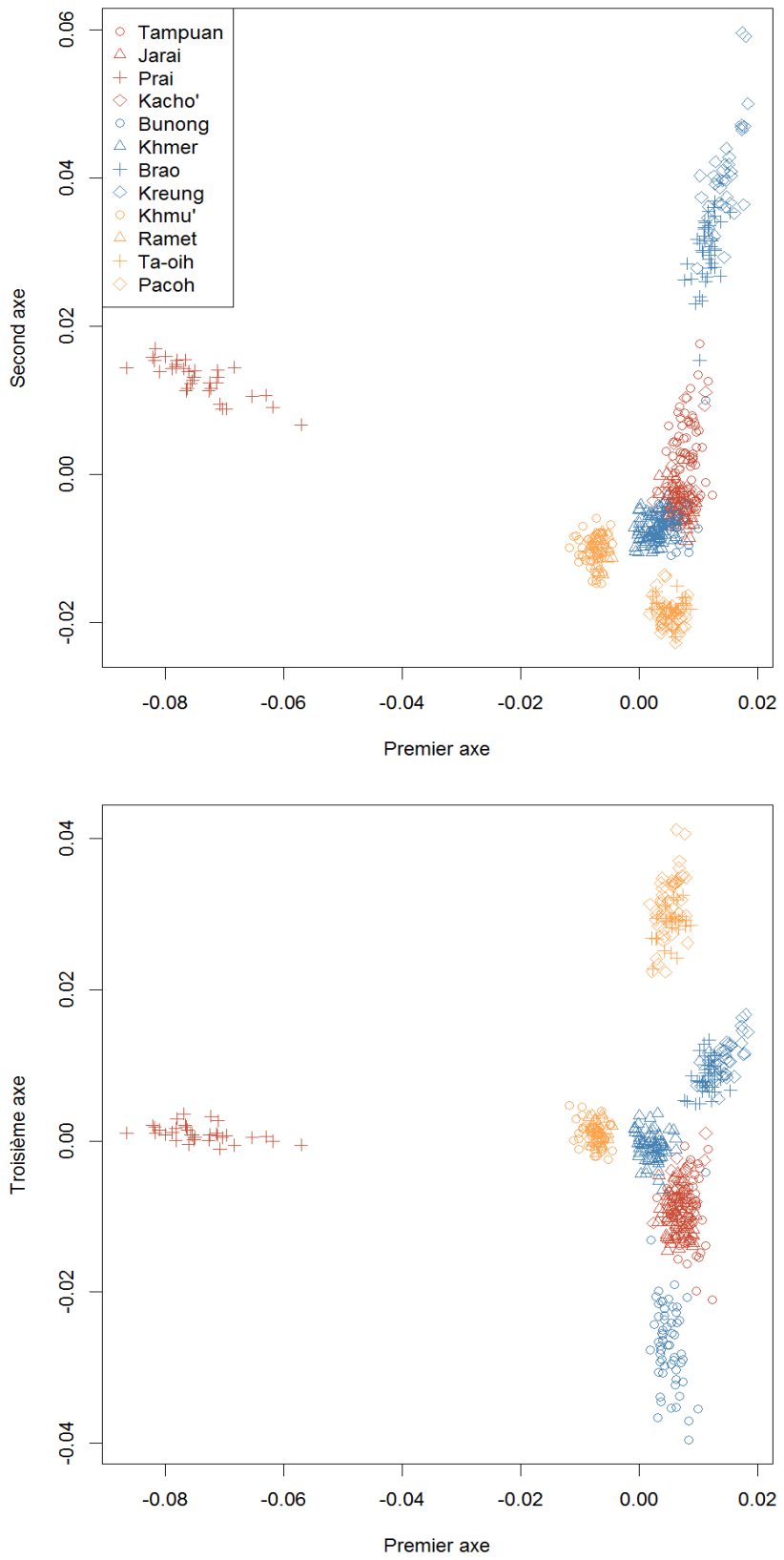


Figure 3-3 : MDS des allele sharing distances (ASD) entre les individus des 12 populations. (A) Projection sur les deux premiers axes de la MDS. (B) Projection sur le premier et troisième axe de la MDS.

3.3 Taille efficace des populations

Nous avons voulu vérifier que les populations étudiées ont une taille efficace similaire et donc qu'une force de dérive similaire s'exerce sur elles afin de pouvoir distinguer l'effet des systèmes de parenté de celui de la dérive sur la consanguinité. La Figure 3-4 présente les estimations de taille efficace des populations. Les Prai sont la seule population avec une taille efficace (Ne = 7182, intervalle de confiance à 95% [6382 ; 7982]) significativement inférieure à toutes les autres populations (p-values < 0.05). La dérive est donc plus importante chez les Prai que les autres populations. Par ailleurs, les Khmers qui sont la population majoritaire au Cambdoge, ont la plus grande taille de population (Ne = 15 503, intervalle de confiance à 95% [14 376 ; 16 630].

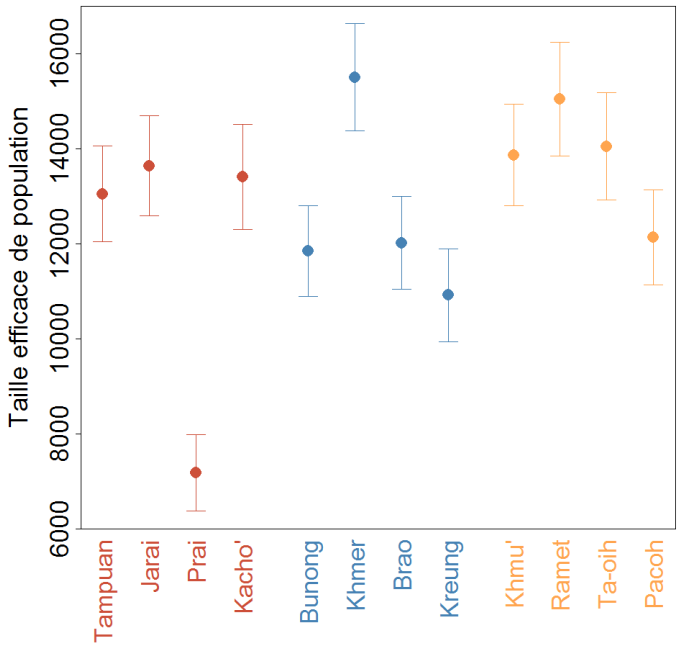


Figure 3-4 : Taille efficace des populations. Les tailles efficaces des populations ont été estimées selon la méthode de McVean pour chaque chromosome (Auton et McVean, 2007). *La moyenne sur les 22 chromosomes est représentée avec l'intervalle de confiance à 95%. Les populations sont regroupées par système de parenté, les populations M sont en rouge, les populations C en bleu et les populations P en orange.*

3.4 Hétérozygotie haplotypique

Tout comme pour la taille efficace des populations, les Prai sont la seule population avec une hétérozygotie (H = 0,580, intervalle de confiance à 95% [0,575 ; 0,586]) significativement inférieure à toutes les autres populations (p-values < 0,05, Figure 3-5).

Ces valeurs singulières des Prai en termes de taille efficace et d'hétérozygotie haplotypique ont probablement pour origine l'histoire démographique patriculière de cette population (voir chapitre 1 partie 2.1). Cette différence d'histoire démographique rend sa comparaison avec les autres populations difficile et nous avons donc retiré les individus Prai des analyses de consanguinité qui suivent.

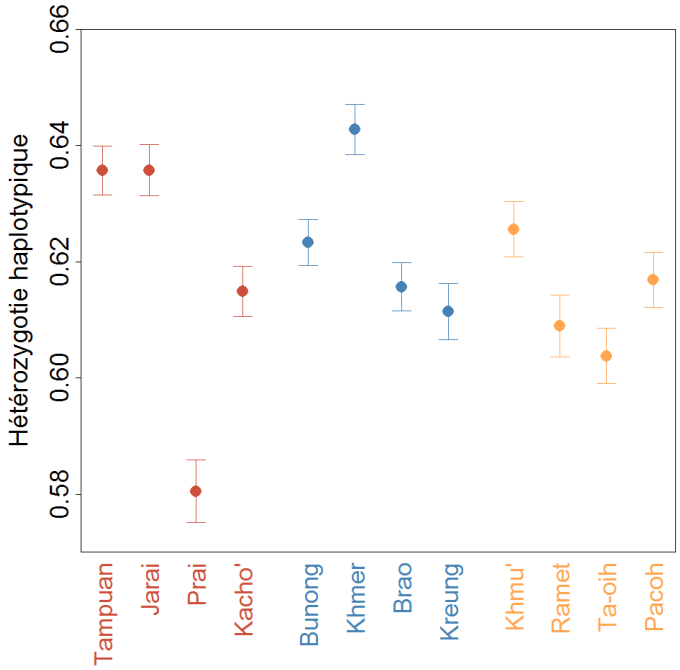


Figure 3-5 : Hétérozygotie haplotypique des populations. L'hétérozygotie haplotypique a été estimée selon la méthode proposée par Li et al. (2008) pour chaque chromosome. La moyenne sur les 22 chromosomes est représentée avec l'intervalle de confiance à 95%. Les populations sont regroupées par système de parenté, les populations M sont en rouge, les populations C en bleu et les populations P en orange.

3.5 Coefficient de consanguinité

La Figure 3-6 présente les coefficients de consanguinité individuels estimés par FEstim par ordre croissant dans chaque population. Dans toutes les populations, une proportion des individus ont un coefficient de consanguinité non nul. Cette proportion est la plus élevée chez les Pacoh avec 67% des individus ayant un coefficient de consanguinité nul.

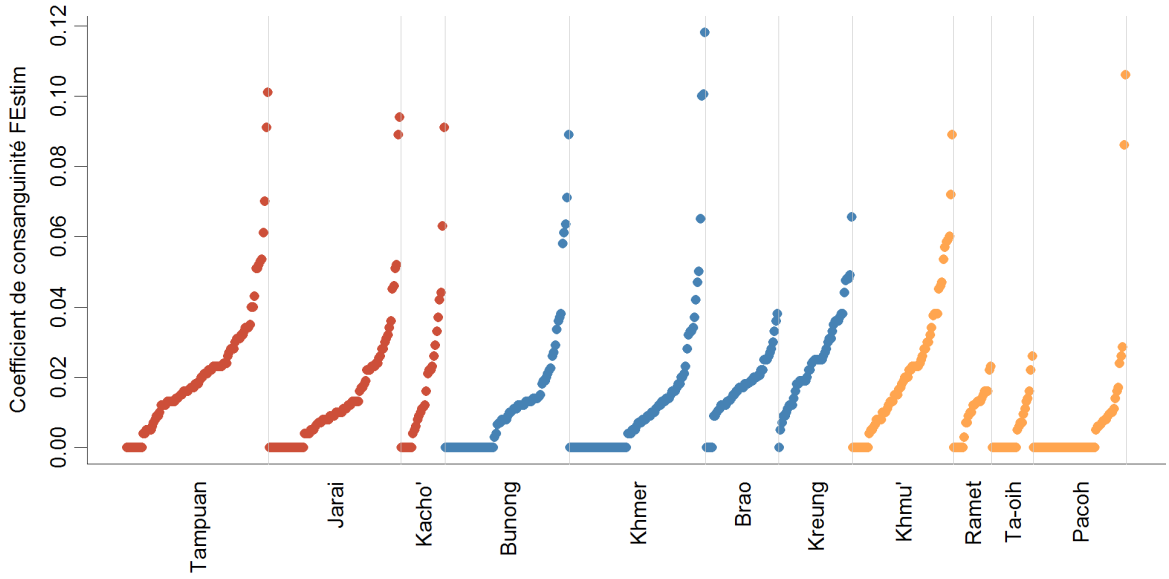


Figure 3-6 : Coefficient de consanguinité estimé par FEstim. Chaque point représente un individu dans la population, les individus sont classés par ordre de coefficient de consanguinité croissant. Les populations M sont en rouge, les populations C en bleu et les populations P en orange.

Le modèle linéaire mixte indique que les populations M et C ont des coefficients de consanguinité significativement plus élevés (0,018 et 0,017 en moyenne respectivement) que les populations P (0,011 ; p-values < 0,05, Figure 3-7). Les populations M et C ont des coefficients de consanguinité similaires (p-value = 0,99). Une analyse de la variance (ANOVA) nous indique que 27% de la variance du coefficient de consanguinité entre populations peut être expliquée par le système de parenté.

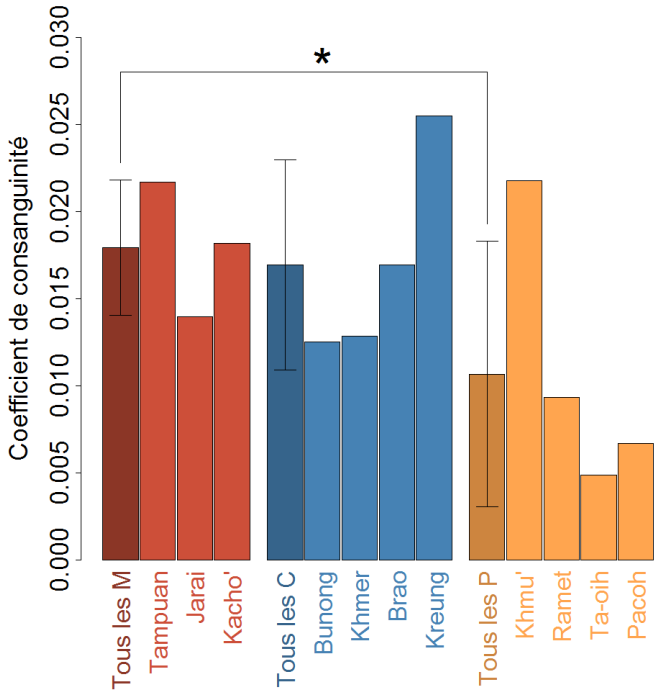


Figure 3-7 : Coefficient de consanguinité moyen par population. Le coefficient de consanguinité a été estimé par FEstim. Les populations sont regroupées par systèmes de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire mixte.

Les différences significatives du coefficient de consanguinité entre les populations M et C et les populations P sont cohérentes avec les résultats obtenus dans le chapitre 1 sur le taux d'endogamie de village. En effet, les populations M et C ont un taux d'endogamie supérieur aux populations P, ce qui pourrait potentiellement être lié à l'existence du « puzzle matrilinéaire » chez les populations M et C. Pour confirmer que l'endogamie de village est une variable ethno-démographique permettant d'expliquer les différences de consanguinité entre population, nous avons estimé le coefficient de corrélation entre ces deux variables à l'échelle des populations (Figure 3-8). Le taux d'endogamie de village est significativement corrélé au coefficient de consanguinité (ρ = 0,73, p-value = 0,015).

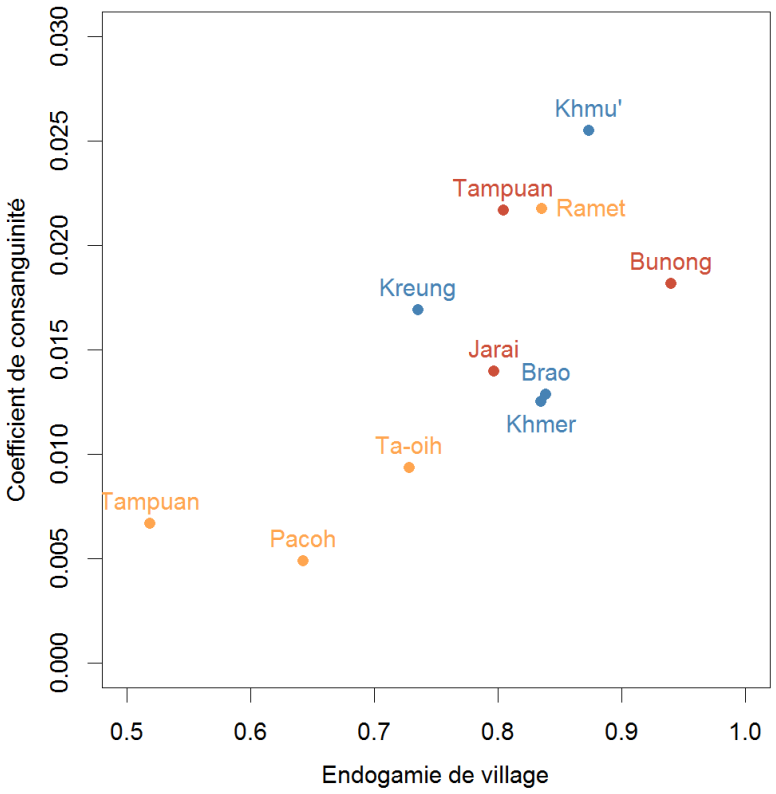


Figure 3-8 : Coefficient de consanguinité en fonction de l'endogamie de village à l'échelle des populations. Les deux variables sont positivement corrélées (ρ = 0,73, p-value = 0,015). Les populations M sont représentées en rouge, les populations M en bleu et les populations P en orange.

Par ailleurs, nous avons également testé si les populations unilinéaires ont une consanguinité plus élevée que les populations cognatiques. Pour cela nous avons regroupé les populations M et P. Il n'y a pas de différence significative de niveau de consanguinité moyen entre les populations unilinéaires (0,14) et cognatiques (0,17, p-value = 0,15).

3.6 Longueur des segments d'homozygotie

La Figure 3-9 présente pour les 11 populations, la longueur totale des ROHs longs (les figures pour les ROHs courts et moyens sont présentées en annexe 6). Les ROHs courts mesurent en moyenne 322,1 kb (intervalle de confiance à 95% [321,7 ; 322,4]), les ROHs moyens 840 kb (intervalle de confiance à 95% [839,2 ; 841,7]) et les ROHs longs 4 093 kb (intervalle de confiance à 95% [4012 ; 4174]).

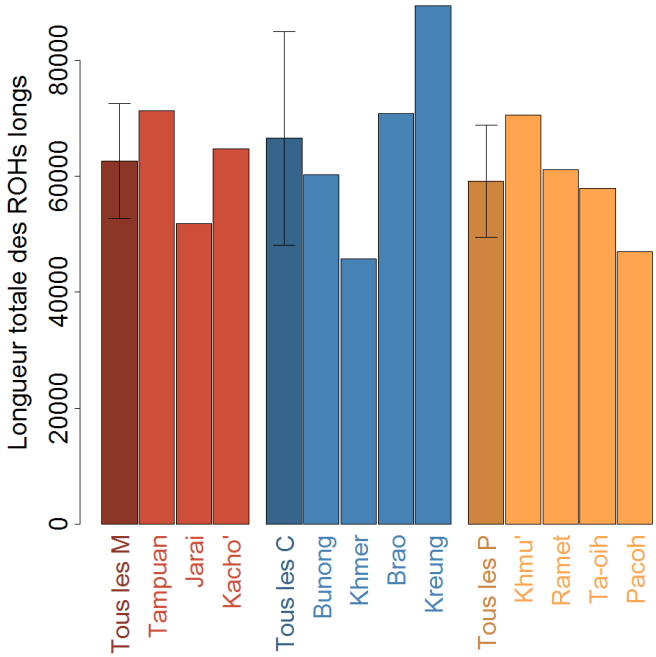


Figure 3-9 : Longueur totale individuelle des ROHs longs par population en kilobase (kb). Les populations sont regroupées par systèmes de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté.

La longueur totale des ROHs n'est pas significativement différente entre les systèmes de parenté dans les trois classes de taille (p-values > 0,05 pour toutes les comparaisons). Nous nous attendions à ne pas observer de différence significative entre systèmes de parenté pour la longueur totale des ROHs courts ou moyens qui sont liées à la consanguinité de dérive. Cependant, il est surprenant de constater que la longueur totale des ROHs longs, censée être un estimateur de la consanguinité due à des unions entre apparentés proches, n'est pas significativement différente entre les systèmes de parenté alors que le coefficient de consanguinité de FEstim l'est. Cependant, lorsque nous intéressons aux ROH « extra longs », définis ici comme ayant une longueur supérieure à 5Mb, la longueur totale des ROHs chez les individus des populations M devient significativement supérieure à celle des individus des

populations P (p-value = 0,009). Les populations C n'ont pas une longueur totale des ROHs « extra longs » significativement différente des populations M (p-value = 0,41) ou P (p-value = 0,14 ; Figure 3-10).

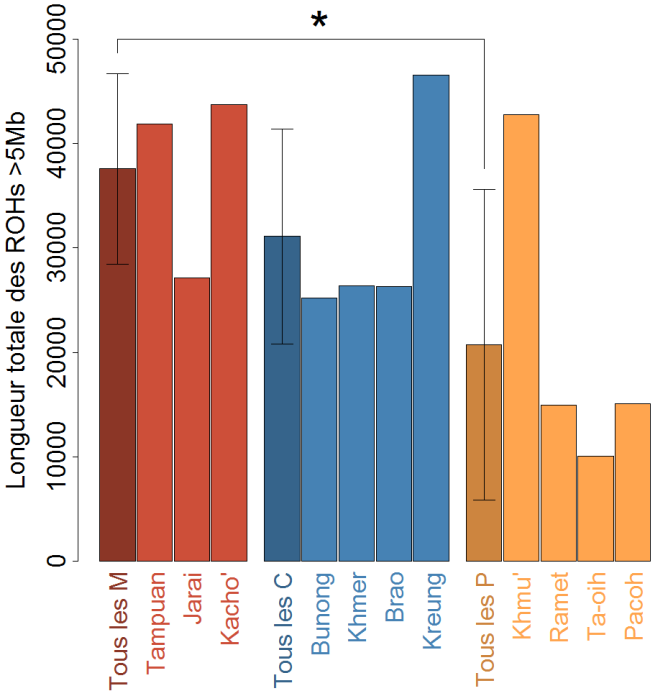


Figure 3-10 : Longueur totale individuelle des ROHs d'une longueur supérieure à 5Mb par population en kilobas (kb). Les populations sont regroupées par systèmes de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté. L'astérisque représente les comparaisons significativement différentes au seuil de 0,05 évaluées par modèle linéaire mixte.

Les ROHs d'une longueur supérieure à 5Mb ont été montrés comme étant fortement corrélés à la consanguinité estimée à partir de la généalogie (Kirin et al., 2010). De plus, la longueur totale des ROHs avec une longueur supérieure à 5Mb est significativement corrélée au taux d'endogamie de village ($\rho = 0,78$, p-value = 0,004, Figure 3-11 A) alors que la longueur totale des ROHs longs ne l'est pas ($\rho = 0,45$, p-value = 0,17, Figure 3-11 B).

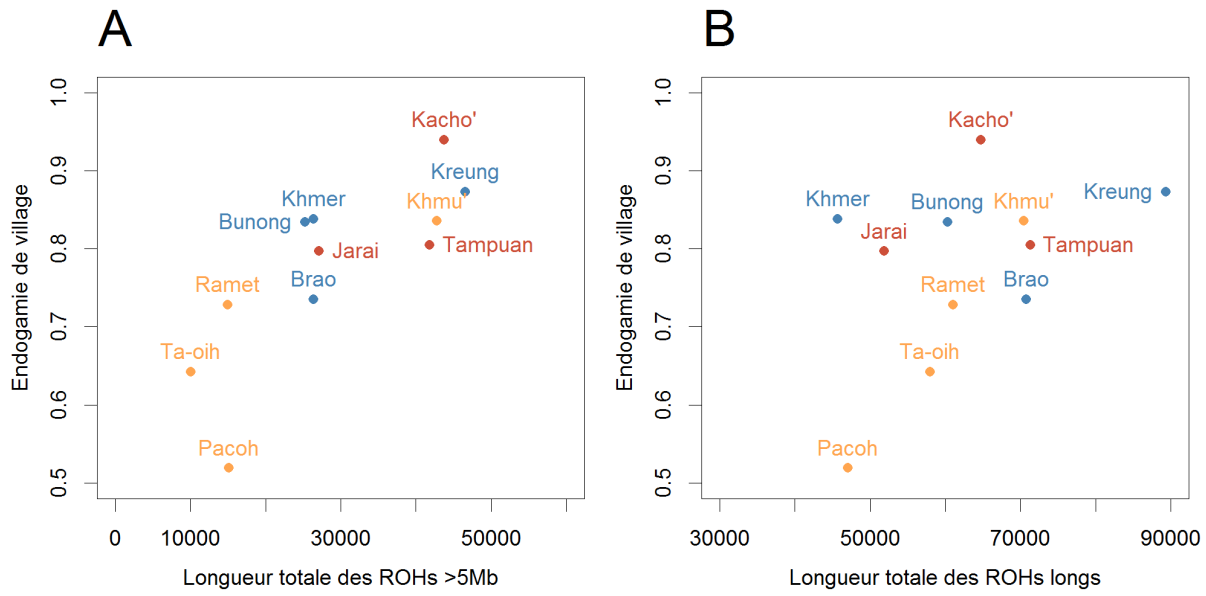


Figure 3-11 : Longueur totale des ROHs en fonction de l'endogamie de village à l'échelle des populations. (A) Longueur totale des ROHs >5Mb (ρ = 0,78, p-value = 0,004). (B) longueur totale des ROHs longs (ρ = 0,45, p-value = 0,17). Les populations M sont représentées en rouge, les populations M en bleu et les populations P en orange.

4. Discussion

Dans ce chapitre, nous nous sommes intéressés à l'impact du système de parenté sur la diversité génétique autosomale et plus particulièrement la consanguinité. Notre but était notamment de tester deux hypothèses. Premièrement, nous nous sommes demandé si nous pouvons détecter l'effet de l'endogamie de village potentiellement liée au « puzzle matrilinéaire ». Nous avons montré dans le chapitre 1 que les populations M et C ont un taux d'endogamie de village plus important que les populations P. Nous nous attendions donc à ce que le coefficient de consanguinité reflète cette différence d'endogamie de village, et que les populations M et C aient des coefficients de consanguinité supérieurs aux populations P.

Deuxièmement, nous avons testé si l'effet de la préférence pour les mariages entre cousins de premier degré chez les populations unilinéaires (M et P) était détectable par des mesures génétique de consanguinité. Nous avons supposé d'après la littérature ethnographique qu'il y avait une préférence plus importante pour les mariages entre cousins de premier degré chez les populations unilinéaires que chez les populations cognatiques. Nous nous attendions alors à ce que les populations unilinéaires aient des coefficients de consanguinité supérieurs aux populations cognatiques.

Avant de tester ces deux hypothèses, nous avons au préalable réalisé un certain nombre d'analyses descriptives afin de s'assurer que les histoires démographiques des populations M, C et P à l'étude étaient comparables, puisque la démographie peut également impacter les niveaux de consanguinité estimés à partir des données génétiques. C'est donc un facteur potentiellement confondant à prendre en compte ici (Jacquard, 1968; Pemberton et al., 2012a). Nos analyses MDS effectuées sur les Fst à l'échelle des populations et sur les *allele sharing distance* (ASD) à l'échelle des individus montrent une séparation entre les populations sur ces deux estimateurs de distance génétique. Les Prai se distinguent des autres populations sur le premier axe sur les deux analyses. Les autres populations sont distribuées sur le second axe avec aux extrêmes les Kreung et les Brao d'un côté et les Pacoh et les Ta-oih de l'autre. Le calcul des distances de lien sur la MDS fondée sur les ASD nous permet de montrer qu'il est possible de distinguer les populations étudiées sur la base de différences génétiques. Nous observons des distances moyennes de lien significatives entre les paires de populations. Il existe 3 exceptions à cette observation générale. Les distances moyennes de lien entre paires d'individus ne sont pas significativement différentes entre : Tampuan et Kacho', Khmu' et Ramet, Ta-oih et Pacoh. Nous remarquons que ces paires de population sont non seulement

proches génétiquement, mais aussi géographiquement (voir carte en introduction). La proximité géographique explique en partie la proximité génétique, les populations pouvant se mélanger et échanger du matériel génétique plus facilement.

Nous pouvons par ailleurs noter que ces paires de populations partagent aussi pour certaines les noms de clan : les Khmu' et les Ramet (Evrard, 2006), les Tampuan et les Kacho' (données enquêtes de terrain). Ce partage des clans pourrait aussi faciliter le mélange de ces populations expliquant leur similarité génétique et/ou être le témoin d'un passé commun. A l'inverse, les Jarai qui sont géographiquement proches des Tampuan et Kacho' et qui partagent aussi les mêmes noms de clans peuvent être génétiquement séparés des deux autres populations. Les Jarai sont une population particulièrement intéressante. Elle possède le même système de parenté que ses voisines les Tampuan et les Kacho' ainsi que les mêmes clans, mais parle une langue de la famille austronésienne alors que les deux autres populations parlent une langue de la famille austro-asiatique. Deux hypothèses s'opposent pour expliquer la différence des Jarai avec les deux autres populations. La première est que les Jarai étaient anciennement une population très proche des Tampuan et Kacho', parlant une langue austro-asiatique et qu'il y aurait eu un remplacement linguistique lorsque les Cham (une population austronésienne) sont arrivés dans la région. La seconde est que les Jarai étaient anciennement une population proche des Cham donc austronésienne qui est arrivée dans la région et qui a incorporé le système de parenté des populations avoisinantes, à savoir les Tampuan et Kacho'. Les statistiques descriptives que nous avons estimées ici ne permettent évidemment pas de départager entre ces deux hypothèses, mais il est intéressant de noter que les Tampuan et les Kacho' se confondent génétiquement alors que les Jarai peuvent être distingués.

L'étude de la taille efficace et de l'hétérozygotie haplotypique ont mis en évidence l'existence d'une histoire démographique particulière chez les Prai. Les Prai ont des estimations significativement inférieures à toutes les autres populations pour ces deux variables. De plus, sur les analyses MDS, les Prai se distinguent sur le premier axe et montrent l'existence d'une distance génétique plus importante avec les autres populations. De ce fait, cette population a été retirée du jeu de données pour les analyses concernant l'influence du système de parenté sur la consanguinité. Cette observation n'est pas surprenante, nous avons déjà mentionné dans le premier chapitre que les Prai étaient une population démographiquement particulière. Elles sont tout d'abord fortement isolées des populations voisines et vivent principalement sur les crêtes de montagne difficiles d'accès. Cet isolement géographique peut en partie expliquer l'existence d'une histoire démographique particulière chez cette population. Deuxièmement les villages

Prai échantillonnés sont un très petit échantillon de la population Prai, beaucoup plus large, qui s'étend très loin en Thaïlande. Cette population est donc connue pour être historiquement et géographiquement plus éloignée des autres.

Après ces analyses descriptives de la diversité génétique autosomale des populations à l'étude et l'exclusion des Prai du fait de leur histoire démographique particulière, nous nous sommes intéressés aux questions principales de ce chapitre.

La comparaison des coefficients de consanguinité montre que les populations M et C ont des coefficients de consanguinité plus élevés que les populations P. Il semble donc que l'endogamie de village plus élevée dans les populations M et C (chapitre 1), probablement en lien avec le « puzzle matrilinéaire » et donc la difficulté des hommes à maintenir leur autorité sur leur famille natale et conjugale s'ils se marient dans un village différent, accroissent dans ces populations les niveaux de consanguinité estimés à partir des données génétiques. Ces résultats sont à notre connaissance les premiers montrant l'effet du « puzzle matrilinéaire » sur la consanguinité génétique. L'association entre l'endogamie et le système de parenté matrilinéaire et matrilocal a été montrée par Murdock en 1949 (Murdock, 1949). Cependant, aucune étude génétique ne s'était jusqu'à présent intéressée à l'impact de cette endogamie sur les patrons de diversité génétique.

Plus généralement, nous observons une corrélation significative et positive entre le taux d'endogamie de village et le coefficient de consanguinité à l'échelle des populations, ce qui soutient l'idée que l'endogamie de village est bien un processus façonnant les niveaux de consanguinité dans ces populations.

Nous remarquons que les populations C ont un taux d'endogamie de village et un coefficient de consanguinité similaires aux populations M. Les hommes dans les populations C auraient donc également une préférence élevée pour l'endogamie de village malgré le fait qu'il n'existe pas de groupes de filiation chez ces populations. Les hommes dans les populations matrilocales pourraient préférer se marier au sein de leur village afin de garder leur position d'autorité dans leur village natal. Les Brao et les Kreung sont des populations cognatiques et multilocales à prédominante matrilocale. Les conjoints vont habiter alternativement avec les parents du mari et de la femme pendant les premières années du mariage. Ainsi, le couple vit matrilocalement la moitié du temps, ce qui semble être suffisant pour entrainer une préférence pour l'endogamie de village aussi élevé que chez les populations matrilinéaires. De plus, si les conjoints viennent de deux villages différents, les déplacements entre les foyers des deux familles seraient moins

pratiques que s'il suffisait de se déplacer au sein du même village. Cela pourrait aussi expliquer le taux d'endogamie de village élevé chez ces populations. Ainsi, il semblerait que l'hypothèse du « puzzle matrilinéaire » pourrait tout aussi bien être élargie à toutes les populations matrilocales ou multilocales, ce qui nous appelle à interroger les anthropologues sur cette question et renommer cette hypothèse élargie le « puzzle matrilocale ».

Les Khmu sont une population particulière en termes d'endogamie de village et de coefficient consanguinité comparée aux autres populations P. Alors que les autres populations P ont des faibles taux d'endogamie et coefficients de consanguinité, les Khmu ont au contraire un fort taux d'endogamie qui va de pair avec un fort coefficient de consanguinité. Ainsi, pour ces deux estimateurs, les Khmu sont plus similaires aux populations M qu'aux populations P. Ces résultats semblent être cohérent avec l'observation ethnologique. Une préférence pour l'endogamie de village a été notée sur le terrain par Olivier Evrard (communication personnelle), cependant nous n'avons pas d'explication sur la raison de ce taux d'endogamie de village élevé chez les Khmu'.

La longueur totale des ROHs longs (définis par la méthode de Pemberton et al. (2012a)) a été décrit comme un estimateur de la consanguinité récente résultante d'unions entre apparentés proches (Pemberton et al., 2012a). Or, nous n'observons pas de différences significatives entre les systèmes de parenté, ce qui ne concorde pas avec les résultats obtenus pour les coefficients de consanguinité estimés par FEstim. Cependant, les populations M ont une longueur totale de ROH « extra-longs » (définis ici comme mesurant 5Mb et plus) significativement plus élevée que les populations P (p-value = 0,009) La proportion du génome compris dans un ROH d'une longueur supérieur à 5 Mb (un estimateur corrélé à la longueur totale) est fortement corrélée à la consanguinité estimée à partir de généalogie dans les populations européennes (r= 0,86, McQuillan et al., 2008). Ainsi, il nous semble que le seuil de 5Mb permet de mieux évaluer la consanguinité générée par des unions entre apparentés proches que le seuil des longs ROHs utilisé par Pemberton et al. (2012a).

La littérature ethnologique sur les populations à l'étude ici (Izikowitz, 1951; LeBar et al., 1964; Dournes, 1972; Bourdier, 2006; Evrard, 2006), suggère que les populations unilinéaires possèdent une préférence plus marquée pour les mariages entre cousins de premier degré que les populations cognatiques. Nous avons testé cette hypothèse en comparant les coefficients de consanguinité entre ces deux groupes de population. Nous n'observons pas de différence de consanguinité entre les populations unilinéaires et cognatiques. Nous n'avons donc pas détecté

d'effets de la préférence des mariages entre apparentés proches sur la consanguinité dans les populations étudiées. Plusieurs raisons peuvent expliquer cette absence de différence.

Tout d'abord, nous n'avons pas pu analyser de façon quantitative les types de mariages dans les différentes populations. Nos suppositions découlent donc des règles décrites dans la littérature ethnographique. Ainsi, il se peut que les populations unilinéaires et cognatiques étudiées dans cette thèse ne se distinguent en réalité pas en termes de taux de mariages entre cousins de premier degré. Il est alors normal de ne pas observer de différence en termes de consanguinité entre les populations unilinéaires et cognatiques.

D'autre part, même s'il existe une préférence pour le mariage entre cousins de premier degré tous les couples dans la population ne peuvent pas suivre cette préférence en cas de restriction du marché matrimonial satisfaisant ces conditions. Ainsi, pour la préférence de mariage avec la cousine croisée matrilatérale (la fille du frère de la mère), une approche par simulation montre que ces mariages représentent uniquement 30% du marché matrimoniale (Guillot et al., 2015). Par ailleurs, la même étude montre qu'un taux faible de mariages avec la cousine croisée matrilatérale ne laisse pas de signature génétique sur la diversité autosomale. Ainsi, nous pourrions ne pas observer de différences au niveau génétique même s'il existe des préférences différentes entre les systèmes de parenté car tous les couples ne peuvent pas se marier selon cette préférence.

Enfin, une troisième raison est que la préférence pour le mariage entre cousins de premier degré chez les unilinéaires s'étendrait également aux cousins classificatoires (van den Berghe, 1980). Nous parlons de lien de parenté classificatoire avec Ego lorsque le lien biologique de parenté n'existe pas, mais la personne est considérée comme étant, vis-à-vis d'Ego dans un rapport identique ou équivalent (Godelier, 2004). Ainsi, chez les Khmu' par exemple, « Le système de parenté prescrit le mariage avec la cousine croisée matrilatérale (fille de frère de mère), réelle ou classificatoire […]. Dans ce système, le mariage avec une réelle fille de frère de mère représente une solution idéale mais le mariage peut avoir lieu avec n'importe quelle femme de catégorie et de génération identique. » (Evrard, 2006). Le mariage avec une femme appartenant au clan de la mère et de même génération représenterait alors un mariage avec une cousine croisée matrilatérale. Ainsi les individus en se mariant à des cousines classificatoires, respectent la préférence de mariage, mais au niveau génétique, le couple n'est pas apparenté proche. Nous ne nous attendons pas à un excès de consanguinité dans cette population dans ce cas. En outre, dans ce cas, capturer la présence de mariages entre cousins classificatoires s'avère très difficile par une enquête par questionnaire en absence de reconstruction généalogique à l'échelle – au

moins- du village. Ceci expliquerait également l'absence de résultats de notre étude ethno-démographique (chapitre 1).

Les résultats de ce chapitre montrent que le "puzzle matrilinéaire" pourrait potentiellement modeler la diversité autosomale dans certaines populations humaines. Nous avons donc mis à jour l'existence d'un nouvel aspect du système de parenté qui pourrait avoir une influence sur la diversité autosomale. Les études précédentes ont montré que la consanguinité dans les populations pouvaient être lié à un taux de mariages entre apparentés proches plus élevés (McQuillan et al., 2008; Kirin et al., 2010; Leutenegger et al., 2011; Pemberton et al., 2012a; Yang et al., 2014; Gazal et al., 2015). Nos résultats suggèrent que l'endogamie de village potentiellement liée au « puzzle matrilinéaire » pourrait également avoir pour effet d'augmenter la consanguinité dans certaines populations, ajoutant ainsi un niveau de complexité supplémentaire aux interactions entre systèmes de parenté et diversité génétique humaine.

Conclusion

L'objectif de cette thèse est d'étudier l'influence des systèmes de parenté sur la diversité génétique humaine. En effet, jusqu'à présent, les études dans ce champ de recherche se sont principalement intéressées à la résidence, beaucoup moins à la filiation, et très peu à l'alliance. En outre, ces études n'ont généralement pas intégré le fait que ces règles interagissent et se croisent. Elles constituent ensemble, et non de façon isolée, un système de parenté. C'est de leur combinaison que nait la vaste gamme de systèmes de parenté pratiquée par les populations humaines. Par ailleurs, ces études ont jusqu'à présent catégorisé les populations selon la règle énoncée par les informateurs ou observée par l'ethnologue (c'est-à-dire selon une norme sociale) et non selon les pratiques réelles des individus (pouvant plus ou moins s'écarter de la norme selon les populations). Par exemple, dans cette catégorisation, la patrilocalité et la matrilocalité ont été considérées comme des opposés symétriques de l'un par rapport à l'autre : les femmes migrent dans les populations patrilocales et les hommes dans les populations matrilocales. Or, des asymétries peuvent potentiellement exister entre les systèmes de parenté.

Une raison pouvant générer des asymétries entre les systèmes de parenté matrilinéaires et patrilinéaires que nous avons étudiée tout au long de cette thèse est le « puzzle matrilinéaire» (Richards, 1950; Schneider et Gough, 1961). Le pouvoir politique et l'autorité au sein des groupes de filiation mais aussi au sein du village appartiennent dans quasiment toutes les sociétés humaines aux hommes, quel que soit le système de parenté. Or dans les populations matrilinéaires, la filiation est transmise en lignée maternelle. Ainsi, pour reprendre les mots de Nicole-Claude Mathieu que nous avions déjà cité en introduction : « les hommes sont pris entre leur rôle de frère et de mari, entre leurs attaches et devoirs envers leur propre lignage, leurs sœurs et leurs neveux/nièces utérins, d'une part, et la « nécessité » (sic) d'asseoir leur autorité sur l'épouse et les enfants dans l'unité conjugale, d'autre part » (Mathieu, 2007). De ce fait, les hommes dans les populations matrilinéaires et matrilocales pourraient préférer s'écarter de la règle matrilocale et s'installer patrilocalement avec leur épouse afin de rester dans leur village natal et maintenir leur position d'autorité sur leur groupe de filiation. Il n'existe pas de raison similaire poussant les femmes à se marier matrilocalement dans les populations patrilocales. Ainsi, la gestion sociale du « puzzle matrilinéaire» pourrait entrainer une flexibilité de la résidence plus importante chez les populations matrilinéaires que les populations patrilinéaires. Une autre manière pour un homme dans une population matrilinéaire de maintenir

conjointement sa position d'autorité sur sa famille natale et sa famille conjugale serait de se marier localement, c'est-à-dire avec une femme vivant dans son village natal. Il en résulterait une endogamie de village plus importante chez les populations matrilinéaires que les populations patrilinéaires. Murdoch avait en effet noté que les systèmes matrilocaux sont effectivement plus endogames que les autres systèmes de parenté : « lorsque la résidence est matrilocale, un homme en se mariant, s'installe rarement dans une nouvelle communauté. Il prend ses possessions du foyer de ses parents et se déplace, pour ainsi dire, de l'autre côté de la rue, dans la partie du village où sa femme et la famille de sa femme habitent. » (traduction personnelle, Murdock, 1949).

Nous avons donc, dans cette thèse, pris en compte différents aspects du système de parenté en compte (résidence, filiation et alliance) et entrepris une approche pluridisciplinaire quantitative qui nous a permis de caractériser ethno-démographiquement divers systèmes de parenté afin de mieux comprendre l'impact des systèmes de parenté sur les patrons de diversité génétique. Nous nous sommes appuyés sur l'étude comparative de 12 populations d'Asie du Sud-Est possédant des règles de filiation différentes (patrilinéaire, matrilinéaire et cognatique) combiné à des règles de résidence variées (patrilocale, matrilocale et multilocale à dominance matrilocale). Cette région du monde est particulièrement adaptée à l'étude de l'influence des systèmes de parenté sur la diversité génétique car les populations ont des modes de vie similaires dans un environnement relativement homogène, ce qui permet de mieux isoler l'influence des systèmes de parenté. Par ailleurs, elle abrite des populations à filiation matrilinéaire et résidence matrilocale, un système de parenté peu étudié en génétique des populations.

Grâce à cette approche, nous avons tout d'abord dans le chapitre 1 montré que les systèmes de parenté se distinguent bien les uns des autres pour certaines variables ethno-démographiques d'importance pour l'évolution génétique des populations (c'est-à-dire déterminant notamment où et avec qui les individus se reproduisent et élèvent leurs enfants). Nous avons en outre mis en évidence qu'il existait une asymétrie au niveau de la flexibilité de la résidence entre les systèmes de parenté : les populations patrilinéaires et patrilocales (P) ont une proportion de couples matrilocaux plus importante que la proportion de couples patrilocaux chez les populations matrilinéaires et matrilocales (M). Les populations P présentent également une proportion de couples néolocaux plus importante que les populations M. Cette flexibilité de la règle de résidence plus importante chez les populations P que M se traduit par des taux de migration de femmes plus élevés chez les populations P que M - comme attendu du fait des règles de résidence respectives de ces populations - mais des taux de migration d'hommes

similaires entre les systèmes de parenté. La flexibilité de la règle de résidence plus importante chez les populations P que M est surprenante contraire à l'attendu ethnologique de l'hypothèse du « puzzle matrilinéaire ». Nous nous attendions en effet à une flexibilité de la résidence plus importante chez les populations M qui permettrait à une partie des hommes de se marier de façon patrilocale afin de rester à proximité de leur groupe de filiation et d'exercer leur fonction d'autorité sur leur groupe de filiation.

En revanche, notre analyse ethno-démographique montre que, si les hommes dans les populations M ne se marient pas plus fréquemment de façon patrilocale, ils pratiquent néanmoins plus fréquemment l'endogamie de village. En conséquence les taux d'endogamie de village sont plus élevés chez les populations M que P. Il s'agirait d'une manière de faire face au « puzzle matrilinéaire ». En effet, se marier à une femme née dans le même village permet à un homme de rester à proximité de son groupe de filiation maternel tout en rejoignant le foyer de sa femme.

Par ailleurs, en termes de dynamique des groupes de filiation, nous avons montré que les populations patrilinéaires tendent à moins bien connaître les clans de leurs ascendants paternels en comparaison à la connaissance des clans des ascendants maternels dans les sociétés matrilinéaires. Ces résultats semblent indiquer que les populations P sont plus flexibles que les populations M au niveau des groupes de filiation en plus de l'être au niveau de la résidence. Les clans jouent un rôle moins important en tant qu'unité structurante de la population chez les populations P que chez les populations M. Ces résultats sont corroborés par l'étude de Bérénice Alard qui a montré, à partir de ces mêmes données, que les femmes appartenant au même clan matrilinéaire sont plus apparentées entre elles en termes de lignées mitochondriales que les hommes appartenant au même clan patrilinéaire ne le sont en terme de chromosomes Y (Alard et al., in prep). La littérature ethnographique indique également que cette flexibilité des groupes de filiation semble correspondre à la façon dont est pensée la parenté dans ces populations : chez les Khmu' qui sont une population patrilinéaire, la filiation dépend tout d'abord du lieu de résidence et des individus partageant le même habitat. Ainsi, lorsqu'un homme habite avec la famille de sa femme, il intégrera le clan et le lignage de sa femme (Evrard, 2006).

Par ailleurs, nous avons proposé qu'il existe une dynamique de fission de clans particulière chez les Prai et les Pacoh. Ces deux populations ont un grand nombre de clans par rapport aux autres populations unilinéaires ce qui témoigne possiblement de l'existence d'une dynamique de fission particulièrement active.

Dans le chapitre 2, nous nous sommes intéressés à l'effet des différences entre systèmes de parenté mise en évidence dans le chapitre 1 sur la diversité uniparentale. Nous avons montré que les différences en termes de taux de migration sexe spécifiques se reflètent dans la diversité uniparentale. La diversité de l'ADN mitochondrial est plus élevée chez les populations P que M comme l'est le taux de migration de femmes. La diversité du chromosome Y est en revanche similaire entre les systèmes de parenté, comme l'est le taux de migration d'hommes. Nous avons aussi montré que la dynamique de fission linéaire plus active chez les Prai et les Pacoh semble générer dans ces populations une diversité mitochondriale (pour les Prai) et du chromosome Y (pour les Pacoh) particulièrement faible.

Enfin, dans le chapitre 3, nous avons étendu notre étude à l'effet du système de parenté sur la diversité autosomale. Deux questions nous ont intéressés : pouvons-nous détecter l'effet de la préférence pour des mariages entre cousins de premier degré a priori plus importante chez les populations unilinéaires (matrilinéaires et patrilinéaires) comparé aux populations cognatiques ? Pouvons-nous détecter l'effet du « puzzle matrilinéaire » qui entraine une endogamie plus importante chez les populations M et C par rapport aux populations P ?

La littérature ethnographique prédit qu'il existe une préférence plus importante pour les mariages entre cousins chez les populations unilinéaires comparées aux populations cognatiques. Nos données ethno-démographiques ne nous ont pas permis de vérifier si c'est également le cas des populations étudiées dans cette thèse. Nous avons malgré tout comparé les niveaux de consanguinité entre populations, sous l'hypothèse qu'une forte fréquence de mariage entre cousins augmente la consanguinité des populations. Cependant, nous n'avons pas détecté de différences en termes de consanguinité entre les populations unilinéaires et cognatiques. En revanche, les populations M et C, qui ont un taux d'endogamie de village plus élevé, ont une consanguinité plus élevée que les populations P et le taux de consanguinité moyen des populations est significativement corrélé au taux d'endogamie de village. Ainsi, nos résultats montrent pour la première fois qu'il existe un effet de l'endogamie de village possiblement lié au « puzzle matrilinéaire » sur la consanguinité.

Ce travail de thèse avait pour objectif d'approfondir nos connaissances sur l'influence du système de parenté sur la diversité génétique. Il a été rendu possible par l'approche pluridisciplinaire croisant des données génétiques et ethno-démographiques mis en œuvre par les membres de mon équipe d'accueil Anthropologie évolutive de l'UMR 7206. Sans l'étude quantitative des variables ethno-démographiques, nous n'aurions pas pu détecter les asymétries existantes entre les systèmes de parenté patrilinéaires et matrilinéaires ; rendant ainsi plus

ardues, imprécises, et parfois fausses les interprétations des données génétiques sur la base d'une classification uniquement catégorielle des systèmes de parenté. Plus généralement, cette étude nous a permis de mettre en évidence l'existence d'une grande variabilité ethno-démographique entre les populations partageant le même système de parenté, démontrant ainsi la limite d'une approche se limitant à un seul aspect de la parenté, et argumentant pour une meilleur intégration des interactions entre résidence, filiation ou alliance.

En se fondant sur les résultats obtenus tout au long de cette thèse, nous pouvons penser à plusieurs pistes pour améliorer nos connaissances sur l'influence du système de parenté sur la diversité génétique. Tout d'abord, à l'échelle des populations que nous avons étudiées, l'obtention de données plus précises et complètes sur les choix de conjoint nous permettrait d'étudier l'impact de la préférence pour des mariages entre cousins sur la diversité génétique. Cela nécessite une meilleure appréhension ethnologique et linguistique de la parenté dans chacune des populations afin de revisiter les méthodes d'enquête relatives à cette question. L'idéale serait évidemment de retracer les liens généalogiques entre les conjoints afin d'étudier la préférence pour les mariages entre cousins sur plusieurs générations. Malheureusement le temps nécessaire à une telle enquête ethnologique est peu compatible sur le terrain avec celui de l'enquête génétique.

En s'éloignant des populations que nous avons étudiées dans cette thèse, une autre manière d'approfondir notre connaissance sur l'influence du système de parenté sur la diversité génétique serait d'étudier des populations ayant un système de parenté disharmonique, c'est-à-dire des populations matrilinéaires et patrilocales en plus de populations ayant des systèmes harmoniques. Dans les populations harmoniques, l'effet de la résidence et de la filiation sont synergiques, la patrilocalité et la patrilinéarité réduisent la diversité du chromosome Y et la matrilocalité et la matrilinéarité celle de l'ADN mitochondrial. Il est alors difficile de distinguer les deux effets. L'inclusion de populations avec un système de parenté matrilinéaire et patrilocal permettrait, en les comparant à des populations matrilinéaires et matrilocales, d'isoler l'effet de la résidence. En les comparant avec des populations patrilinéaires et patrilocales cela permettrait d'isoler l'effet de la filiation sur la diversité génétique.

Il serait aussi intéressant d'étudier dans d'autres régions du monde des populations matrilinéaires qui ont été jusqu'à présent peu utilisées dans les études de génétique des populations. En effet, nous avons supposé que la différence de consanguinité entre les systèmes de parenté était générée par la différence d'endogamie de village qui découlerait de l'existence du « puzzle matrilinéaire ». A notre connaissance, il s'agit de la première étude mettant en

évidence ce potentiel effet du « puzzle matrilinéaire ». Il est donc nécessaire de reconfirmer nos résultats chez d'autres populations matrilinéaires.

Nous avons mentionné rapidement dans l'annexe 2 que nous avions entrepris une ébauche d'approche par modélisation. Nous pensons que l'approche par modélisation prenant en compte le système de parenté des populations simulées serait un outil particulièrement puissant et complémentaire à l'approche pluridisciplinaire utilisée dans cette thèse dans l'étude de l'influence du système de parenté sur la diversité génétique. Nous pourrions alors étudier toutes les combinaisons de règles de filiation, résidence et alliance. Par ailleurs, il serait aussi possible de faire varier la flexibilité de ces différentes règles et étudier son impact sur la diversité génétique. Une approche similaire à celle proposée ici a été mis en œuvre pour étudier l'influence de la préférence de mariage avec la cousine croisée matrilatérale (la fille du frère de la mère) sur la diversité génétique (Guillot et al., 2015). Cette étude montre que cette préférence de mariage réduit la diversité génétique dans la population mais uniquement lorsqu'elle est suivie strictement. Lorsque tous les individus ne suivent pas la préférence, l'effet sur la diversité génétique s'estompe.

Ainsi, nous pouvons nous diriger dans différentes directions dans le but d'approfondir nos connaissances sur l'influence du système de parenté sur la diversité génétique : mieux caractériser les populations étudiées, utiliser les mêmes méthodes dans d'autres régions du monde sur des populations ayant des systèmes de parenté différents ou développer une autre approche fondée sur la modélisation.

Plus généralement, ce travail de thèse a permis d'approfondir notre connaissance sur l'influence du système de parenté sur la diversité génétique des populations humaines. Une partie des résultats confirment les observations faites dans d'autres populations, comme l'impact de la règle de résidence sur la diversité uniparentale, mais nous montrons qu'il est nécessaire de quantifier le comportement ethno-démographique pour pouvoir prendre la pleine mesure de l'influence du système de parenté. De plus, nous ajoutons un niveau de complexité supplémentaire en montrant que la matrilocalité est associée à l'endogamie de village, ce qui augmente la consanguinité dans les populations matrilocales actuelles. Ce résultat nous paraît important dans le contexte de l'étude de l'évolution génétique des populations humaines à long terme. En effet, il semblerait que les populations prénéolithiques étaient moins patrilocales que les populations actuelles, avec possiblement une fréquence plus importante de multilocalité et de matrilocalité (Marlowe, 2004). L'impact de ces modes de résidence passés sur les profils génétiques uniparentaux a été discutée (Rasteiro et al., 2012). Nos résultats suggèreraient donc

que ces modes de résidence pourraient avoir façonné les profils de consanguinité génétique de nos ancêtres. Ainsi, il nous semble que l'intégration des systèmes de parenté dans les modèles de génétique des populations humaines permettrait de mieux comprendre l'évolution génétique humaine à court mais aussi long terme.

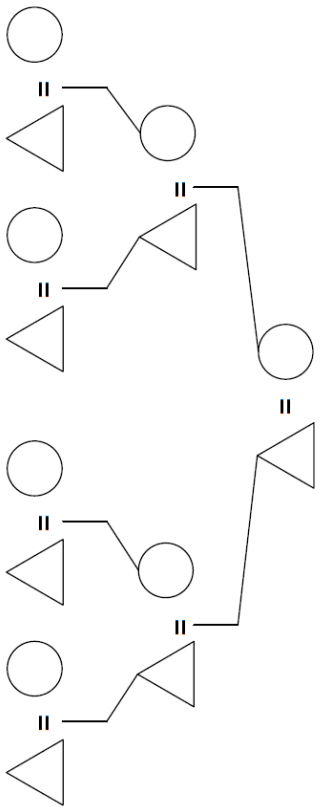
Annexe

Annexe 1 - Questionnaire ethno-démographique

A

Village: n° Family: GPS: 1- General information

Ego: M – F Age: Photos:



Résidence

Ego	Desc.G	Birth P.	Lang.	Spouse	Desc.G	Birth P.	Lang.
∟ EF				∟ SpF			
∟ EFF				∟ SpFF			
∟ EFM				∟ SpFM			
∟ EM				∟ SpM			
∟ EMF				∟ SpMF			
∟ EMM				∟ SpMM			

Annexe 1-A : Première page du questionnaire ethno-démographique. L'informateur nous renseigne sur son couple et les ascendants des deux conjoints.

B

2. Matrimonial and reproductive history of EGO

Rank Current Spouse: N of Spouse: N of children:

N°	Desc. G	Birth P	N°	Desc. G	Birth P
∟ SpF			∟ SpF		
∟ SpM			∟ SpM		
Poly – Rem.Div – Rem.DC		Link:	Poly – Rem.Div – Rem.DC		Link:
N°			N°		
∟ SpF			∟ SpF		
∟ SpM			∟ SpM		

Annexe 1-B : Deuxième page du questionnaire ethno-démographique. L'informateur nous renseigne sur son histoire reproductive.

C

3. The married children of EGO

Desc. G	Birth P	Desc. G	Birth P
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	

Annexe 1-C : Troisième page du questionnaire ethno-démographique. L'informateur nous renseigne sur les conjoints de ses enfants.

D

4. Matrimonial and reproductive history of EGO's PARENT

Rank Current Spouse: N of Spouse: N of children:

	Desc. G	Birth P		Desc. G	Birth P
N°			N°		
∟ SpF			∟ SpF		
∟ SpM			∟ SpM		
Poly – Rem.Div – Rem.DC		Link:	Poly – Rem.Div – Rem.DC		Link:
N°			N°		
∟ SpF			∟ SpF		
∟ SpM			∟ SpM		

Annexe 1-D : Quatrième page du questionnaire-ethno-démographique. L'informateur nous renseigne sur l'histoire reproductive d'un de ses parents (souvent présent lors de l'entretien).

E

5. The bothers and sisters of EGO

Desc. G	Birth P	Desc. G	Birth P
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	
∟ SpM		∟ SpM	
N°		N°	
∟ SpF		∟ SpF	

Annexe 1-E : Cinquième page du questionnaire ethno-démographique. L'informateur nous renseigne sur les conjoints de ses frères et sœurs.

Annexe 2 - Simulation de la dynamique de clan dans un village

1. Introduction

Nous présentons dans cette annexe un travail exploratoire que nous avons effectué au cours de la thèse. Il s'agit de la première étape d'un projet de construction d'un modèle de simulation de populations humaines intégrant les systèmes de parenté. Cette première étape modélise la dynamique des clans d'une population unilinéaire dans un village.

2. Matériels et méthodes

2.1 Présentation du modèle

Nous avons étudié la question de l'évolution du nombre de clans dans une population par une approche de modélisation agents-centrée (*agent-based model*) à générations non chevauchantes. Nous simulons des individus vivants tous dans le même village et chacun appartenant à un groupe de filiation (clan) qui est transmis du père aux enfants (filiation patrilinéaire). A chaque génération, les individus forment des couples et engendrent des enfants qui seront les individus de la génération suivante. Les couples sont formés, soit en suivant une règle d'évitement de l'inceste, soit en suivant une règle d'exogamie de clan. Dans le premier cas, un homme ne peut pas se marier avec ses sœurs. Dans le second cas, un homme ne peut pas se marier avec une femme appartenant à son clan (ses sœurs appartiennent évidemment à son clan).

2.2 Génération des couples et formation de la génération suivante

A chaque nouvelle génération, les individus sont initialement non appariés. Nous tirons aléatoirement un individu et un conjoint lui est ensuite assigné aléatoirement parmi tous les conjoints possibles selon la règle simulée (interdiction d'inceste ou exogamie de clan). S'il n'existe pas de conjoint possible pour un individu, nous le marquons comme célibataire. Il ne participera pas à la formation de la génération suivante. Lorsque tous les individus ont trouvé

un conjoint ou ont été marqués comme célibataire, nous passons à la phase de création de la génération suivante.

Pour chaque couple, le nombre d'enfants a été tiré aléatoirement dans une loi de Poisson de paramètre λ. Le paramètre de la loi de Poisson λ dépend du nombre de couples formées.

$$\lambda = \frac{Taille\ initiale\ du\ village}{Nombre\ de\ couples\ formé}$$

Cette paramétrisation permet de garder constant la taille du village sur le long terme, tout en autorisant de petites fluctuations à chaque génération autour de la taille initiale du village. Le sexe des enfants a été choisi de manière équiprobable et les enfants ont été assignés au clan de leur père (mais le sexe n'ayant aucune influence sur le modèle, les résultats sont généralisables aux populations matrilinéaire).

2.3 Paramétrisation du modèle

Nous avons simulé des scénarios considérant des tailles de village différentes (1400, 700, 350 et 140) et intégrant ou non l'exogamie de mariage ou l'interdiction d'inceste (8 scénarios différents au total). Le nombre initial de clans a été fixé à 20. A la première génération, les individus sont répartis de façon égale dans les différents clans du village. Nous simulons 100 générations, ce qui représente 2500 ans si nous supposons un temps de génération de 25 ans. Chaque scénario a été simulé 100 fois.

2.4 Simulation de l'apport de nouveau clan dans le village

Nous avons simulé l'apparition de nouveaux clans dans le village en assignant un nouveau clan à un homme pris au hasard dans la population, et ce à chaque génération. Nous avons alors analysé la dynamique du nombre de clans avec le temps pour chaque scénario.

Les simulations ont été codées en perl v5.16.3 sous la distribution ActivePerl pour Windows.

3. Résultats

La figure S2-1 présente l'évolution du nombre de clans dans les villages pour les différents scénarios simulés. Dans tous les scénarios, le nombre de clans dans le village diminue au cours du temps. La vitesse de déclin dépend évidemment de la taille du village. Plus le village est grand, plus la diminution est lente. C'est un résultat connu en génétique des populations : si nous considérons que les clans sont des allèles d'un locus donné, plus la taille de la population (ici le village) est petite, plus la dérive génétique (ici de clan) est forte, et donc plus la réduction de la diversité génétique (ici le nombre de clans) au cours du temps est rapide.

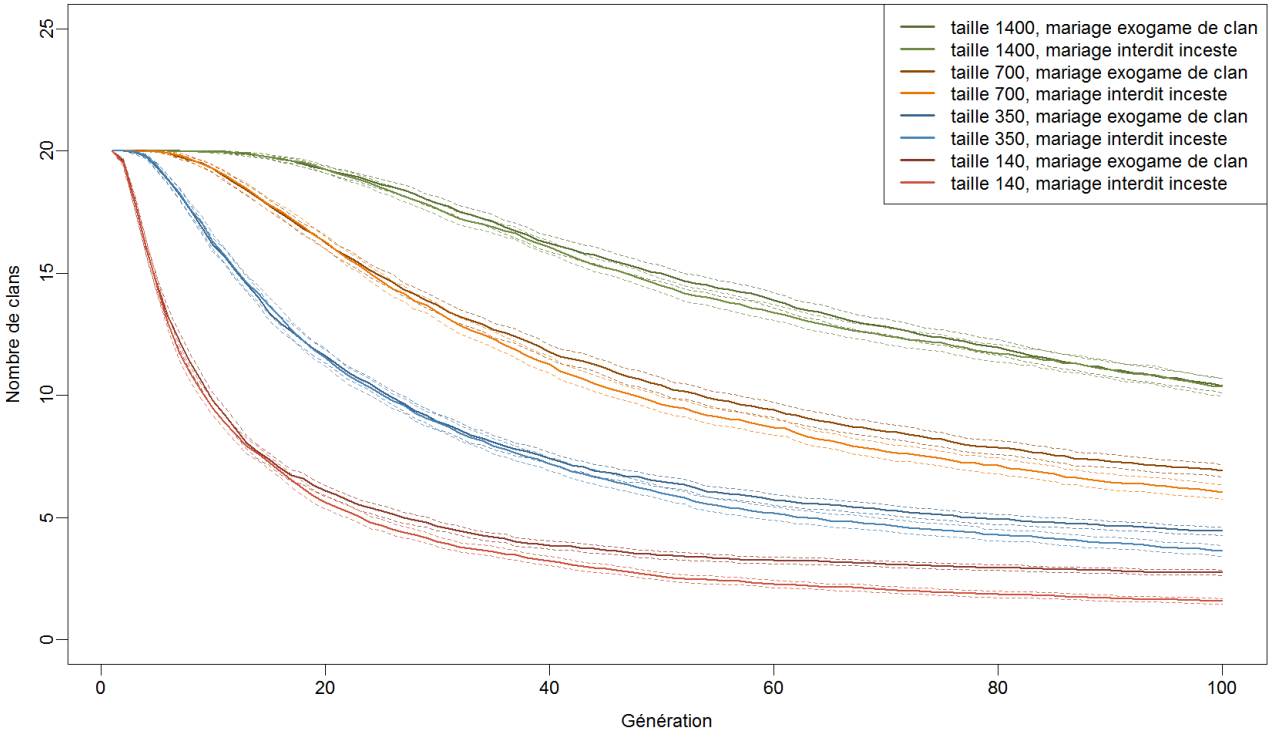


Figure S2-1 : Evolution du nombre de clans au cours du temps. Le nombre moyen de clans sur les 100 simulations pour les 8 scénarios différents sont représentés par des lignes de couleurs différentes. Les lignes en pointillés représentent les intervalles de confiance à 95% de la valeur moyenne pour chaque scénario.

Ce qui est plus intéressant est la comparaison des deux scénarios qui diffèrent par la règle de mariage en vigueur (exogamie de clan ou non). Nous pouvons voir que l'exogamie de clan permet de maintenir un nombre de clans légèrement supérieur dans le temps pour les trois tailles de village les plus petites. Les scénarios avec exogamie de clan ont un nombre de clans significativement supérieur aux scénarios sans exogamie à partir de la génération 26, 65 et 81 respectivement pour les villages de taille 140, 350 et 700 (p-values <0,05 après correction de

Bonferroni pour tests multiples). En outre, cet effet est d'autant plus grand que la taille du village est petite.

Nous nous intéressons maintenant au cas dans lequel de nouveaux clans apparaissent dans un village où l'exogamie de clan structure les mariages (Figure S2-2). Nous ne présentons pas les résultats sur les villages avec uniquement l'interdit de l'inceste car les différences entre les scénarios avec la règle d'exogamie de clan sont similaires à celles présentées précédemment. Sur le long terme, l'apport d'un nouveau clan à chaque génération dans le village, permet de maintenir un plus grand nombre de clans et ce de façon plus stable au cours du temps. Cependant, cet apport seul ne permet pas de maintenir le nombre de clans à la valeur initiale de 20. L'apport de nouveau clan permet de maintenir le nombre de clans à 12,5 dans les villages 350 individus et à 9,6 dans les villages de 140 individus.

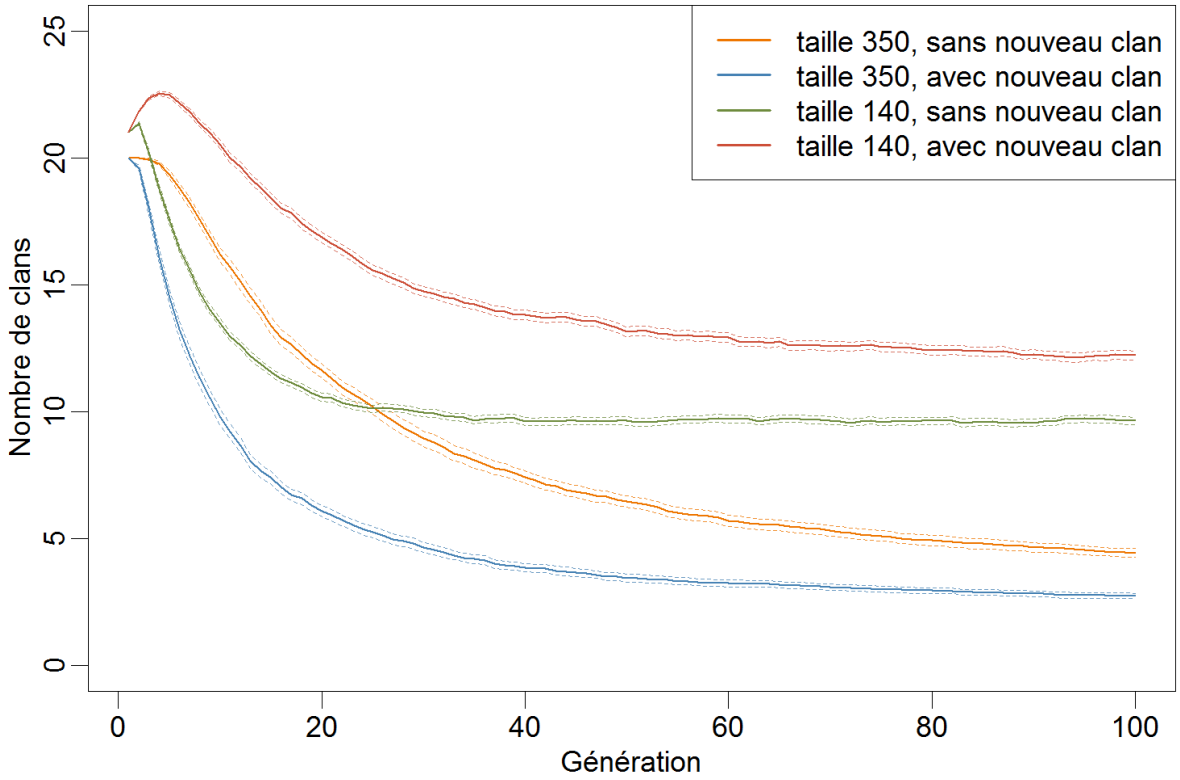


Figure S2-2 : Evolution du nombre de clans au cours du temps avec ou sans apport de nouveau clan dans le village. Le nombre moyen de clans sur les 100 simulations pour les 4 scénarios différents sont représentés par des lignes de couleurs différentes. Les lignes en pointillés représentent les intervalles de confiance à 95% de la valeur moyenne pour chaque scénario.

4. Conclusion

Par ce travail exploratoire sur la dynamique des clans, nous mettons en évidence que :

- La règle d'exogamie de clan permet de maintenir au cours du temps un nombre de clans plus élevés pour des villages de tailles inférieures à 700 individus.
- L'apport d'un nouveau clan à chaque génération permet de maintenir un nombre de clans constant au sein du village (12,5 dans les villages de 350 individus et 9,6 dans les villages de 140 individus).

Annexe 3 - Distribution des clans dans les villages Pacoh et Prai

Le premier diagramme circulaire sur chaque page présente la répartition des clans dans la population en entier. Les diagrammes suivants présentent la répartition des clans dans chaque village de la population. Les couleurs sont conservées entre les diagrammes (un clan aura la même couleur dans tous les diagrammes). Le nombre d'individu appartenant au clan est indiqué à l'extérieur de chaque section.

La figure S3-1 sur la page suivante présente la répartition des clans chez les Pacoh. La figure S3-2 sur la page d'après, la répartition des clans chez les Prai.

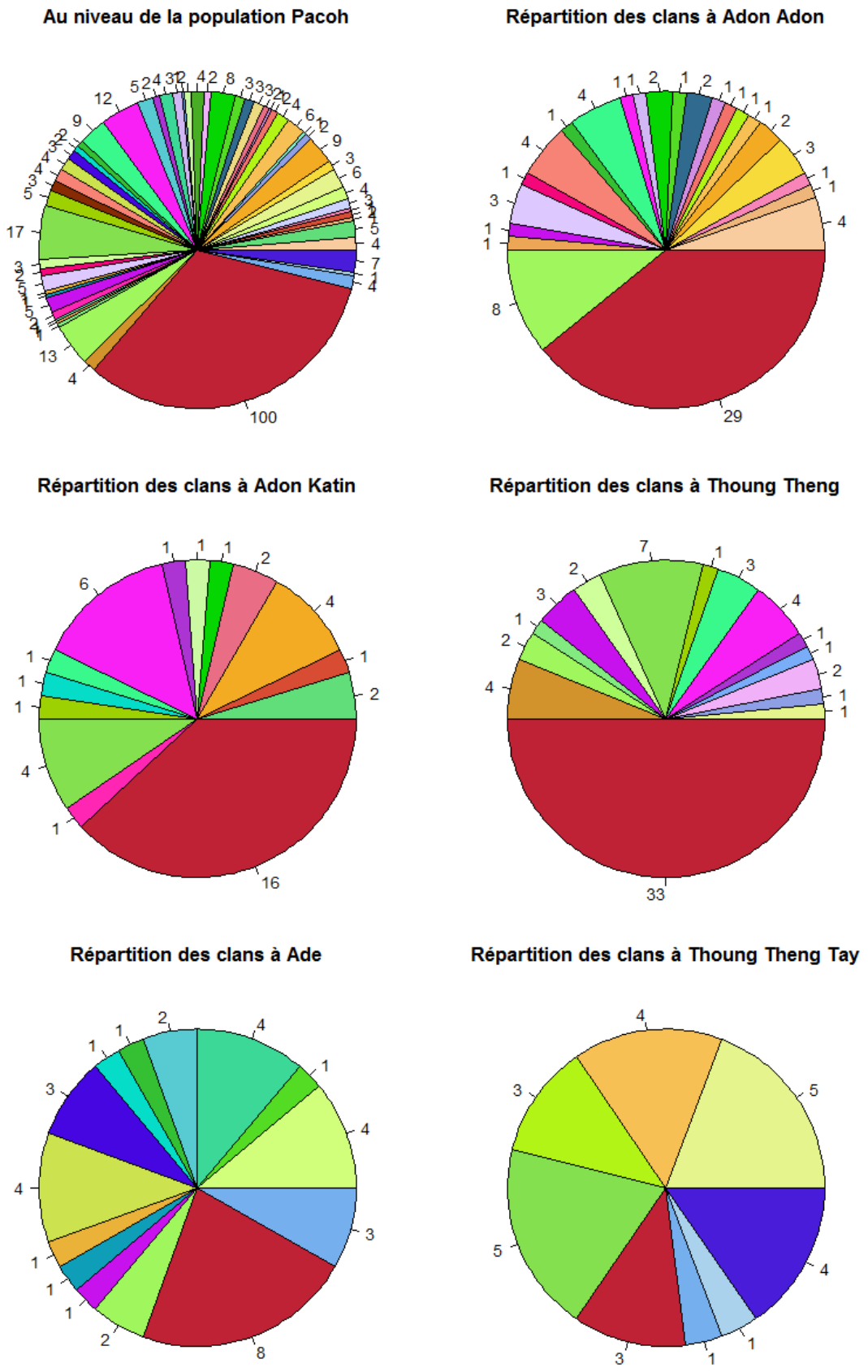


Figure S3-1 : Répartition des clans dans les villages Pacoh

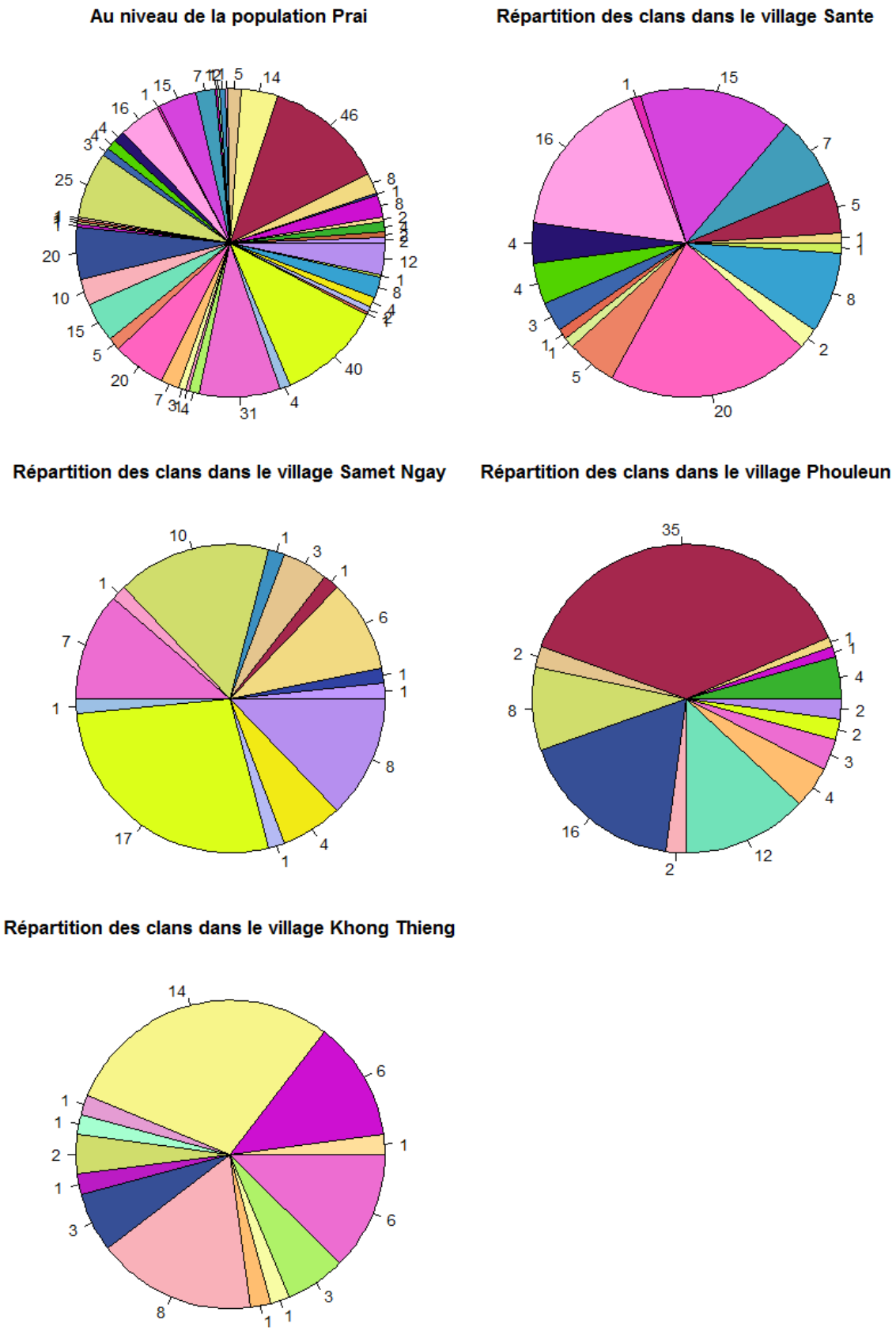


Figure S3-2 : Répartition des clans dans les villages Prai

Annexe 4 - Article : Matrilineal puzzle shapes genetic inbreeding patterns in human populations

Introduction

In matrilineal populations, which represent 12 to 17% of the world populations (1, 2), descent and inheritance are transmitted through women. However, in most cases, authority over matrilineal groups and political power at the village level lies in the hands of men (3–5). Matrilineal descent is usually combined with matrilocal residence (the husband settles with his wife's family after marriage), and men from such societies face a dilemma that has been named the "matrilineal puzzle" (6, 7)): by following the matrilocal residence rule, men move out to settle with their wife, possibly in a different village, which makes it harder form them to exert their influence on their natal descent group and natal village. The further away they marry, the harder it will be for them to maintain their influence over their natal descent group and village (8). Consequently, in these societies, men tend to be reluctant to marry far away from their natal village, and a significant preference for marrying within the same village has been reported in these societies (9, 10).

Although this preference has been known for over 50 years, no study has addressed the question of its potential impact on the genetic evolution of these populations. Indeed, it is likely that such preference for local endogamy would increase the proportion of consanguineous marriages (11) and thus the genetic inbreeding level in these matrilineal populations in comparison to populations where such preference does not exist, in particular patrilineal and cognatic populations.

In the past years, new tools have been made available to evaluate inbreeding level from genome-wide single nucleotide polymorphisms (SNPs) data (12–14). In this study, we propose to revisit

the matrilineal puzzle hypothesis using such genetic tools and more precisely to test whether matrilinearity is indeed associated with higher inbreeding levels. To do so, we collected ethno-demographic data as well as high density autosomal SNP data from 12 populations from South-East Asia exhibiting a wide variety of social organizations. More precisely, we sampled four populations (M) with matrilineal descent and matrilocal residence (Jarai, Tampuan, Prai and Kacho'), and four populations (P) with patrilineal descent and patrilocal residence (Khmu', Lamet, Ta-oih and Pacoh). We included to this dataset four populations (C) with cognatic descent and predominantly matrilocal residence (Khmer, Bunong, Brao and Kreung) in order to test whether matrilocality alone could generate a similar level of constraints on men dispersal as than when it is associated with matrilinearity. Indeed, in populations with no matrilineal descent group (cognatic populations), but following a matrilocal residence rule, we could expect men to be equally reluctant to marry far away from their natal village, in order to keep the influential status that they have acquired in their natal village. All these populations live in tropical environments, have similar lifestyles and speak related languages from the Austro-Asiatic family (with the exception of the Jarai who speak an Austronesian language). We analyzed ethno-demographic data collected for over 3500 couples to estimate the proportion of village endogamy under each social organization and we analyzed genetic data to investigate inbreeding levels in these populations.

Results

In this study, we focus on the influence of social organization on inbreeding levels by comparing South-East Asian populations with matrilineal descent and matrilocal residence (M), cognatic descent and predominant matrilocal residence (C) and patrilineal descent and patrilocal residence (P).

Estimation of effective population size

First of all, we checked that all the studied populations have comparable effective sizes since this parameter is known to influence inbreeding levels (15), with smaller effective population sizes associated with higher inbreeding. Among all studied populations (table 1), Prai had a significantly lower effective population size compared to all other populations (7181 compared to 13 228 (s.d.±1403) in average for the other populations; p-values < 0.05; table 2). This suggests a particular demographic history for this population (potentially as a result of past bottlenecks). Consequently, we removed this population from our dataset.

Estimation of village endogamy

We estimated village endogamy rates from 532 interviews conducted by our research team, providing ethno-demographic information for 3530 couples (figure 1). When both spouses were born in the same village, we considered the couple as being endogamous at the village level. Village endogamy rates were compared using generalized linear mixed model. As expected under the matrilineal puzzle hypothesis, village endogamy was significantly higher in M than P populations (0.87 versus 0.73 receptively, p-value = 0.029). In C populations, village endogamy was similar to M populations (0.84, p-value = 0.62) and the difference with P populations was marginally significant (p-value = 0.066). This suggests that the matrilocal

residence alone (with no matrilineal descent groups but cognatic descent) generates a similar level of constraints on men dispersal as this rule associated with matrilineal descent groups.In addition, one-way analysis of variance (ANOVA) at the population level indicated that social organization explained 43% of the variance in village endogamy rate thus showing that social organization is a good predictor of village endogamy.

Estimation of inbreeding coefficients

We tested whether M and C populations exhibit higher inbreeding levels in comparison to P populations as a result of their higher local endogamy prevalence. To do so, we used the FEstim software (13) in the FSuite pipeline (16). M and C populations had similar mean inbreeding coefficients (0.018 and 0.017 respectively, p-value = 0.99, figure 2), and both had higher mean coefficient compated to P populations (0.011, both p-values < 0.05). In addition, one way ANOVA indicated that 27% of the variance in inbreeding coefficients was explained by social organization at the population level.

To further investigate the influence of village endogamy on inbreeding level and confirm that village endogamy is the social component that explained the difference in proportion of inbred individuals between social organizationwe measured the correlation coefficient between these two parameters at the population level (figure 3): village endogamy rate was significantly correlated to inbreeding coefficient (Spearman's $\rho = 0.71$, p-value = 0.02).

Discussion

In this study, we tested whether the preference for local endogamy in matrilineal populations reported over 50 years ago by ethnographic works (8, 9) could be found in our South East Asian datasets and whether it shaped genetic inbreeding levels. Analyzing a data set of 11 populations having different social organizations, we observed that the matrilineal and matrilocal populations (M) under study had higher level of village endogamy than patrilineal and patrilocal populations (P) and that such higher village endogamy reflected itself into higher mean inbreeding coefficient. Interestingly, we did not see higher endogamy and inbreeding in M populations in comparison to cognatic populations with predominant matrilocal residence (C). In addition, we showed that social organization is a good predictor of inbreeding levels in these populations.

The higher endogamy and genetic inbreeding in matrilineal populations likely resulted from what have been called the matrilineal puzzle. In matrilineal populations, under the matrilocal residence rule, men settle with their wife's family and divide authority with their brother or father in law while they want to maintain their influence in their natal kin group and village. The further away they marry, the more difficult it will be to maintain their authority position in their natal village. Consequently, they tend to favor local marriages. Ethnographic works in the matrilineal populations under study confirmed the existence of such preference: the Jarai marry according to a "the closest, the safest" rule (17) and in the Tampuan population, the councils of elders are reluctant to integrate in their village a man coming from a distant village, a preference still in action these days (Bourdier personal communication).

In patrilineal and patrilocal populations, such puzzle does not occur as most men settle with their wife in their natal village whether they marry with a woman from the same or from another village, with no risk of losing their position of influence or leadership. This leads to comparatively higher rate of village exogamy and lower inbreeding levels in these patrilineal

populations. In our dataset, we noticed one exception to this general observation: the Khmu', a patrilineal and patrilocal population, exhibits a matrilineal-like rate of village endogamy and a matrilineal-like inbreeding level. The reasons for these differences with the other P populations remain to be investigate.

As pointed out above, the cognatic populations (C) under study exhibit a similar rate of village endogamy and inbreeding as the matrilineal populations (M). These cognatic populations follow predominantly a matrilocal residence rule. To be more precise, two of them (Khmer and Bunong) are matrilocal and two of them (Brao and Kreung) are multilocal with final settlement in the wife's family. Our results suggest that men from these societies are similarly discouraged to disperse to another village as in matrilineal and matrilocal societies despite the absence of descent groups. It is most likely in order to keep their influence on their natal village. In the two multilocal populations, the couple live alternatively with the husband's and wife's family, which seems to constrain men dispersal as efficiently as in matrilocal populations. In addition, the inconvenience of having to move from one village to another when going from one family's home to the other family's home could have also favored local endogamy. As such, the "matrilineal puzzle" seems to not only to affect matrilineal populations but all types of matrilocal as well as multilocal populations. Consequently, it could be renamed the "matrilocal puzzle" to express the fact that it affects not only matrilineal populations, but more generally matrilocal populations.

Village endogamy may not be the only coping mechanism to the matrilineal puzzle. An exemption of matrilocality for one of an only brother sibship of a matrilineage who stays in his natal village to deal with his descent group matters, may be an alternative mechanism (3, 6). However, our ethno-demographic data do not support such hypothesis since a lower exemption to matrilocality was observed in matrilineal and matrilocal populations than exemption to patrilocality in patrilineal and patrilocal populations.

Our interdisciplinary study not only support the matrilineal puzzle hypothesis but also suggest that such puzzle shapes genetic diversity patterns in some human populations, thus identifying a new factor linked to social organization that may contribute to genetic diversity patterns among human populations. Influence of social organization on genetic diversity have drawn the attention of human population geneticist for the past decades and the impact of descent (18–20), residence (21–26), alliance rules (27–29) have been extensively documented (30). It has previously been shown that genetic inbreeding levels are greatly influenced by the prevalence of consanguineous marriages in human populations (12, 31–36). Our study shows that the association of matrilocality with local endogamy may also contributes to some extent to higher inbreeding levels in human populations, thus adding an additional layer of complexity to the interactions between socio-cultural factors and human genetic diversity patterns.

Method

Sampled populations

Twelve populations from Cambodia and Laos were sampled in 61 villages during 3 field missions carried out between 2011 and 2012: the Tampuan, Jarai, Kacho', Bunong, Khmer, Brao and Kreung from Cambodia and the Khmu', Ramet, Ta-oih, Pacoh and Prai from Laos. The populations were chosen for their differences in social organizations. Most of them have been the focus of ethnographic works, describing in details their social organization (Table 1). The Tampuan, Jarai, Kacho' and Prai have matrilineal descent and matrilocal residence (5, 17, 37, 38), the Bunong and the Khmer have cognatic descent and matrilocal residence (39–42) the Brao and Kreung have cognatic descent and multilocal residence (38, 43–45) and the Khmu', Ramet, Ta-oih and Pacoh have patrilineal descent and patrilocal residence (38, 46–49).

Ethno-demographic data

We interviewed 532 individuals, conjointly with their spouse when possible, and collected ethno-demographic information for them, their spouse and their family members (parents, grand-parents, siblings, children and their respective spouses). This procedure allowed us to gather ethno-demographic information for 3530 couples. We estimated the proportion of couples in which both spouses were born in the same village (village endogamy). We used logistic regression to assess the influence of social organization (M, P and C) on village endogamy rates with the 'glmer' function in 'lme4' package v1.1-9 in R (50). We incorporated population as a fixed effect and village of residence and family as random effects in this model in order to take into account potential bias stemming from our sampling scheme. P-values were estimated with the 'lsmeans' function in the 'lsmeans' package v2.27-2 in R. We performed an

one-way analysis of variance (ANOVA) to estimate the proportion of variance of village endogamy explained by social organisation (M, P and C) at the population level.

DNA samples

Individuals with all four grandparents from the same population were sampled. When possible, we sampled a couple with one of their child. We collected two saliva samples for each individual (4mL each). Samples were kept in equivalent volume of lysis buffer with 300µL of 10% SDS and 60µL of proteinase K (20mg/mL). DNA was extracted from saliva samples using a standard ethanol precipitation protocol. All participants provided written informed consents and the study was approved by the National Ethic Comities for Health Research in Cambodia and Laos as well as by the Comité Opérationnel pour l'Ethique (CNRS, France).

SNP genotyping

Samples were genotyped on Illumina Omni1 (529 individuals) and Omni2.5 SNP arrays (226 individuals). SNPs present on both chips were retained, leading to a data set of 701 163 SNPs for 753 individuals. After quality control (Figure S1), the dataset contained 598 764 SNPs for 743 individuals.

We then used the method described in Conomos et al. (51) in order to check if any siblings were present in the dataset. We removed 24 individuals in order to get a sibling free dataset. The dataset contained 598 764SNPs for 719 individuals.

Effective population size

In order to verify that all populations had similar past demographic history, we estimated the effective population size of each population using the method described in Auton et al. (52)

(table 2). Effective population sizes were compared between populations by a Welch's t-test with a Bonferroni correction for multiple testing.

To disentangle the influence of social organization from the influence of other factors, such as variation in effective sizes that could also impact inbreeding levels, we removed the Prai population from the dataset. The final genetic dataset included 675 individuals from 11 populations.

Genetic inbreeding coefficients estimation

We used the FEstim software (13) integrated in the FSuite pipeline (16) to estimate the inbreeding coefficients, test whether individuals were inbred by a likelihoo-ration test and infer the most likely parental mating type for each individual. Genetic map were retrieved from the shapeit homepage. The --hotspots option with the hg19 build was used when creating the 100 submaps. In order to estimate allele frequencies, we generated a dataset with only unrelated individuals at the second cousin level We estimated Rousset's relatedness coefficient (53) for all pairs of individuals in each population. For each pair with a coefficient higher or equal to 1/64 (corresponding to the coefficient expected between second cousins), we removed one of the individuals, drawn at random. This step was repeated until no pair had a coefficient higher than or equal to the limit. The unrelated dataset included 457 individuals on which we estimated allele frequencies. Inbreeding coefficients for the 675 individuals included in the final dataset were estimated using the allele frequency generated using the unrelated dataset.

We then used a generalized linear model to assess the influence of social organization (M, P and C) on mean inbreeding coefficients. As for the village endogamy rates, we incorporated population as a fixed effect and village of residence and family as random effects in this model in order to take into account potential bias stemming from or sampling scheme. P-values were estimated with the 'lsmeans' function in the 'lsmeans' package v2.27-2 in R.

We performed an one way ANOVA to estiamte the proportion of variance of mean inbreeding coefficient explained by social organization (M, P and C) at the population level.

Pearson's correlation coefficient between village endogamy rate and mean inbreeding coefficient was estimated at the population level.

All statistical analysis were performed in R v3.2.2 (50).

References

1. Godelier M (2004) *Métamorphoses de la parenté* (Fayard, Paris).

2. Murdock GP, White DR (1969) Standard Cross-Cultural Sample. *Ethnology* 8:329‑369.

3. Fox R (1972) *Anthropologie de la parenté. Une analyse de la consanguinité et de l'alliance* (Gallimard, Paris).

4. Zelditch M (1955) Role differentiation in the nuclear family: a comparative study. *Family Socialization and Interaction Process*, éd Parsons T, Bales RF (Free Press, Glencoe), p 307‑352.

5. Bourdier F (2006) *The Mountain of Precious Stones* (Phnom Penh: Center for Khmer Studies., Ratanakiri: Cambodia: Essays in Social Anthropology).

6. Richards AI (1950) Some Types of Family Structure Amongst the Central Bantu. *African Systems of Kinship and Marriage*, éd Radcliffe-Brown A, Forde D (Oxford University Press, London), p 207‑251.

7. Schneider DM (1961) The distinctive features of matrilineal descent groups. *Matrilineal Kinship*, éd Schneider DM, Gough K (University of California Press, Derkeley), p 1‑29.

8. Kloos P (1963) Matrilocal residence and local endogamy: Environmental knowledge or leadership. *Am Anthropol* 65(4):854‑862.

9. Murdock GP (1949) *Social Structure* (Macmillan company, London).

10. Johnston FE, Kensigner KM, Jantz RL, Walker GF (1969) The population structure of the Peruvian Cashinahua: Demographic, genetic and cultural interrelationships. *Hum Biol* 41(1):29‑41.

11. Gilbert JP, Hammel E a. (1966) Computer Simulation and Analysis of Problems in Kinship and Social Structure. *Am Anthropol* 68(1):71‑93.

12. Leutenegger A-L, Sahbatou M, Gazal S, Cann H, Génin E (2011) Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? *Eur J Hum Genet* 19(5):583‑7.

13. Leutenegger A-L, et al. (2003) Estimation of the inbreeding coefficient through use of genomic data. *Am J Hum Genet* 73(3):516‑23.

14. Leutenegger A-L, et al. (2006) Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. *Am J Hum Genet* 79(1):62‑6.

15. Jacquard A (1968) Evolution des populations d'effectif limité. *Popul (French Ed* 23(2):279‑300.

16. Steven G, Mourad S, Marie-Claude B, Emmanuelle G, Anne-Louise L (2014) FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics* (here cM):1‑2.

17. Dournes J (1972) *Coordonées, structures jörai familiales et sociales* (Muséum national

d'Histoire naturelle (Travaux et Mémoires de l'Institut d'Ethnologie n°77), Paris).

18. Chaix R, et al. (2007) From social to genetic structures in central Asia. *Curr Biol* 17(1):43‑48.

19. Ségurel L, et al. (2008) Sex-specific genetic structure and social organization in Central Asia: Insights from a multi-locus study. *PLoS Genet* 4(9):e1000200.

20. Sanchez-Faddeev H, et al. (2013) The influence of clan structure on the genetic variation in a single Ghanaian village. *Eur J Hum Genet* 21(10):1134‑1139.

21. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat Genet* 29(1):20‑1.

22. Hamilton G, Stoneking M, Excoffier L (2005) Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc Natl Acad Sci U S A* 102(21):7476‑7480.

23. Besaggio D, et al. (2007) Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evol Biol* 7 Suppl 2(Suppl 2):S12.

24. Kumar V, et al. (2006) Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genet* 2(4):420‑424.

25. Gunnarsdóttir ED, et al. (2011) Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. *Nat Commun* 2:228.

26. Rasteiro R, Bouttier P -a., Sousa VC, Chikhi L (2012) Investigating sex-biased

migration during the Neolithic transition in Europe, using an explicit spatial simulation framework. *Proc R Soc B Biol Sci* 279(1737):2409‑2416.

27. Neel J V (1970) Lessons from a « primitive » people. *Science (80-)* 170(960):815‑822.

28. Guillot EG, et al. (2015) Relaxed Observance of Traditional Marriage Rules Allows Social Connectivity without Loss of Genetic Diversity. *Mol Biol Evol* 32(9):2254‑2262.

29. Kayser M, et al. (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72(2):281‑302.

30. Heyer E, Chaix R, Pavard S, Austerlitz F (2012) Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 21(3):597‑612.

31. Bittles AH, Hamamy H (2010) Endogamy and Consanguineous Marriage in Arab Populations. *Genetic Disorders Among Arab Populations*, éd Teebi AS (Springer Berlin Heidelberg), p 85‑108.

32. Gazal S, Sahbatou M, Babron M-C, Génin E, Leutenegger A-L (2015) High level of inbreeding in final phase of 1000 Genomes Project. *Sci Rep* 5:17453.

33. Pemberton TJ, et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91(2):275‑92.

34. Kirin M, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.

35. McQuillan R, et al. (2008) Runs of homozygosity in European populations. *Am J Hum Genet* 83:359‑372.

36. Yang X, et al. (2014) The influence of admixture and consanguinity on population

genetic diversity in Middle East. *J Hum Genet* 59(11):615‑622.

37. Dessaint WY (1981) The T'in (Mal), dry rice cultivators of Northern Thailand. *J Siam Soc* 69:107‑137.

38. LeBar M, Hickey GC, Musgrave JK (1964) *Ethnic groups of mainland Southeast Asia* (Human Relations Area Files).

39. Ebihara M (1977) Residence Patterns in a Khmer Peasant Village. *Ann N Y Acad Sci* 293(1):51‑68.

40. Ledgerwood JL (1995) Khmer Kinship - the Matriliny Matriarchy Myth. *J Anthropol Res* 51(3):247‑261.

41. Martel G (1975) *Lovea. Village des environs d'Angkor. Aspects démographiques, économiques et sociologiques du monde rural cambodgien dans la province de Siem-Réap* (Ecole française d'Extrême-Orient, Paris).

42. UNDP Cambodia (2010) *Phnong Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village* (Phnom Penh).

43. Baird IG (2008) Various Forms of Colonialism : the Social and Spatial Reorganisation of the Brao in Southern Laos and Northeastern Cambodia. Dissertation (University of British Columbia).

44. Matras-Troubetzkoy J (1983) *Un Village En Foret. L'essartage Chez Les Brou Du Cambodge* éd Selaf (Peeters, Paris).

45. UNDP Cambodia (2010) *Kreung Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village* (Phnom Penh).

46. Evrard O (2006) *Chroniques des cendres* (IRD Edition, Paris).

47. Izikowitz KG (1951) *Lamet hill peasants in French Indochina* (Etnografiska Museet, Göteborg).

48. Schmutz J (2013) The Ta'Oi Language and People. *Mon-Khmer Stud* 42:i‑xiii.

49. Lindell K, Samuelsson R, Tayanin D (1979) Kinship and Marriage in Northern Kammu Villages: The Kinship Model. *Sociologus* 29(1):60‑84.

50. R Development Core Team (2008) R: A Language and Environment for Statistical Computing.

51. Conomos MP, Reiner AP, Weir BS, Thornton TA (2016) Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet* 98(1):127‑148.

52. Auton A, McVean G a T (2007) Recombination rate estimation in the presence of hotspots. *Genome Res* 17:1219‑1227.

53. Rousset F (2002) Inbreeding and relatedness coefficients: what do they measure? *Heredity (Edinb)*:371‑380.

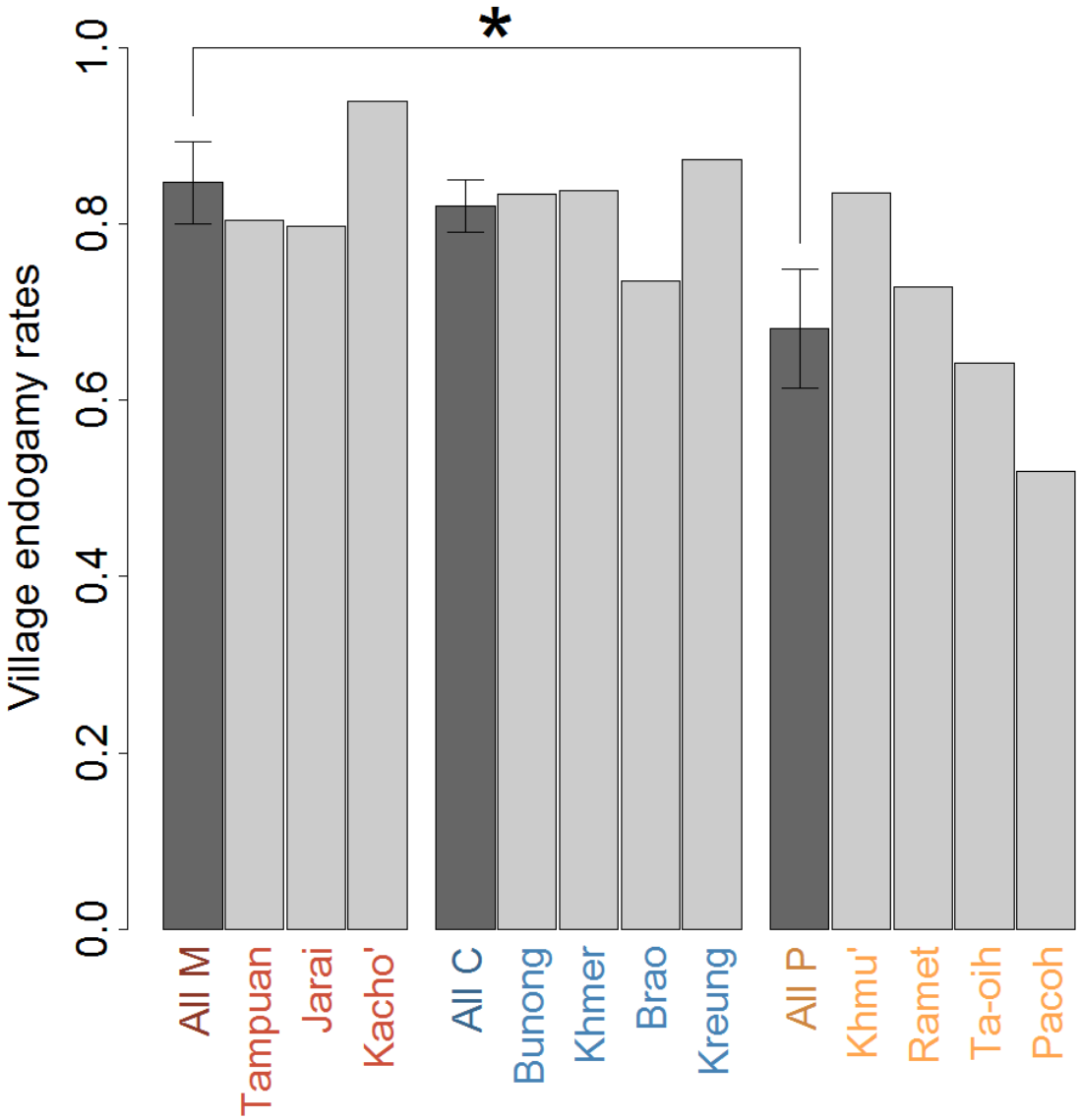


Figure 1: Village endogamy rates. The first dark bar in each group represents the mean in each group with standard error. Asterisks indicate statistical significance (p-value < 0.05) assessed by generalized linear mixed model.

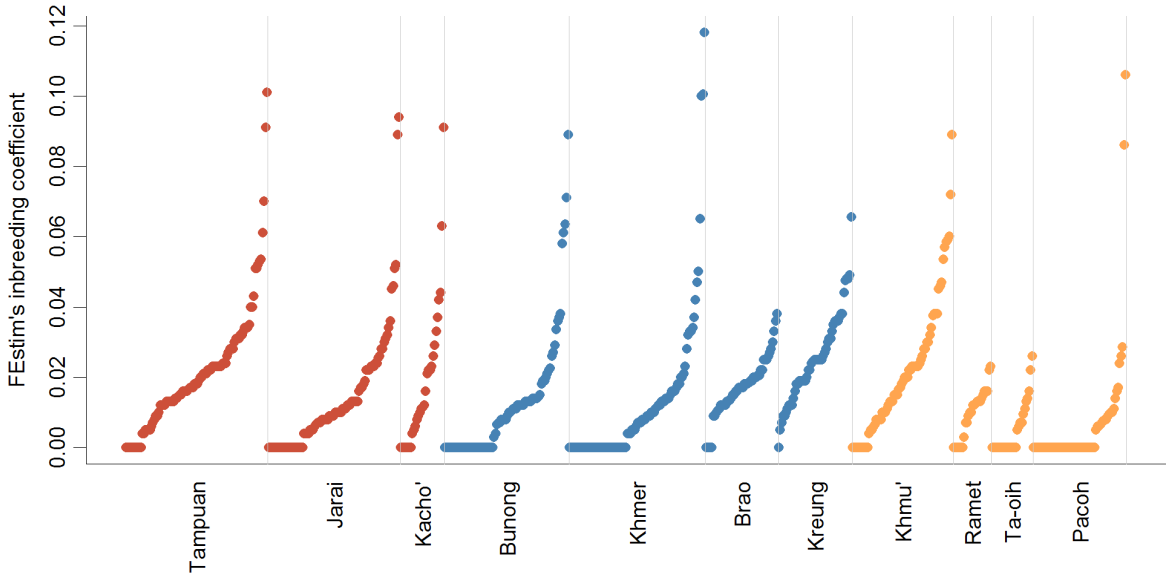


Figure 2 Individual inbreeding coefficient from FEstim. Individuals are sorted in accending order of inbreeding coefficient in each population.

Figure 3: Relationship between village endogamy rate and inbreeding coefficient at the population level. M populations are in red, C populations in blue and P populations in yellow.

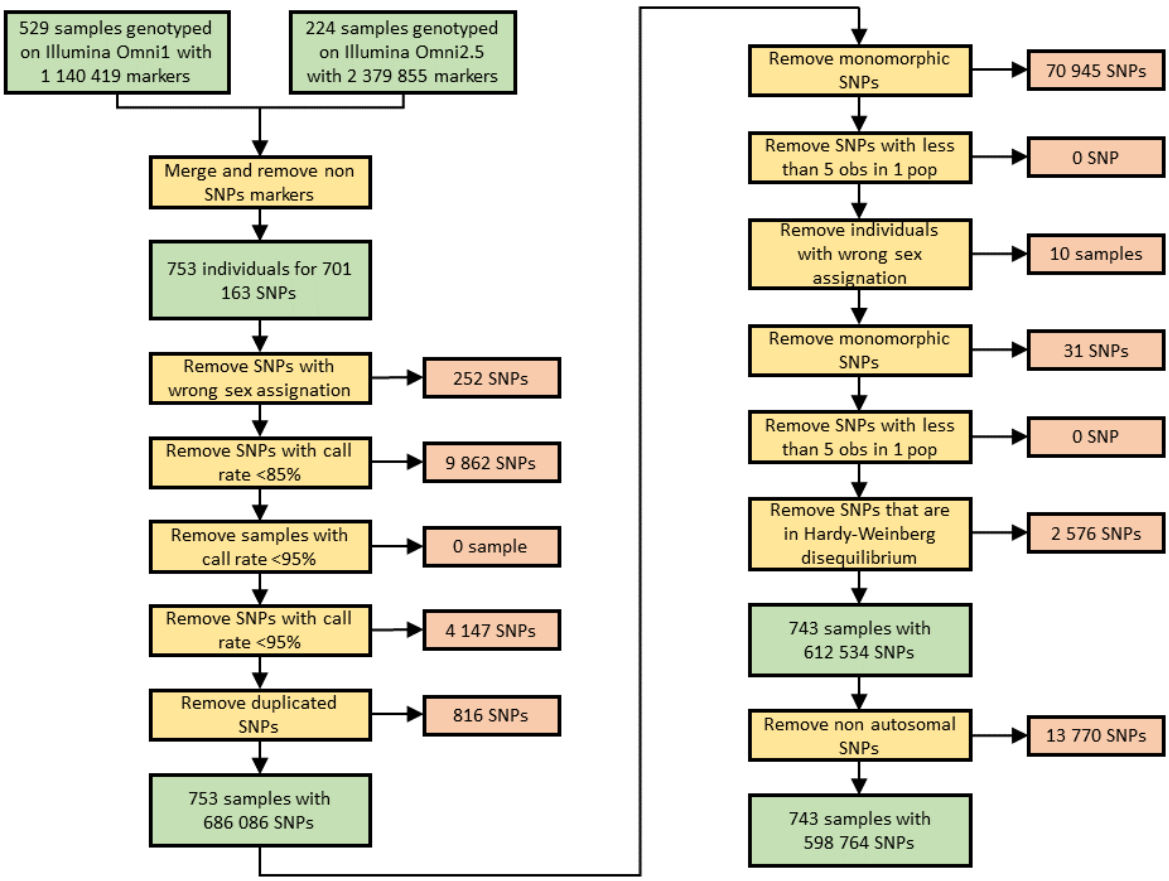
Table 1 Description of the studied populations with sampling information.

Population	Descent rule	Residence rule	Abbreviation (Group)	sampled villages	mtDNA samples	Y chromosome samples	ethno-demographic interviews
Tampuan	Matrilineal[1]	Matrilocal[4]	M	8	121	61	65
Jarai	Matrilineal	Matrilocal		6	85	56	56
Prai	Matrilineal	Matrilocal		4	73	38	37
Kacho'	Matrilineal	Matrilocal		3	41	24	27
Bunong	Cognatic[2]	Matrilocal	C	5	49	45	45
Khmer	Cognatic	Matrilocal		5	57	41	44
Brao	Cognatic	Multilocal[5]		6	34	39	39
Kreung	Cognatic	Multilocal		3	34	36	36
Khmu'	Patrilineal[3]	Patrilocal[6]	P	8	43	121	65
Ramet	Patrilineal	Patrilocal		4	21	40	40
Ta-oih	Patrilineal	Patrilocal		4	21	33	34
Pacoh	Patrilineal	Patrilocal		5	40	49	44
Total				61	619	583	532

[1]matrilineal descent: descent group affiliation is transmitted to the children through the mother. [2]cognatic descent: recognition of descent from both sides of the family in the absence of any specified lines of descent. [3]patrilineal descent: descent group affiliation is transmitted to the children through the father. [4]matrilocal residence: the husband moved to his wife's natal village after marriage. [5]multilocal residence: the couple lives alternatively in the husband's and wife natal villages before settling definitively in one place. [6]patrilocal residence: the wife moves to her husband's natal village after marriage.

Social organisation	Population	Effective population size	Confidence interval
M	Tampuan	13052	12040-14064
	Jarai	13641	12592-14690
	Prai	7182	6382-7982
	Kacho’	13407	12301-14512
C	Bunong	11850	10895-12804
	Khmer	15503	14376-16630
	Brao	12023	11052-12995
	Kreung	10920	9939-11901
P	Khmu’	13868	12798-14938
	Ramet	15051	13855-16248
	Ta-oih	14050	12922-15179
	Pacoh	12141	11140-13143

Table 2 Effective population size of the studied populations



Supplementary Figure 1 Quality control pipeline.

Annexe 5 - Contrôle qualité effectué par Romain Laurent

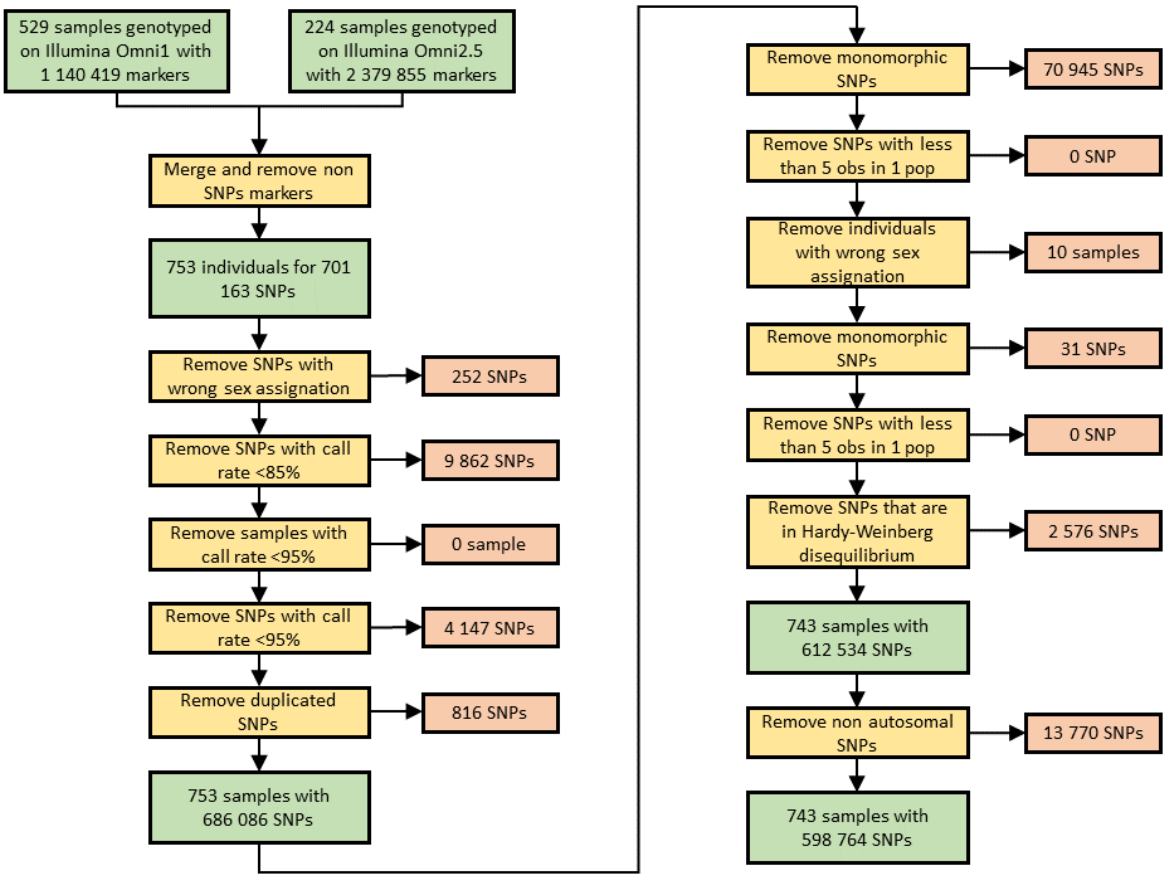


Figure S5-1 : Schéma représentant les différentes étapes du contrôle qualité effectué par Romain Laurent pour les données de génotypage.

Annexe 6 - Longueur totale des ROHs court et moyen

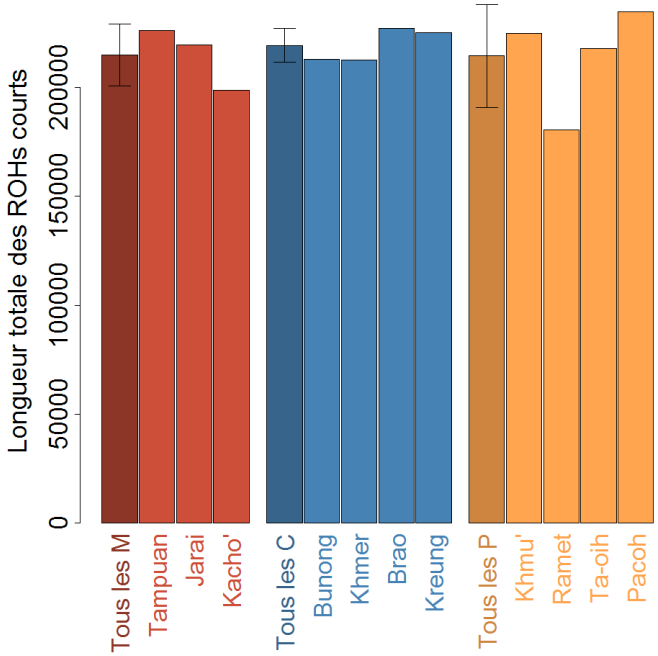


Figure S6-1 : Longueur totale individuelle des ROHs courts par population. Les populations sont regroupées par systèmes de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté.

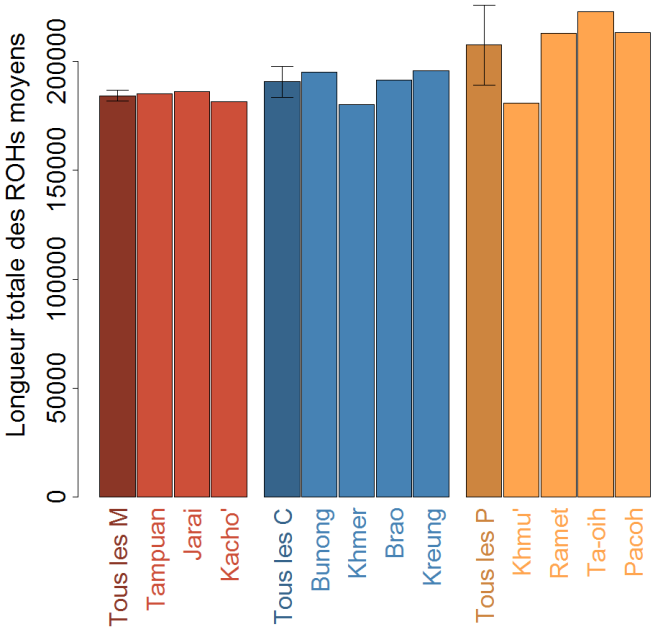


Figure S6-2 : Longueur totale individuelle des ROHs courts par population. Les populations sont regroupées par systèmes de parenté et la première barre de chaque groupe représente la valeur moyenne (avec l'écart-type) par système de parenté.

Références bibliographiques

Alard B, Ly G, Laurent R, Lafosse S, Bandine-Rolland C, Monidarin C, Diffloth G, Evrard O, Bourdier F, Pavard S, Chaix R. Social and genetic kinship in matrilineal and patrilineal human clans from Southeast Asia. prep.

Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Res 19:795‑803.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res 17:1219‑1227.

Baird IG. 2008. Various Forms of Colonialism : the Social and Spatial Reorganisation of the Brao in Southern Laos and Northeastern Cambodia.

van den Berghe PL. 1980. Incest and exogamy: A sociobiological reconsideration. Ethol Sociobiol 1:151‑162.

van den Berghe PL, Barash DP. 1977. Inclusive fitness and human family structure. Am Anthropol 79:809‑823.

Besaggio D, Fuselli S, Srikummool M, Kampuansai J, Castrì L, Tyler-Smith C, Seielstad M, Kangwanpong D, Bertorelle G. 2007. Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. BMC Evol Biol 7 Suppl 2:S12.

Billari FC, Prskawetz A, Diaz BA, Fent T. 2007. The « wedding-ring »: An agent-based marriage model based on social interaction. Demogr Res 17:59‑81.

Bittles AH. 2008. A community genetics perspective on consanguineous marriage. Community Genet 11:324‑330.

Bittles AH, Black ML. 2010. Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci 107:1779‑1786.

Bittles AH, Hamamy H. 2010. Endogamy and Consanguineous Marriage in Arab Populations. In: Teebi AS, éditeur. Genetic Disorders Among Arab Populations. Springer Berlin Heidelberg. p 85‑108.

Bourdier F. 2006. The Mountain of Precious Stones. Ratanakiri: Cambodia: Essays in Social Anthropology: Phnom Penh: Center for Khmer Studies.

Broman KW, Weber JL. 1999. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet 65:1493‑1500.

Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han L, Vives L, O'Roak BJ, Sudmant PH, Shendure J, Abney M, Ober C, Eichler EE. 2012. Estimating the human mutation rate

using autozygosity in a founder population. Nat Genet 44:1277‑81.

Cann HM, Toma C de, Cazes L, others. 2002. A Human Genome Diversity Cell Line Panel. Science (80-) 296:261‑262.

Casselberry SE, Valavanes N. 1976. « matrilocal » Greek peasants and a reconsideration of residence terminology. Am Ethnol 3:215‑226.

Chaix R, Austerlitz F, Khegay T, Jacquesson S, Hammer MF, Heyer E, Quintana-Murci L. 2004. The genetic or mythical ancestry of descent groups: lessons from the Y chromosome. Am J Hum Genet 75:1113‑1116.

Chaix R, Quintana-Murci L, Hegay T, Hammer MF, Mobasher Z, Austerlitz F, Heyer E. 2007. From social to genetic structures in central Asia. Curr Biol 17:43‑48.

Chelhod J. 1965. Le mariage avec la cousine parallèle dans le système arabe. L'homme 5:113‑173.

Conomos MP, Reiner AP, Weir BS, Thornton TA. 2016. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet 98:127‑148

Cuisenier J. 1962. Endogamie et exogamie dans le mariage arabe. L'Homme 2:80‑105.

Dessaint WY. 1981. The T'in (Mal), dry rice cultivators of Northern Thailand. J Siam Soc 69:107‑137.

Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglià A, Tofanelli S, Spedini G, Capelli C. 2004. Variation of female and male lineages in sub-Saharan populations: The importance of sociocultural factors. Mol Biol Evol 21:1673‑1682.

Destro Bisol G, Capocasa M, Anagnostou P. 2012. When gender matters: New insights into the relationships between social systems and the genetic structure of human populations. Mol Ecol 21:4917‑4920.

Diamond JM. 1975. The island dilemma: Lessons of modern biogeographic studies for the design of natural reserves. Biol Conserv 7:129‑146.

Dournes J. 1972. Coordonées, structures jörai familiales et sociales. Paris: Muséum national d'Histoire naturelle (Travaux et Mémoires de l'Institut d'Ethnologie n°77).

Ebihara M. 1977. Residence Patterns in a Khmer Peasant Village. Ann N Y Acad Sci 293:51‑68.

Evrard O. 2006. Chroniques des cendres. Paris: IRD Edition.

Feuk L, Feuk L, Carson AR, Carson AR, Scherer SW, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet 7:85‑97.

Frankham R, Driscoe DA, Ballou JD. 2002. Introduction to conservation genetics. Cambridge: Cambridge University Press.

Friedberg EC. 2003. DNA damage and repair. Nature 421:436‑40.

Gazal S, Sahbatou M, Babron M-C, Génin E, Leutenegger A-L. 2015. High level of inbreeding in final phase of 1000 Genomes Project. Sci Rep 5:17453.

Gibson J, Morton NE, Collins A. 2006. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 15:789‑95.

Gilbert JP, Hammel E a. 1966. Computer Simulation and Analysis of Problems in Kinship and Social Structure. Am Anthropol 68:71‑93.

Gluckman M. 1950. Kinship and Marriage among the Lozi of Northern Rhodesia and the Zulu of Natal. In: Radcliffe-Brown A, Forde D, éditeurs. African Systems of Kinship and Marriage. London: Oxford University Press. p 206.

Godelier M. 2004. Métamorphoses de la parenté. Paris: Fayard.

Goldstein DB, Chikhi L. 2002. Human Migrations and Population Structure: What We Know and Why it Matters. Annu Rev Genomics Hum Genet 3:129‑152.

Guillot EG, Cox MP. 2014. SMARTPOP: inferring the impact of social dynamics on genetic diversity through high speed simulations. BMC Bioinformatics 15:175.

Guillot EG, Hazelton ML, Karafet TM, Lansing JS, Sudoyo H, Cox MP. 2015. Relaxed Observance of Traditional Marriage Rules Allows Social Connectivity without Loss of Genetic Diversity. Mol Biol Evol 32:2254‑2262.

Gunnarsdóttir ED, Nandineni MR, Li M, Myles S, Gil D, Pakendorf B, Stoneking M. 2011. Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. Nat Commun 2:228.

Ben Halim N, Nagara M, Regnault B, Hsouna S, Lasram K, Kefi R, Azaiez H, Khemira L, Saidane R, Ammar S Ben, Besbes G, Weil D, Petit C, Abdelhak S, Romdhane L. 2015. Estimation of Recent and Ancient Inbreeding in a Small Endogamous Tunisian Community Through Genomic Runs of Homozygosity. Ann Hum Genet 79:402‑417.

Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. Proc Natl Acad Sci U S A 102:7476‑7480.

Hammer MF, Mendez FL, Cox MP, Woerner AE, Wall JD. 2008. Sex-biased evolutionary forces shape genomic patterns of human diversity. PLoS Genet 4:e1000202.

Henry J-P, Gouyon P-H. 2003. Précis de gnéétiques des populations. Paris: Dunod.

Heyer E, Chaix R, Pavard S, Austerlitz F. 2012. Sex-specific demographic behaviours that shape human genomic variation. Mol Ecol 21:597‑612.

International Human Genome Sequencing C. 2001. Initial sequencing and analysis of the human genome. Nature 409:860‑921.

Izikowitz KG. 1951. Lamet hill peasants in French Indochina. Göteborg: Etnografiska Museet.

Jacquard A. 1968. Panmixie et consanguinité. Quelques précisions de langage. Popul (french Ed 23:1065‑1090.

Jacquard A. 1969. Systèmes de mariages et structures génotypiques. Popul (french Ed 24:919‑934.

Johnston FE, Kensigner KM, Jantz RL, Walker GF. 1969. The population structure of the Peruvian Cashinahua: Demographic, genetic and cultural interrelationships. Hum Biol 41:29‑41.

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. Am J Hum Genet 72:281‑302.

Kimura M, Crow JF. 1963. The measurement of effective population number. Evolution (N Y) 17:279‑288.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. 2010. Genomic runs of homozygosity record population history and consanguinity. PLoS One 5:e13996.

Kloos P. 1963. Matrilocal residence and local endogamy: Environmental knowledge or leadership. Am Anthropol 65:854‑862.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WSW, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K. 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471‑5.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. 2002. A high-resolution recombination map of the human genome. Nat Genet 31:241‑247.

Kopelman NM, Stone L, Wang C, Gefel D, Feldman MW, Hillel J, Rosenberg N a. 2009. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. BMC Genet 10:80.

Kramer KL, Greaves RD. 2011. Postmarital residence and bilateral kin associations among hunter-gatherers: Pumé foragers living in the best of both worlds. Hum Nat 22:41‑63.

Kumar V, Langstieh BT, Madhavi K V., Naidu VM, Singh HP, Biswas S, Thangaraj K, Singh L, Reddy BM. 2006. Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. PLoS Genet 2:420‑424.

Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D, Ghirotto S, Brunelli A, Stoneking M. 2017. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai???Kadai languages. Hum Genet 136:85‑98.

Lachance J, Tishkoff SA. 2013. SNP ascertainment bias in poopulation genetic analyses: Whi it is important, and how to correct it. BioEssays 35:780‑786.

Laland KN, Odling-Smee J, Myles S. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. Nat Rev Genet 11:137‑48.

LeBar FM, Hickey GC, Musgrave JK. 1964. Ethnic groups of mainland Southeast Asia. Human Relations Area Files.

Ledgerwood JL. 1995. Khmer Kinship - the Matriliny Matriarchy Myth. J Anthropol Res 51:247‑261.

Leutenegger A-L, Prum B, Génin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson E a. 2003. Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73:516‑23.

Leutenegger A-L, Sahbatou M, Gazal S, Cann H, Génin E. 2011. Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? Eur J Hum Genet 19:583‑7.

Lévi-Strauss C. 1983. Le regard éloigné. Paris: Plon.

Levinson D, Malone MJ. 1980. Toward explaining human culture : a critical review of the findings of worldwide cross-cultural research. New Haven: Human Relations Area Files.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Supplementary data. Science 319:1100‑4.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5:13.

Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, Morozova I, Denisova G, Miścicka-Sliwka D. 2004. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. Hum Biol 76:877‑900.

Marlowe FW. 2000. Paternal investment and the human mating system. Behav Processes 51:45‑61.

Marlowe FW. 2004. Marital residence among foragers. Curr Anthropol 45:277‑284.

Martel G. 1975. Lovea. Village des environs d'Angkor. Aspects démographiques, économiques et sociologiques du monde rural cambodgien dans la province de Siem-Réap. Paris: Ecole française d'Extrême-Orient.

Mathieu N-C. 2007. Une maison sans fille est une maison morte. Paris: Maison des sciences de l'homme.

Matras-Troubetzkoy J. 1983. Un Village En Foret. L'essartage Chez Les Brou Du Cambodge. (Selaf, éditeur.). Paris: Peeters.

McDougall I, Brown FH, Fleagle JG. 2008. Sapropels and the age of hominins Omo I and II, Kibish, Ethiopia. J Hum Evol 55:409‑420.

McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin C, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, MacLeod A, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF. 2008. Runs of homozygosity in European populations. Am J Hum Genet 83:359‑372.

Montinaro F, Davies J, Capelli C. 2016. Group membership, geography and shared ancestry: Genetic variation in the Basotho of Lesotho. Am J Phys Anthropol 160:156‑161.

Murdock GP. 1949. Social Structure. London: Macmillan company.

Murdock GP. 1967. Ethnographic Atlas. Pittsburgh: University of Pittsburgh Press.

Murdock GP, White DR. 1969. Standard Cross-Cultural Sample. Ethnology 8:329‑369.

Nasidze I, Ling EYS, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M. 2004. Mitochondrial DNA and Y-chromosome variation in the Caucasus. Ann Hum Genet 68:205‑221.

Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. Genetics 97:145‑163.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20‑1.

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg N a, Li JZ. 2012a. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet 91:275‑92.

Pemberton TJ, Li F-Y, Hanson EK, Mehta NU, Choi S, Ballantyne J, Belmont JW, Rosenberg N a, Tyler-Smith C, Patel PI. 2012b. Impact of restricted marital practices on genetic variation in an endogamous Gujarati group. Am J Phys Anthropol 149:92‑103.

Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J. 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208‑219.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559‑75.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing.

Rasteiro R, Bouttier P -a., Sousa VC, Chikhi L. 2012. Investigating sex-biased migration during the Neolithic transition in Europe, using an explicit spatial simulation framework. Proc R Soc B Biol Sci 279:2409‑2416.

Richards AI. 1950. Some Types of Family Structure Amongst the Central Bantu. In: Radcliffe-Brown A, Forde D, éditeurs. African Systems of Kinship and Marriage. London: Oxford University Press. p 207‑251.

Rousset F. 2002. Inbreeding and relatedness coefficients: what do they measure? Heredity (Edinb):371‑380.

Salem AH, Badr FM, Gaballah M, Pääbo S. 1996. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. Am J Hum Genet 59:741‑743.

Sanchez-Faddeev H, Pijpe J, van der Hulle T, Meij HJ, van der Gaag KJ, Slagboom PE,

Westendorp RGJ, de Knijff P. 2013. The influence of clan structure on the genetic variation in a single Ghanaian village. Eur J Hum Genet 21:1134‑1139.

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. 2005. Calibrating a coalescent simulation of human genome sequence variation. Genome Res 15:1576‑1583.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat Genet 46:919‑925.

Schmutz J. 2013. The Ta'Oi Language and People. Mon-Khmer Stud 42:i‑xiii.

Schneider DM, Gough K. 1961. Matrilineal Kinship. Berkeley: University of California Press.

Sear R. 2008. Kin and Child Survival in Rural Malawi. Hum Nat 19:277‑293.

Ségurel L, Martínez-Cruz B, Quintana-Murci L, Balaresque P, Georges M, Hegay T, Aldashev A, Nasyrova F, Jobling MA, Heyer E, Vitalis R. 2008. Sex-specific genetic structure and social organization in Central Asia: Insights from a multi-locus study. PLoS Genet 4:e1000200.

Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278‑280.

Smouse PE, Vitzthum VJ, Neel J V. 1981. The impact of random and lineal fission on the genetic divergence of small human groups: A case study among the Yanomama. Genetics 98:179‑197.

Sprenger G. 2008. Do the Rmeet have clans? In: New research on Laos. . p 559‑578.

Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in disease. Annu Rev Med 61:437‑455.

Steven G, Mourad S, Marie-Claude B, Emmanuelle G, Anne-Louise L. 2014. FSuite: exploiting inbreeding in dense SNP chip and exome data. Bioinformatics [Internet]:1‑2.

The 1000 Genome Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061‑73.

Timm NH. 2002. Applied Multivariate Analysis. New York: Springer-Verlag.

UNDP Cambodia. 2010a. Phnong Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village. Phnom Penh.

UNDP Cambodia. 2010b. Kreung Ethnicity Documentation of Customary Rules Indigenous People in Pu-Trou village. Phnom Penh.

Verdu P, Pemberton TJ, Laurent R, Kemp BM, Gonzalez-Oliver A, Gorodezky C, Hughes CE, Shattuck MR, Petzelt B, Mitchell J, Harry H, William T, Worl R, Cybulski JS, Rosenberg N a, Malhi RS. 2014. Patterns of admixture and population structure in native populations of Northwest North America. PLoS Genet 10:e1004530.

Wakeley J. 2004. Metapopulation models for historical inference. Mol Ecol 13:865‑875.

Walker RS, Beckerman S, Flinn M, Gurven M, von Rueden CR, Kramer KL, Greaves RD, Córdoba L, Villar D, Hagen EH, Koster JM, Sugiyama L, Hunter TE, Hill K. 2013. Living

with Kin in Lowland Horticultural Societies. Curr Anthropol 54:96‑103.

Watkins JC. 2004. The role of marriage rules in the structure of genetic relatedness. Theor Popul Biol 66:13‑24.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution (N Y) 38:1358‑1370.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36:1122‑1125.

Wilkins JF, Marlowe FW. 2006. Sex-biased migration in humans: what should we expect from genetic data? Bioessays 28:290‑300.

Yang X, Al-Bustan S, Feng Q, Guo W, Ma Z, Marafie M, Jacob S, Al-Mulla F, Xu S. 2014. The influence of admixture and consanguinity on population genetic diversity in Middle East. J Hum Genet 59:615‑622.

Zhang X, Qi X, Yang Z, Serey B, Sovannary T, Bunnath L, Seang Aun H, Samnom H, Zhang H, Lin Q, van Oven M, Shi H, Su B. 2013. Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. Nat Commun 4:2599.