



HAL
open science

Inférence des processus démographiques passés à partir de différents marqueurs génétiques pour des populations humaines aux modes de vie contrastés.

Carla Aimé

► **To cite this version:**

Carla Aimé. Inférence des processus démographiques passés à partir de différents marqueurs génétiques pour des populations humaines aux modes de vie contrastés.. Génétique des populations [q-bio.PE]. Université Pierre et Marie Curie, 2013. Français. NNT: . tel-03114172

HAL Id: tel-03114172

<https://hal.parisnanterre.fr/tel-03114172v1>

Submitted on 18 Jan 2021

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.

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ PIERRE ET MARIE CURIE

ÉCOLE DOCTORALE : Diversité du vivant (ED 392)

Par Carla AIME

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : Anthropologie génétique

**Inférence des processus démographiques passés à partir de
différents marqueurs génétiques pour des populations humaines
aux modes de vie contrastés.**

Directeur de recherche : Frédéric Austerlitz

Soutenance le : 17 septembre 2013
Devant la commission d'examen formée de :

M. Michaël BLUM, Chargé de recherche (CNRS), rapporteur.

M. Viktor CERNY, Professeur (Université Charles de Prague), rapporteur.

M. Dominique HIGUET, Professeur (Université Pierre et Marie Curie), examinateur.

M. Etienne KLEIN, Directeur de recherche (INRA), examinateur.

M. Frédéric AUSTERLITZ, Directeur de recherche (CNRS), examinateur.

Mme Évelyne HEYER, Professeure (Muséum National d'Histoire Naturelle), examinateur.

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse Frédéric Austerlitz, pour son encadrement, son soutien et ses précieux conseils.

Je remercie chaleureusement toutes les personnes avec qui j'ai eu la chance de collaborer durant ce travail. Je pense notamment à Guillaume Laval, Etienne Patin, Lluís Quintana-Murci, Laure Ségurel, Paul Verdu, Raphaëlle Chaix, Begoña Martínez-Cruz.

Je remercie également Evelyne Heyer, Guillaume Achaz, Guillaume Laval et Dominique Higuët pour avoir accepté de prendre part chaque année à mon comité de thèse, et pour leur aide et leurs conseils.

Merci aussi à Michaël Fontaine et Friso Palstra, pour leur aide régulière et d'intéressantes discussions scientifiques, ainsi qu'à Julio Pedraza pour sa disponibilité et son aide quant à l'utilisation du cluster de calcul du Museum national d'Histoire naturelle.

Je tiens à remercier Evelyne Heyer, Etienne Klein et Dominique Higuët d'avoir accepté d'être membres de mon jury de thèse, et tout particulièrement Michaël Blum et Viktor Cerny d'être les rapporteurs de mon travail.

J'ai effectué mon travail de thèse au sein de l'UMR « Eco-anthropologie et Ethnobiologie », rattachée au département « Hommes, Natures, Sociétés » (HNS) du Museum national d'Histoire naturelle (MNHN). Je remercie chaleureusement Evelyne Heyer et Serge Bahuchet de m'y avoir accueillie durant ces trois belles années.

Je remercie l'ensemble de l'équipe de génétique des populations humaines, dont chaque membre a largement contribué à faire de mon doctorat une expérience particulièrement enrichissante à la fois scientifiquement et personnellement.

J'ai une pensée également pour d'autres membres du département, notamment Farida Selmet, Taoues Lahrem, Florence Loiseau, Manon Galland, ...

Je remercie également Alexei Drummond et son équipe pour m'avoir accueillie durant deux mois au sein du « Computational Evolution Group » à l'Université d'Auckland (Nouvelle-Zélande), dans le cadre du projet Européen NEFREX (resp. P. Endicott).

Enfin, un immense Merci à ma famille, à Cédric et à mes amis en Corse, à Paris et ailleurs pour votre soutien infailible et vos précieux encouragements ("fà pè u megliu !").

Table des matières

Introduction générale.....	7
Chapitre I : Reconstitution de l’Histoire démographique de populations aux modes de vie contrastés – Analyse de données réelles.....	17
I.A. Introduction	19
I.A.i Régions d’étude.	20
I.A.ii Différents marqueurs étudiés et questions spécifiques.....	22
I.B. Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations, predating the emergence of farming.....	27
I.B.i. Abstract.	27
I.B.ii. Introduction.....	28
I.B.iii. Results.....	30
I.B.iv. Discussion.	41
I.B.v. Material and methods.....	48
I.B.vi. Supplementary Figures.	54
I.B.vii. Supplementary Tables.....	58
I.C. Microsatellite data show recent demographic expansions in sedentary but not in nomadic human populations in Africa and Eurasia.....	89
I.C.i. Abstract.	89
I.C.ii. Introduction.....	90
I.C.iii. Material and Methods.	92
I.C.iv. Results.....	95
I.C.v. Discussion.	97
I.C.vi. Conclusion and further work.	100
I.C.vii. Supplementary Tables.....	101
I.D. Y-chromosome microsatellite markers reveal that sex-specific migration patterns impact coalescent-based demographic inferences in Central Asia.....	111
I.D.i. Abstract.....	111
I.D.ii. Introduction.	112

I.D.iii. Material and Methods.....	114
I.D.iv. Results.....	117
I.D.v. Discussion.....	121
I.D.vi. Supplementary Figures.....	126
I.D.vii. Supplementary Tables.....	129
I.E. Conclusions de l'analyse des données réelles.....	141
Chapitre II : Vers une meilleure interprétation des inférences démographiques à partir des données génétiques - Analyses préliminaires sur données simulées.....	145
II.A. Introduction.....	147
II.B. Test des méthodes à partir de séquences simulées.....	149
II.B.i. Matériel et méthodes.....	149
II.B.ii. Résultats.....	150
II.B.iii. Supplementary Tables.....	163
II.C. Inférences démographiques à partir de différents types de marqueurs simulés sous différents scénarios d'expansions.....	171
II.C.i. Matériel et méthodes.....	171
II.C.ii. Résultats.....	173
II.C.iii. Tableau supplémentaire.....	181
II.D. Discussion et conclusions des analyses sur données simulées.....	183
Conclusion générale.....	187
Bibliographie.....	191

Introduction générale

Comme celles de beaucoup d'espèces, l'Histoire d'*Homo sapiens* est rythmée par des processus démographiques complexes incluant expansions, déclin, migrations, *etc.* (Cavalli-Sforza *et al.*, 1994). L'étude de ces processus fait historiquement appel à des disciplines variées telles que l'archéologie et la paléanthropologie. Depuis ces dernières décennies, la génétique des populations, qui étudie le rôle relatif des forces évolutives qui façonnent notre diversité génétique, n'est pas en reste et fournit de nouveaux outils et de nouvelles perspectives dans ce domaine. Les processus démographiques (fluctuations d'effectifs, intensité des migrations, *etc.*) influencent fortement et durablement la structure génétique des populations, notamment via la dérive génétique. La dérive, qui se définit comme la variation stochastique des fréquences alléliques d'une génération à l'autre dans une population de taille finie, est en effet d'autant plus intense que la taille efficace de la population est faible (Wright, 1931). Des traces des expansions et déclin passés peuvent donc subsister, avec néanmoins une certaine perte d'information, dans la structure du patrimoine génétique des populations actuelles. Ainsi, bien que l'ensemble des êtres humains vivants actuellement possèdent la même information génétique pour plus de 99% de leur génome (Li and Sadler, 1991), étudier la variation néanmoins existante au sein et entre les populations permet de mieux appréhender leur histoire démographique (Menozzi *et al.*, 1978 ; Cavalli-Sforza *et al.*, 1994 ; Beaumont, 1999).

Du point de vue méthodologique, la nature (séquences, microsatellites, puces à ADN) et la quantité de données génétiques disponibles ont rapidement évolué au cours de ces dix dernières années, notamment grâce aux nouvelles techniques de génotypage (Mardis, 2008). Parallèlement, de nouvelles méthodes d'analyses de ces données se sont développées, largement favorisées par un équipement informatique permettant d'exploiter de plus en plus rapidement des jeux de données toujours plus conséquents. Certaines difficultés subsistent cependant, en termes de puissance d'analyse ou encore d'interprétation des résultats. Ainsi, les méthodes utilisées pour inférer les processus démographiques ont dû beaucoup évoluer, et évoluent encore, afin de s'adapter au mieux aux difficultés de toutes sortes. Des processus différents (par exemple une expansion démographique et un évènement de sélection) aboutissent parfois au même signal. Pour y remédier, il faut étudier simultanément plusieurs régions du génome. En effet, les processus démographiques affectent tous les locus de la même façon, contrairement à la sélection (Nielsen, 2001).

Malgré ces difficultés, la génétique des populations, en complément des données archéologiques et paléanthropologiques, a apporté au cours de ces vingt-cinq dernières années de précieuses informations sur les divergences et migrations historiques entre les populations, ainsi que sur les origines de notre espèce. Notamment, elle a permis de confirmer l'origine africaine d'*Homo sapiens* (Cann *et al.*, 1987) et de dater la colonisation hors d'Afrique (« out-of-Africa ») entre 70 000 et 50 000 ans avant notre ère (Cavalli-Sforza and Feldman, 2003 ; Laval *et al.*, 2010). De nombreuses études, que nous détaillerons davantage par la suite, se sont également intéressées à l'histoire de la colonisation de l'Eurasie (*e.g.* Quintana-Murci *et al.*, 1999; Laval *et al.*, 2010), ou encore aux variations d'effectif efficace au sein des populations au cours de leur histoire , incluant notamment les grandes expansions démographiques, sujet qui nous intéressera particulièrement ici (*e.g.* Harpending *et al.* 1993 ; Watson *et al.*, 1997 ; Pritchard *et al.*, 1999 ; Voight *et al.*, 2005; Chaix *et al.*, 2008; Atkinson *et al.*, 2009; Batini *et al.*, 2011).

Reconstituer la démographie passée des populations à partir du polymorphisme actuel : Etat des lieux des outils et méthodes.

Plusieurs types de méthodes ont été proposés pour inférer des processus démographiques passés à partir de données génétiques contemporaines. Elles peuvent être regroupées en trois principales catégories : celles utilisant la simulation d'arbres de coalescence grâce à des algorithmes de type MCMC (« Markov Chain Monte Carlo », Metropolis *et al.*, 1953), les méthodes dites ABC (« Approximate Bayesian Computations » ; Beaumont, 2002), et celles fondées sur le calcul du déséquilibre de liaison (*e.g.* Austerlitz *et al.*, 2003). Nous développerons ici plus particulièrement les méthodes basées sur la simulation d'arbres de coalescence par MCMC, que nous avons utilisées au cours de ce travail.

Ces méthodes trouvent leur origine dans un des corpus théoriques clés de la génétique des populations : la théorie de la coalescence. Proposée par Kingman (1982), et qualifiée par Fu et Li (1999) d'« avancée la plus significative des deux dernières décennies du 20^{ème} siècle en génétique des populations », elle permet de faire le lien entre la structure actuelle du patrimoine génétique et les paramètres démographiques. Deux gènes observés dans une population donnée peuvent être des copies d'un même gène à la génération précédente (ce que nous appellerons par la suite « évènement de coalescence »), avec une probabilité égale à l'inverse du nombre de gènes présent dans la population. Ainsi, les lois de la probabilité permettent de reconstruire par une approche ascendante (c'est-à-dire en remontant dans le temps), des « arbres de coalescence », sorte d'« arbres généalogiques » ayant pour unité non

pas les individus mais les gènes, et dans lesquels les mutations représentent la source de variation. À chaque génération, la probabilité d'un évènement de coalescence dépend de l'effectif efficace de la population : plus la population est petite, plus cette probabilité est importante. Ainsi, la forme de l'arbre de coalescence est fortement influencée par les variations d'effectif efficace au cours de l'Histoire de la population (**Figure 1**).

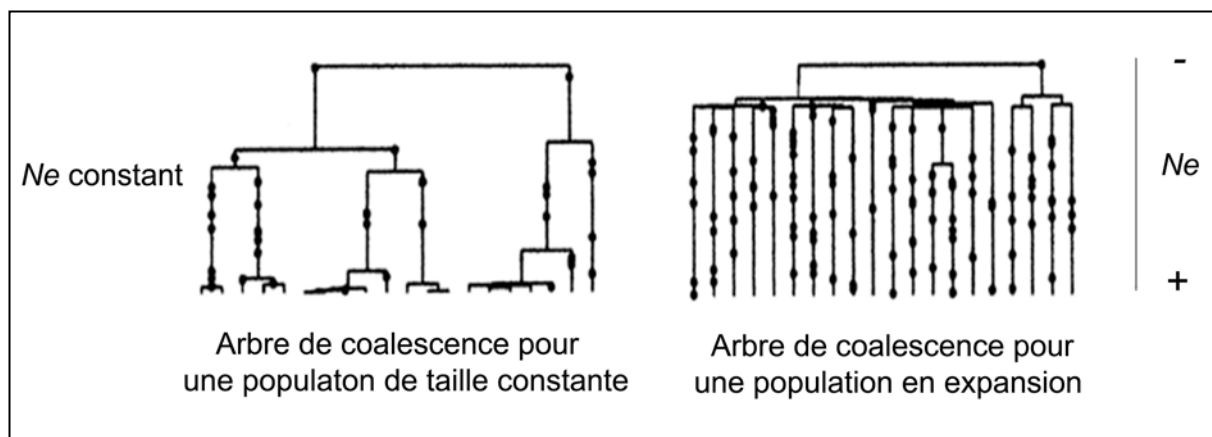


Figure 1. Exemples d'arbres de coalescence pour une population de taille constante versus une population en expansion.

Cependant, il est généralement impossible de reconstruire avec certitude le véritable arbre de coalescence sous-jacent aux données observées, ou de considérer tous les arbres possibles, bien trop nombreux. Des méthodes numériques ont donc été développées pour faire face à ce problème. Plusieurs logiciels (*e.g.* GENETREE, Griffiths et Tavaré, 1994 ; MSVAR, Beaumont, 1999 ; LAMARC, Kuhner, 2006 ; BEAST, Drummond et Rambaut, 2007) utilisent les algorithmes de type MCMC (Markov Chain Monte Carlo, Metropolis *et al.*, 1953) pour simuler des arbres de coalescence possibles à partir de données de polymorphisme, et estimer des paramètres d'intérêt à l'aide de leur vraisemblance (*i.e.* probabilité d'observer les données en fonction des paramètres). Excoffier et Heckel (2006) et Kuhner (2008) présentent plus en détails et comparent ces différents outils. Le principe général est le suivant : les arbres simulés sont modifiés pas par pas, chaque modification étant conservée ou annulée en fonction de son effet positif ou négatif sur la vraisemblance de l'arbre par rapport aux données. On obtient ainsi un échantillon des arbres les plus vraisemblables parmi l'espace des arbres parcourus par l'algorithme. La densité de probabilité des paramètres démographiques, tels que les effectifs efficaces actuel et ancestral de la population ou encore son taux de croissance, est ensuite estimée à partir de la topologie de ces arbres « les plus vraisemblables ».

Ces méthodes impliquent généralement le choix d'un modèle démographique *a priori* (croissance exponentielle, croissance logistique, etc.). Cependant, différents outils permettent ensuite de comparer différents modèles et de sélectionner celui qui s'ajuste au plus près des données observées. On pourra notamment utiliser le DIC (Deviance Information Criteria, Spiegelhalter *et al.*, 2010), qui est une mesure de l'écart du modèle aux données tenant compte du nombre de paramètres. Une autre solution est d'estimer la vraisemblance marginale des différents modèles, c'est-à-dire la probabilité des données observées sachant le modèle considéré (Gelfand et Dey, 1994). Certains auteurs présenteront plutôt le ratio des vraisemblances marginales des modèles comparés, alors appelé « Bayes factor » (Sinsheimer *et al.* 1996; Suchard *et al.*, 2001). Si les méthodes originales d'estimation de la vraisemblance marginale et notamment l'utilisation de la moyenne harmonique (HME) ont longtemps été sujettes à débat et tendent souvent à une surestimation (*e.g.* Xie *et al.*, 2011), d'autres méthodes plus rigoureuses et plus consensuelles peuvent être utilisées, telles que les approches de type « Path-sampling » (Gelman and Meng, 1998) et « Stepping-stone sampling » (Xie *et al.*, 2011). Ces méthodes sont décrites en détail et comparées par Baele *et al.* (2012).

Enfin, des méthodes non-paramétriques existent pour inférer les variations passées de taille efficace dans une population sans fixer de modèle démographique *a priori*, en utilisant l'intervalle de temps entre les événements de coalescence. Ce sont les approches de type « Bayesian Skyline plots », qui permettent de visualiser graphiquement l'évolution de l'effectif efficace au cours du temps (**Figure 2**). Les différentes variantes de cette approche sont décrites et comparées par Ho et Shapiro (2011). On retiendra particulièrement la plus récente, appelée « Extended Bayesian Skyline Plots » (EBSP, Heled and Drummond, 2010). Elle présente en effet deux avantages notables : la possibilité d'analyser conjointement plusieurs locus, et la non-nécessité de fixer *a priori* le nombre de changements démographiques connus par la population.

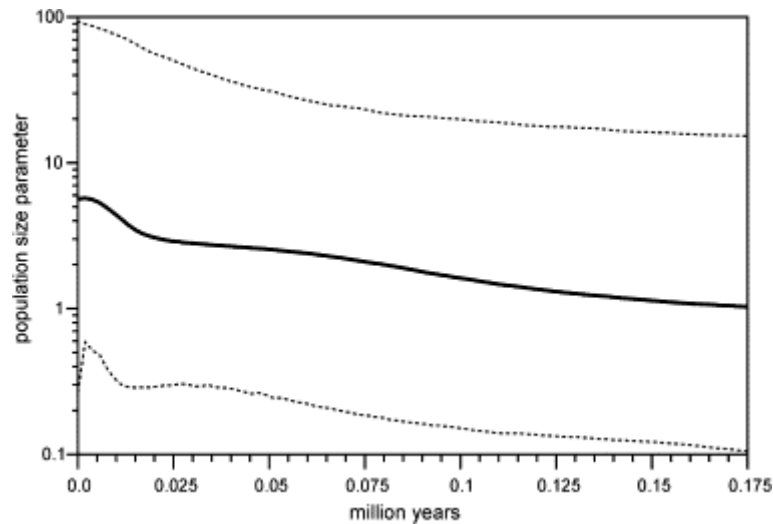


Figure 2 : Un exemple de « Bayesian Skyline plot » pour une population en expansion (Rebernik *et al.*, 2010). Le temps est représenté en abscisse, du présent à gauche au plus loin vers le passé à droite. L'effectif efficace de la population apparaît en ordonnée. Ici, la courbe noire représente la médiane, alors que les courbes bleues représentent l'intervalle HPD (« Highest probability density interval »).

Relations entre mode de vie, démographie et diversité génétique : Le cas de la transition Néolithique.

La relation entre démographie passée et diversité génétique actuelle n'étant donc plus à démontrer, nous nous intéresserons à présent aux liens entre modes de vie et démographie dans l'espèce humaine, et à ce que l'étude de la diversité génétique peut nous apprendre sur cette question.

Notre espèce a connu, au cours de son histoire, des périodes de transitions culturelles et technologiques majeures. L'une d'elle attire depuis longtemps une attention toute particulière chez les archéologues et les paléanthropologues, mais aussi les généticiens des populations : la « transition Néolithique ». Débutée entre 11 000 et 3 000 ans avant notre ère selon les régions du monde, elle est marquée par le début de la domestication des plantes et des animaux, entraînant de nombreuses populations humaines à passer d'un mode de subsistance basé sur la chasse et la cueillette à l'avènement de l'agriculture et de l'élevage (*e.g.* Bocquet-Appel and Bar-Yosef, 2008).

En ce qui concerne l'agriculture, les plus anciens vestiges archéologiques (poteries, systèmes d'irrigation, *etc.*) ont été découverts au Proche-Orient et datent d'au moins 11 000 ans (*e.g.* Bocquet-Appel et Bar-Yosef, 2008). Cependant, plusieurs foyers d'apparition

indépendants sont à dénombrer, notamment en Chine, en Nouvelle-Guinée, et plus tardivement en Amérique centrale et en Afrique de l'Ouest (**Figure 3**).

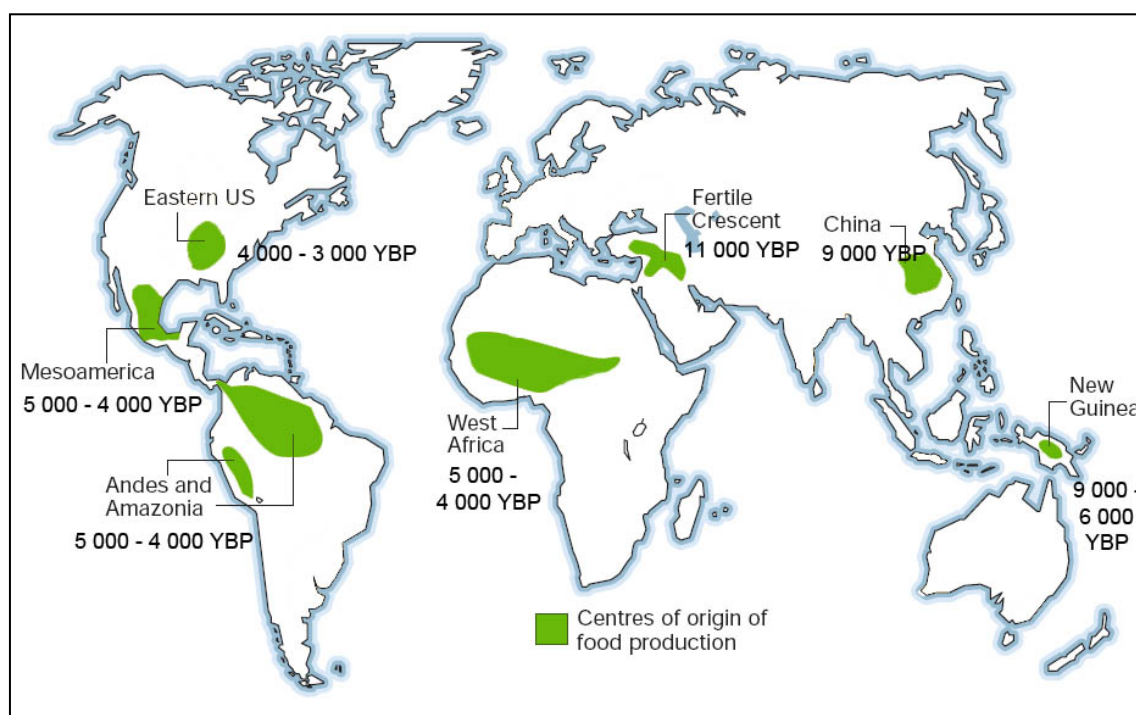


Figure 3. Les foyers d'apparitions de l'agriculture au Néolithique (d'après Diamond, 2002 et Bocquet-Appel et Bar-Yosef, 2008). YBP= « Years Before Present ».

L'agriculture s'est ensuite répandue de proche en proche dans le monde entier, soit de manière démique (migration d'individus) soit par diffusion culturelle (diffusion des savoirs et des techniques d'une région à l'autre sans déplacement de populations). En Europe par exemple, des études génétiques ont montré une origine proche-orientale des premières populations agricoles, suggérant que des flux démiques depuis cette région auraient entraîné la diffusion vers l'ouest de l'agriculture il y a environ 9 000 ans (Sokal *et al.*, 1991). Dans tous les cas, l'apparition de l'agriculture ne représente pas seulement un nouveau mode d'alimentation, mais implique également de profonds bouleversements sur le plan du mode de vie, des techniques, de l'organisation sociale, etc. Elle s'accompagne notamment du début de la sédentarisation des populations (Mazoyer and Roudart, 2006).

D'autres populations ont conservé leur mobilité, soit en restant chasseurs-cueilleurs, soit en développant un pastoralisme dit « semi-nomade ». Aujourd'hui encore, des populations ont conservé un mode de vie nomade basé sur la chasse et la cueillette. C'est le cas notamment de certaines populations d'Afrique centrale, couramment appelées « Pygmées », ou encore d'Asie du Sud-Est (les « Négritos »). En ce qui concerne l'élevage semi-nomade,

son origine fait encore largement débat (Cavalli Sforza, 1996). Ce mode de vie semble s'être particulièrement développé dans les régions d'Eurasie les plus arides (déserts, steppes) ou montagneuses, probablement moins propices à l'agriculture (Asie centrale, Altaï, Mongolie, etc.). Néanmoins, la question de son apparition directement à partir d'un mode vie basé sur la chasse, ou secondairement à partir d'un agropastoralisme sédentaire n'est pas encore tranchée (Cavalli Sforza, 1996 ; Renfrew, 1996). Il est possible également que les modalités d'apparitions de l'élevage varient en fonction des régions.

D'après les archéologues et les paléanthropologues, la transition Néolithique se serait accompagnée du début des grandes expansions démographiques humaines, notamment dans les populations sédentarisées (Binford et Chasko, 1976 ; Mazoyer et Roudart, 2006; Bocquet-Appel, 2011). En effet, une augmentation importante des vestiges archéologiques dans les différents foyers d'apparition de l'agriculture met en lumière une importante croissance démographique. Les données paléanthropologiques des nécropoles en Eurasie confirment cette observation. Notamment, Bocquet-Appel (2011) montre une augmentation nette de la proportion des individus immatures (5 à 19 ans) parmi les squelettes correspondant à la période de la transition Néolithique. Cette proportion reflète le degré d'ouverture de la pyramide des âges de la population vivante, causée par la natalité et la fécondité. La croissance Néolithique correspond donc à une augmentation de la fertilité, plutôt qu'à une diminution de la mortalité. A l'inverse, les populations de chasseur-cueilleurs nomades auraient gardé un effectif constant, dû aux contraintes liées à leur mode de vie. En particulier, il est connu que l'intervalle moyen entre les naissances est plus long chez les populations de chasseur-cueilleurs nomades que chez les populations d'agriculteurs sédentaires (Short, 1982 ; Diamond and Bellwood, 2003). Diverses hypothèses ont été proposées pour expliquer ce phénomène, telles que la difficulté pour les mères de se déplacer de campement en campement avec plus d'un enfant en bas âge (Blurton Jones, 1986), ou encore des théories liées à l'alimentation. Notamment, la consommation d'aliments riches en hydrates de carbones, tel que les céréales, favoriserait une récupération rapide des femmes après un accouchement (Valeggia and Ellison, 2009). En ce qui concerne les éleveurs nomades, l'impact de ce mode de vie sur la démographie des populations n'a, à notre connaissance, pas été spécifiquement étudié.

Du côté des généticiens, le point de vue est un peu différent. En effet, peu d'études de génétique des populations ont permis de détecter des expansions démographiques correspondant à la période de la transition Néolithique. A l'inverse, de nombreux auteurs

montrent des signaux d'expansions plus anciennes, apparues dès le Paléolithique moyen ou supérieur, dans de nombreuses populations d'Eurasie et d'Afrique. En Afrique, les études basées sur l'analyse des séquences ADN autosomales ou mitochondriales indiquent un début d'expansion entre 80 000 et 25 000 ans avant notre ère (*e.g.* Harpending *et al.*, 1993; Watson *et al.*, 1997; Voight *et al.*, 2005; Laval *et al.*, 2010; Batini *et al.*, 2011). Pritchard *et al.* (1999) fournit une estimation plus récente, entre 15 000 et 12 000 ans avant notre ère, en analysant des marqueurs microsatellites portés par le chromosome Y. Enfin, en analysant séparément différents haplogroupes de l'ADN mitochondrial, Atkinson *et al.* (2009) montrent plusieurs vagues d'expansions successives durant la période paléolithique en Afrique, la plus récente datant de 8 000 ans avant notre ère. Dans tous les cas, les expansions inférées par ces auteurs sont donc antérieures à l'apparition de l'agriculture dans la région, datée d'il y a environ 4 000 à 5 000 ans (Bocquet-Appel et Bar-Yosef, 2008). En ce qui concerne l'Eurasie, Chaix *et al.* (2008) infèrent des débuts d'expansions allant de 63 000 à 17 000 ans avant notre ère à partir de la région contrôle HVS-I de l'ADN mitochondrial, et de 40 000 à 11 000 ans avant notre ère à partir de marqueurs microsatellites du chromosome Y. Ils observent une vague d'expansion d'est en ouest, avec des débuts d'expansions plus anciens en Asie de l'Est et de plus en plus récents en allant vers l'Europe. Harpending *et al.* (1993) inféraient également une croissance antérieure à l'agriculture en Europe à partir de l'ADN mitochondrial. A notre connaissance, seule une étude récente basée sur l'analyse de séquences mitochondriales complètes a montré un épisode d'expansion plus récent chez les premiers agriculteurs d'Europe (Fu *et al.*, 2012).

Le fait que les études génétiques soient aussi rares à démontrer une croissance Néolithique pourrait être dû à des difficultés méthodologiques. Notamment, il est connu que les événements anciens d'expansion peuvent masquer le signal des événements plus récents (Lavery, 1996), dû au fait que la majorité des événements de coalescence vont alors correspondre aux événements les plus anciens (la population étant alors plus petite). Néanmoins, l'ensemble des études précédentes nous fournit deux informations essentielles : (1) Contrairement au point de vue des paléoanthropologues, qui considèrent que le nomadisme est généralement associé à des tailles de population constantes, les généticiens des populations montrent un début des expansions démographiques humaines précédant la transition Néolithique et les premières sédentarizations de populations (2) L'étude de marqueurs génétiques différents entraîne l'inférence de dates de début d'expansion différentes. Notamment, les inférences provenant de l'analyse des microsatellites du chromosome Y sont généralement plus récentes que celles provenant de l'étude de séquences

autosomales ou mitochondriales. Cependant, il est encore difficile de savoir si cela reflète plusieurs vagues d'expansions successives ou seulement un artefact méthodologique lié aux différentes propriétés de chaque marqueur. Nous reviendrons sur cette question au cours des travaux présentés ici.

Plus étonnant encore, en Afrique, les études de génétique des populations n'ont détecté d'expansions paléolithiques que dans les populations ultérieurement sédentarisées au Néolithique. En effet, Batini *et al.* (2011) détectent une expansion paléolithique chez les populations actuellement agricoles, mais pas chez les populations restées chasseur-cueilleurs. Ces résultats posent la question de la chronologie de la transition de mode vie au Néolithique et de la transition démographique. Ils nous ramènent à une théorie proposée par Sauer (1952) pour expliquer l'apparition indépendante de l'agriculture dans plusieurs régions du monde. Selon lui, la transition Néolithique auraient été favorisée par des expansions démographiques antérieures : les populations humaines auraient commencé leur croissance démographique au Paléolithique, cette croissance auraient favorisé l'apparition de l'agriculture dans les populations où elle était la plus forte. Cette hypothèse est également défendue par Cavalli-Sforza (1996), qui suggère que les populations humaines auraient connu une croissance démographique d'un facteur cinq entre la sortie d'Afrique il y a environ 60 000 ans et le début de l'Holocène il y a 10 000 ans, juste avant l'apparition de l'agriculture.

D'après Cohen (1977), la pression démographique, alliée à des conditions climatiques favorables, aurait favorisé l'émergence de nouvelles techniques aboutissant à de nouveaux modes de subsistance, incluant la domestication des plantes et des animaux. Ce point de vue est donc apparemment supporté par les résultats des études récentes en génétique des populations. Cependant, certaines difficultés méthodologiques viennent semer le doute sur cette interprétation. Notamment, selon Ray *et al.* (2003) et Excoffier (2004), les signaux d'expansions spatiales anciennes pourraient être effacés ou atténués dans les populations présentant un fort degré d'isolation (moins de 20 immigrants par génération). Or, les populations de chasseur-cueilleurs d'Afrique centrale présentant un degré d'isolation largement supérieur à celui de leurs voisins agriculteurs (Verdu *et al.*, 2009), cela constitue une explication alternative pour l'absence de signaux d'expansion paléolithique dans ces populations. Cette question reste pour l'instant ouverte, et sera traitée au cours de cette thèse.

Objectifs et organisation de la thèse.

Le travail présenté ici se compose de deux parties :

Dans un premier temps, nous avons utilisé plusieurs méthodes basées sur la théorie de la coalescence pour analyser différents jeux de données réelles afin de mieux comprendre l'histoire démographique de populations humaines d'Afrique et d'Eurasie aux modes de vie contrastés (agriculteurs sédentaires, chasseurs-cueilleurs nomades, éleveurs semi-nomades). Nous avons ainsi pu comparer l'histoire démographique de ces populations et dater le début des expansions inférées. Nous avons analysé différents types de marqueurs génétiques aux propriétés (taux de mutation, effectif efficace, mode de transmission) différentes afin de comparer les résultats obtenus et d'apporter un éclairage nouveau aux différentes questions soulevées précédemment sur l'expansion des populations humaines. Enfin, nous avons discuté nos résultats à la lumière des informations disponibles sur les populations et régions étudiées (données historiques, anthropologiques, paléoclimatiques, etc.) afin de mieux comprendre et interpréter les résultats observés. Nous avons également analysé les degrés d'isolation et migration dans les populations étudiées afin d'identifier de potentiels facteurs confondants et biais méthodologiques.

Dans un second temps, nous avons appliqué les mêmes méthodes à l'étude de données simulées sous différents scénarii démographiques, afin de (1) tester nos méthodes et leur limites, (2) mieux comprendre les différences entre les résultats obtenus avec les différents marqueurs. Nous avons notamment comparé le pouvoir de détection d'expansions plus ou moins récentes et intenses, aussi que la capacité à détecter et distinguer deux phases d'expansions successives avec les différents types de données.

Chapitre I : Reconstitution de l'Histoire démographique de populations aux modes de vie contrastés – Analyse de données réelles.

I.A. Introduction

Comme expliqué en introduction générale, la répartition actuelle du polymorphisme génétique dans les populations humaines est influencée par son histoire démographique, et différentes méthodes basées sur la théorie de la coalescence permettent d'inférer les grands changements démographiques passés à partir de données génétiques actuelles. Dans cette première partie, nous avons analysé différents jeux de données déjà disponibles au laboratoire ou publiés dans la littérature, afin d'apporter un éclairage aux problématiques suivantes : (1) Y a-t-il une association entre modes de vie et histoire démographique dans les populations humaines ? (2) Les grandes expansions ont-elles commencé avant ou pendant la transition Néolithique et les premières sédentarisation ? (3) Différents marqueurs génétiques apportent-ils des informations similaires ou complémentaires sur le passé démographique des populations étudiées ? (4) Dans quelle mesure les différences démographiques observées entre des populations aux modes de vie différents peuvent-elles être liées à des facteurs confondants tels que l'isolation et les migrations récentes ?

Ce travail a donné lieu à la rédaction de trois articles scientifiques. Le premier, qui décrit l'analyse de séquences autosomales et mitochondriales, est en cours de considération (version révisée) par la revue *Molecular Biology and Evolution*. Le second porte sur l'analyse de données microsatellites autosomales, et a été soumis à *European Journal of Human Genetics*. Le troisième, qui sera bientôt soumis également à une revue internationale à comité de lecture, concerne l'étude de marqueurs microsatellites du chromosome Y. Les régions étudiées sont les mêmes pour les trois études : L'Afrique, l'Eurasie, et plus particulièrement l'Asie centrale. En Afrique, nous avons comparé des populations d'agriculteurs sédentaires et de chasseurs cueilleurs nomades. En Eurasie et en Asie centrale, nous avons pu comparer l'histoire démographique de populations d'agriculteurs sédentaires et d'éleveurs nomades, ce qui, à notre connaissance, n'avait encore fait l'objet d'aucune étude. Les populations étudiées, les méthodes employées ainsi que certaines questions spécifiques varient en fonction des jeux de données et des particularités de chaque marqueur, et seront détaillées ci-après.

I.A.i Régions d'étude.

En Afrique.

La majorité des populations africaines étudiées sont originaires d'Afrique centrale (Gabon, Cameroun, Nigeria, République centrafricaine et République démocratique du Congo), bien que le jeu de données utilisé pour les séquences autosomales inclue également deux populations agricoles sédentaires d'Afrique de l'Est (les Chagga de Tanzanie et les Mozambicains). L'Afrique centrale est aujourd'hui peuplée par de nombreuses populations d'agriculteurs au mode de vie sédentaire, mais également par le plus grand groupe de populations contemporaines de chasseurs-cueilleurs nomades (Hewlett, 1996, Diamond et Bellwood, 2003). Ces dernières, communément appelés pygmées en référence à leur petite stature, sont présentes sur un vaste territoire de forêt équatoriale allant du bassin du Congo au lac Victoria (Lee et Daly, 1999). Elles interagissent fortement avec leurs voisins agriculteurs, notamment par le commerce (Cavalli-Sforza, 1986). Les études de génétique des populations à partir de données contemporaines sont particulièrement intéressantes dans cette région où l'acidité des sols limite fortement la subsistance de restes humains pour les études paléanthropologiques (Oslisly, 2001). Elles ont notamment permis de dater la divergence génétique entre les ancêtres des agriculteurs et des chasseur-cueilleurs actuels il y a environ 60 000 ans, donc bien avant l'apparition de l'agriculture (Patin *et al.*, 2009 ; Verdu *et al.*, 2009). Il existe des flux génétiques limités entre les deux groupes, intervenant presque exclusivement depuis les agriculteurs vers les chasseurs-cueilleurs (Verdu *et al.*, 2009).

En Eurasie.

Des populations d'éleveurs nomades ou semi-nomades coexistent avec les populations sédentaires dans différentes parties du continent (Moyen-Orient, Pamir, Russie, Asie centrale, *etc.*). L'histoire démographique des éleveurs nomades est jusqu'alors très mal connue. Nous avons comparé ici pour chaque étude des populations sédentaires et nomades de différentes régions d'Eurasie.

Le cas particulier de l'Asie centrale.

Nous appuyant sur les nombreuses données récoltées par notre laboratoire dans cette région depuis plusieurs années, nous avons pu analyser, en fonction des marqueurs, jusqu'à 30 populations originaires du Kirghizstan, du Tadjikistan et d'Ouzbékistan. Bien que l'Asie centrale soit souvent considérée comme un carrefour des grandes vagues historiques de

migrations en Eurasie (Nei et Roychoudhury, 1993), l'histoire de son peuplement et de sa démographie passée est encore largement méconnue. On divise classiquement les populations de la région en deux groupes sur la base à la fois de leur mode de vie et des données linguistiques (Bregel, 2003). Les populations d'agriculteurs traditionnellement sédentaires parlent des langues dites indo-iraniennes. Elles coexistent avec des populations au mode de vie traditionnellement basé sur l'élevage semi-nomade, du moins jusqu'à leur sédentarisation récente (pour la plupart au 19^{ème} siècle sous le régime russe). Ces dernières populations sont de langues turco-mongoles. La différence linguistique témoigne en faveur d'une origine différente des deux groupes. Cependant, la catégorisation n'est pas toujours aussi simple. Par exemple, les sociétés turkmènes, de langue turco-mongole, sont divisées en agriculteurs et éleveurs, et les individus alternent constamment entre ces deux modes de vie (Wood, 2002). Cependant, nous classerons ici ces dernières parmi les populations d'éleveurs compte tenu de l'importance de l'élevage dans leur régime alimentaire.

Selon Martinez-Cruz *et al.* (2011), les populations d'éleveurs turco-mongoles seraient génétiquement plus proches des populations de l'Est de l'Eurasie, ce qui corrobore l'hypothèse d'un flux génique lié aux grandes invasions venues des empires mongols pendant les deux premiers siècles avant Jésus Christ (Gumilyov, 1967). Les populations agricoles indo-iraniennes seraient à l'inverse plus proches des populations européennes, et établies dans la région au moins depuis le Néolithique, comme en témoignent les données archéologiques (Brunet, 1999). Les populations étudiées ici appartiennent aux ethnies suivantes : les Kirghiz, les Ouzbeks, les Kazakhs, les Karakalpaks et les Turkmènes (Turco-mongols) et les Tadjiks (Indo-iraniens).

I.A.ii Différents marqueurs étudiés et questions spécifiques.

Les marqueurs génétiques utilisés ici (séquences autosomales neutres, séquences mitochondriales hypervariables, microsatellites autosomaux et microsatellites du chromosome Y) possèdent des propriétés différentes en termes de modèles de mutation, d'effectif efficace et de mode de transmission. Notamment, le taux de mutation des microsatellites est beaucoup plus élevé que celui des séquences, et celui des séquences mitochondriales hypervariables est plus élevé que celui des séquences autosomales. En théorie, un taux de mutation plus élevé devrait permettre d'inférer plus précisément des événements plus récents, comme cela a déjà été montré sur des données simulées dans le cadre des méthodes ABC (Cornuet *et al.*, 2008). Par ailleurs, les autosomes sont transmis à la fois par les deux parents (marqueurs biparentaux) tandis que l'ADN mitochondrial n'est transmis que par la lignée maternelle et le chromosome Y n'est transmis que par la lignée paternelle. De précédentes études portant sur ces deux derniers marqueurs ont ainsi permis de révéler des différences de signaux démographiques selon que l'on s'intéresse à l'histoire paternelle ou l'histoire maternelle, notamment en Asie centrale (Chaix *et al.*, 2007). Ainsi, l'étude de plusieurs marqueurs peut permettre de répondre à des questions différentes et complémentaires.

Les séquences autosomales et mitochondriales.

L'ADN mitochondrial, notamment le premier segment de la région contrôle hypervariable dite HVS-I (positions 16067 à 16384), présente l'avantage d'un taux de mutation relativement élevé. En effet, en fonction des méthodes utilisées, les estimations varient dans la littérature entre 5×10^{-6} (Forster *et al.*, 1996) et 10^{-5} (Howell *et al.*, 1996; Heyer *et al.*, 2001) par génération et par site pour la région HVS-I. L'ADN mitochondrial a notamment permis de confirmer l'origine africaine d'*Homo sapiens* (*e.g.* Cann *et al.*, 1987 ; Wilson et Cann, 1992 ; Cavalli-Sforza et Feldman, 2003), de localiser plus précisément cette origine en Afrique de l'Est (Quintana-Murci *et al.*, 1999), mais aussi de mieux comprendre le processus de colonisation de l'Eurasie (*e.g.* Pakendorf et Stoneking, 2005) et de détecter des expansions paléolithiques en Afrique et en Eurasie (*e.g.* Watson *et al.*, 1997; Chaix *et al.*, 2008 ; Batini *et al.*, 2011). Cependant, ce type de marqueur ne peut fournir qu'une vision partielle de l'histoire démographique des populations. Premièrement, il s'agit d'un marqueur uni-parental, qui ne nous renseigne que sur l'histoire des lignées maternelles. Deuxièmement, dû à l'absence de recombinaison, une séquence mitochondriale constitue en fait un seul locus, donc un seul réplicat du processus de coalescence. Finalement, on ne peut exclure des effets confondants de la sélection. En effet, des études (*e.g.* Pakendorf et Stoneking, 2005) ont

montré que l'ADN mitochondrial peut être soumis aux processus sélectifs. Or, la sélection peut amener à des signaux similaires à ceux d'une expansion démographique (Nielsen, 2001).

A l'inverse, les séquences autosomales offrent la possibilité d'analyser plusieurs locus indépendants, représentant autant de répliquats du processus de coalescence. Cela permet notamment de minimiser les effets confondants de la sélection, car celle-ci, contrairement aux processus démographiques, agit indépendamment sur chaque locus. De plus, il est possible, comme nous l'avons fait ici, d'utiliser uniquement des locus supposés neutres (*i.e.* suffisamment éloignés des locus sous sélection connus pour ne pas être en déséquilibre de liaison avec ceux-ci). Les séquences autosomales sont transmises à la fois par la lignée paternelle et la lignée maternelle. Elles présentent un taux de mutation de 1×2.10^{-8} à $2,5 \times 10^{-8}$ par génération et par site (Pluzhnikov *et al.*, 2002 ; Conrad *et al.*, 2011; Scally and Durbin, 2012), plus faible que la région HVS-I. Par le passé, elles ont notamment permis d'affiner la datation de l'exode d'*Homo sapiens* hors d'Afrique il y a 60 000 ans et de dater la divergence des ancêtres des Européens et des Asiatiques modernes il y a environ 22 500 ans (Laval *et al.*, 2010). Concernant l'Afrique, elles ont permis de dater la divergence entre les ancêtres des chasseurs-cueilleurs actuels et ceux des agriculteurs il y a environ 60 000 ans, et la divergence entre les chasseurs-cueilleurs de l'est et de l'ouest de l'Afrique centrale il y a environ 20 000 ans (Verdu *et al.*, 2009 ; Patin *et al.*, 2009). L'étude de séquences autosomales a également permis de détecter des expansions paléolithiques en Afrique ayant débuté il y a 50 000 à 25 000 ans (Voight *et al.*, 2005 ; Laval *et al.*, 2010).

Dans une première étude, présentée ci-après (**Chapitre I.B**), nous avons décidé d'utiliser conjointement ces deux types de marqueurs, afin d'obtenir des informations complémentaires sur l'histoire démographique ancienne des populations étudiées et de dater le début des grandes expansions humaines. D'un point de vue méthodologique, il s'agissait également de comparer les inférences obtenues à partir des deux types de marqueurs pour mieux comprendre l'impact de leurs propriétés respectives. Nous avons utilisé trois approches différentes basées sur la théorie de la coalescence : (1) des tests de neutralité (D de Tajima, 1989 ; D et F de Fu et Li, 1993 ; F_s de Fu, 1997), afin de détecter les changements démographiques *via* leur effet sur le spectre des fréquences alléliques, (2) Une approche paramétrique implémentée dans le logiciel BEAST v1.6 (Drummond et Rambaut, 2007), basée sur la simulation d'arbre de coalescence par MCMC pour inférer les paramètres démographiques, et (3) la méthode non-paramétrique EBSP (« Extended Bayesian Skyline Plots », Heled et Drummond, 2010) pour visualiser l'évolution de l'effectif efficace au cours

du temps. Les tests de neutralité sont généralement utilisés, comme leur nom l'indique, pour détecter des traces de sélection. Cependant, sous l'hypothèse de neutralité, ils nous renseignent également sur les processus démographiques. En effet, des évènements d'expansion se traduisent alors par une valeur négative des tests, alors que les évènements de contraction entraînent des valeurs positives. Les méthodes basées sur la simulation d'arbres de coalescence ont fait l'objet d'une description plus détaillée précédemment (**Introduction générale**). Notons que la sélection du meilleur modèle avec la méthode paramétrique a été effectuée par les méthodes de « Path sampling » et « Stepping-stone sampling » décrite par Baele *et al.* (2012), plutôt que grâce à la méthode originellement implémentée dans BEAST (estimation des Bayes factors utilisant les moyennes harmoniques). En effet, Baele *et al.* (2012) montrent une meilleure performance de ces méthodes sur des données simulées.

Les microsatellites autosomaux.

Les microsatellites autosomaux présentent un taux de mutation beaucoup plus élevé que les données de séquences, de l'ordre de 10^{-4} à 10^{-3} par génération et par site (Dib *et al.*, 1996 ; Jorde *et al.*, 1995 ; Weber et Wong, 1993). Des études récentes portant sur une espèce animale (le marsouin commun, Fontaine *et al.*, 2012) ou sur des données simulées (Cornuet *et al.*, 2008) ont montré que ce type de marqueurs peut nous renseigner sur des évènements plus récents que les marqueurs de type séquences. De plus, comme les séquences autosomales, ils offrent la possibilité d'analyser plusieurs locus indépendants.

Nous avons donc, dans un deuxième temps, décidé d'utiliser des marqueurs microsatellites autosomaux afin de tenter d'obtenir des informations sur l'histoire démographique plus récente des populations, et notamment d'adresser la question de l'impact de la transition Néolithique sur la démographie des populations humaines (**Chapitre I.C**). Les données microsatellites autosomales étant généralement moins informatives que les données de type séquences, la méthode non-paramétrique utilisée précédemment (EBSP) se révèle inefficace pour détecter des changements démographiques à partir ce type de données. Nous avons donc utilisé ici uniquement la méthode paramétrique, toujours grâce au logiciel BEAST v.1.6. Notons que la sélection du meilleur modèle a été effectuée ici à l'aide du Deviance Information Criteria (DIC, Spiegelhalter, 2002). En effet, l'article de Baele *et al.* (2012) proposant une implémentation des méthodes de « Path sampling » et « Stepping-stone sampling » (utilisées précédemment pour les séquences) n'était pas encore publié au moment de nos analyses sur les microsatellites autosomaux. Cependant, comme présenté dans le

chapitre suivant (**Chapitre II.B**), nos analyses sur des données simulées suivant des scénarios d'expansion, de contraction ou d'effectif constant ont démontré l'efficacité du DIC pour distinguer le modèle approprié.

Les microsatellites du chromosome Y.

Les microsatellites du chromosome Y présentent un taux de mutation élevé, dont les estimations dans la littérature varient entre $0,69 \times 10^{-3}$ (Zhivotovsky *et al.*, 2004) et $2,1 \times 10^{-3}$ (Heyer *et al.*, 1997). Etant donné l'absence de recombinaison, les différents loci sont liés, représentant comme pour l'ADN mitochondrial un seul réplicat du processus de coalescence. Il s'agit également d'un marqueur uni-parental, transmis cette fois-ci par la lignée paternelle.

La comparaison de la diversité génétique du chromosome Y et de l'ADN mitochondrial a souvent permis de détecter des processus démographiques sexe-spécifiques. Par exemple, Seielstad *et al.* (1998) démontrent un taux de migration plus élevé en moyenne chez les femmes que chez les hommes, probablement lié à la règle de patrilocalité (c'est-à-dire le fait que les femmes quittent leur lieu de vie pour rejoindre celui de leur mari, plutôt que l'inverse), qui s'applique dans la majorité des populations humaines. L'effectif efficace des femmes est souvent aussi plus élevé que celui des hommes (Heyer *et al.*, 2012). Au niveau génétique, cela se traduit un niveau de diversité génétique plus élevé pour l'ADN mitochondrial que pour le chromosome Y.

L'étude de ces marqueurs nous a permis ici de nous interroger sur l'impact des processus sexe-spécifiques sur nos inférences, notamment en relation avec les facteurs culturels telles que les règles de mariages (**Chapitre I.D.**). Cette question est particulièrement intéressante dans le cas de l'Asie centrale, comme déjà soulevé par Chaix *et al.* (2007). En effet, cette étude compare la diversité génétique des populations d'Asie centrale et montre une diversité supérieure chez les populations dites « turco-mongoles » (éleveurs nomades) par rapport aux populations dites « indo-iraniennes » (agriculteurs sédentaires). Cette différence serait due à davantage de migrations entre les populations d'éleveurs, organisées selon un système de tribus et clan et dans lesquelles les mariages se font majoritairement entre des tribus différentes (exogamie), et les populations d'agriculteurs, dans lesquelles les mariages ont lieu le plus souvent entre individus d'un même village (endogamie). Cependant, les populations des deux groupes étant patrilocales, cette différence s'observe uniquement lorsque l'on s'intéresse à l'ADN mitochondrial, marqueur transmis par la lignée maternelle. En effet,

ce sont les femmes des populations d'éleveurs qui vont migrer d'une tribu à l'autre par le biais des mariages.

Nous nous sommes interrogés sur l'impact de ces différences de taux de migration entre les sexes sur les inférences démographiques. Premièrement, selon Ray *et al.* (2003) et Excoffier (2004), les signaux d'expansions anciennes pourraient être atténués ou masqués dans les populations les plus isolées (*i.e.* recevant moins de 20 migrants par génération). Dans les populations patrilocales, ce processus est donc susceptible d'atténuer plus particulièrement les signaux d'expansion inférés à partir du chromosome Y. Deuxièmement, les migrations entre populations entraînent un apport d'allèles extérieurs à faible fréquence dans la population receveuse, ce qui pourrait mimer un signal d'expansion démographique, particulièrement pour les éleveurs nomades (exogames et patrilocaux) et lors de l'analyse de données mitochondriales. Les résultats de Chaix *et al.* (2007) vont en ce sens en montrant un signal expansion plus fort chez les agriculteurs que chez les éleveurs avec le chromosome Y, mais des signaux d'expansion similaires entre les deux groupes avec l'ADN mitochondrial. Il était donc important ici de s'intéresser au chromosome Y et de comparer les résultats avec ceux obtenus à partir de l'ADN mitochondrial, afin de vérifier dans quelle mesure et de quelle façon ces derniers sont influencés par les processus migratoires sexe-spécifiques. Nous avons également entrepris la même comparaison pour les autres régions étudiées, afin de détecter d'éventuels processus similaires. Pour l'analyse du chromosome Y, nous avons à nouveau utilisé les méthodes paramétrique et non-paramétrique implémentées dans BEAST. Les différents locus du chromosome Y étant liés, nous avons pu utiliser ici la méthode non-paramétrique dite « Bayesian skyline plots » (Drummond *et al.*, 2005), plutôt que la méthode dite « Extended Bayesian Skyline Plots » (Heled and Drummond, 2010), utilisée précédemment. En effet, comme montré par la suite à partir de données simulées (**Chapitre II.B et C**), la méthode EBSP ne permet généralement pas de distinguer des signaux d'expansion à partir de données microsatellites. Comme pour les microsatellites autosomaux, le meilleur modèle pour la méthode paramétrique a été sélectionné sur la base du DIC.

I.B. Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations, predating the emergence of farming.

Carla Aimé, Guillaume Laval, Etienne Patin, Paul Verdu, Laure Ségurel, Raphaëlle Chaix, Tatiana Hegay, Lluís Quintana-Murci, Evelyne Heyer, Frédéric Austerlitz.

Article soumis en l'état (version révisée) à Molecular Biology and Evolution.

I.B.i. Abstract.

Demographic changes are known to leave footprints on genetic polymorphism. Together with the increased availability of large polymorphism datasets, coalescent-based methods allow inferring the past demography of large sets of human populations from their present-day patterns of genetic diversity. Here, we analyzed both nuclear (20 non-coding regions) and mitochondrial (HVS-I) re-sequencing data to infer the demographic history of 66 African and Eurasian populations presenting contrasted life-styles (nomadic hunter-gatherers, nomadic herders and sedentary farmers). This allowed us to investigate the relationships between life-style and demography, and to address the long-standing debate about the chronology of demographic expansions and the Neolithic transition. In Africa, we inferred expansion events for farmers, but constant population sizes or contraction events for hunter-gatherers. In Eurasia, we inferred higher expansion rates for farmers than herders with HVS-I data, except in Central Asia and Korea. Although isolation and admixture processes could have impacted our demographic inferences, these processes alone seem unlikely to explain the contrasted demographic histories inferred in populations with different life-styles. The small expansion rates or constant population sizes inferred for herders and hunter-gatherers may thus result from constraints linked to nomadism. However, autosomal data showed contractions events for two sedentary populations in Eurasia, which may be linked with founder effects. Finally, the inferred expansions likely predated the emergence of agriculture and herding. This suggests that human populations could have started to expand in Paleolithic times, and that strong Paleolithic expansions in some populations may have ultimately favored their shift towards agriculture during the Neolithic.

I.B.ii. Introduction.

Studying the current distribution of genetic diversity in human populations has important implications for our understanding of the evolution and history of our species. Indeed, within- and among-population genetic diversity has been shaped both by demographic forces, such as gene flow and genetic drift, and by selective processes such as selective sweeps and local adaptation (*e.g.* Balaesque *et al.*, 2007). Cultural factors also have a considerable indirect impact on patterns of genetic diversity, as they can influence both the demographic and adaptive history, as clearly shown for social organization (*e.g.* Oota *et al.*, 2001; Kumar *et al.*, 2006; Heyer *et al.*, 2012) or technological innovations (Ambrose, 2001).

The Neolithic revolution is thought to be one of the most important cultural and technological transitions in human history. During this period, different human populations domesticated plants and animals in several parts of the world independently, including Western Central Africa, the Middle Eastern Fertile Crescent, Eastern Asia and Central America (Bocquet-Appel and Bar-Yosef, 2008). The emergence of farming occurred concomitantly with the sedentarization of most nomadic hunter-gatherer populations. Other populations remained nomadic, but some of them also developed new means of subsistence like nomadic herding. These major shifts in life-style and modes of subsistence involved profound changes in many aspects (diet, technologies, social organization...), which may have strongly impacted demographic patterns. Indeed, according to some archeologists and paleoanthropologists, the major human expansions would have started as a result from the Neolithic transition: sedentarized populations could have experienced strong demographic expansions (*e.g.* Bocquet-Appel, 2011), while nomadic populations may have remained constant because of inherent constraints of their life-style (*e.g.* a longer inter-birth interval, Short, 1982). However, a number of population genetic studies have reported evidence for more ancient expansion processes in many African and Eurasian populations, starting during the Paleolithic period (*e.g.* Chaix *et al.*, 2008; Atkinson *et al.*, 2009; Laval *et al.*, 2010; Batini *et al.*, 2011). These findings seem consistent with the “demographic theory” early suggested by Sauer (1952), according to which human populations could have started to increase before the Neolithic, and these Paleolithic expansions in some populations may have ultimately favored their shift toward farming.

The recent developments in sequencing technologies have allowed the exploration of large multi-locus polymorphism datasets. Together with modern statistical and bioinformatics tools, this can facilitate a better understanding of the forces shaping the distribution of human genetic diversity. In combination with archeological and paleo-anthropological records, it can substantially improve our ability to infer past demographical events (Beaumont, 2004). In this context, the coalescent theory (Kingman, 1982) provides a powerful framework to assess how gene genealogies (*i.e.* “coalescent trees”) are affected by demographic parameters such as the effective size and growth rate of populations. Numerical coalescent-based methods have thus been developed, allowing the inference of demographic parameters from molecular data. Most of these methods are based on the computation of the likelihood of the demographic parameters, *a priori* assuming a specific demographic model. Moreover, non-parametric approaches, such as Extended Bayesian Skyline Plots (EBSP, Heled and Drummond, 2010) allow inference of the demographic history of populations without assuming a specific model, by using the time intervals between serial coalescent events (see Excoffier and Heckel, 2006 and Ho and Shapiro, 2011 for reviews).

Here, we investigated (i) the relationship between life-style (*i.e.* sedentary farming, nomadic herding or nomadic hunting-gathering) and demographic patterns in a large set of African and Eurasian populations, and (ii) the chronology of demographic expansions and the emergence of farming, by comparing inferred expansions onset times with the dating of the most ancient archeological traces of farming and herding (potteries, irrigation structures, animals bones) reported in Bocquet-Appel and Bar-Yosef (2008) for each region. We used three coalescent-based methods: (i) a moment-based approach using neutrality tests (Tajima, 1989; Fu and Li, 1993; Fu, 1997) to detect the effects of past demographic changes on the allelic frequency spectrum (ii) a parametric approach using MCMC coalescent-based simulations to estimate demographic parameters, and (iii) a non-parametric approach using EBSPs (Heled and Drummond, 2010) to infer and visualize the evolution of effective population size through time. In addition, by computing F_{ST} values and immigration rates, we investigated the extent to which the inferred demographic patterns could be explained by spatial expansion processes. Indeed, modeling studies (Ray *et al.*, 2003, Excoffier, 2004) have shown that such processes can produce signals on within-population diversity patterns similar to those obtained with pure demographic expansions. In particular, these studies argue that ancient spatial expansion signals could be attenuated or suppressed in isolated populations.

Beside differences in effective size or population growth rates, differences in immigration rates and degrees of isolation may therefore explain variable signals of expansions among populations detected using genetic data.

We used 20 *a priori* neutral autosomal regions and the hypervariable control region (HVS-1) of the mitochondrial DNA (mtDNA) sequenced in, 404 individuals from 16 populations and 2429 individuals from 61 populations, respectively (**Table I.B-S1**). Given their distinct properties and modes of transmission, we compared the inferences obtained with these two types of markers, in order to gain complementary insights into the past demography of the studied populations. By studying many populations from different geographic areas worldwide, we were able to determine which patterns were observed across all populations and which were specific to a given geographical region. First, we focused on Central Africa, where nomadic hunter-gatherer populations, commonly called Pygmies, coexist with sedentary farmer populations. These two groups are genetically differentiated and seem to have diverged about 60,000 years ago (Patin *et al.*, 2009; Verdu *et al.*, 2009), thus long before the Neolithic sedentarization of farmer populations, which occurred between 5,000 and 4,000 years before present (YPB) in this area (Bocquet-Appel and Bar-Yosef, 2008). Second, we analyzed a sample of populations from several distant geographical regions of Eurasia where sedentary farmers coexist with nomadic herders. This was of particular interest, as to our knowledge the differences in demographic processes between herders and farmers have not been studied yet. Third, we performed a more detailed study in Central Asia, another area of interest as it is thought to have been a major corridor during the successive Eurasian migration waves (Nei and Roychoudhury, 1993).

I.B.iii. Results.

Neutrality tests

Focusing first on Africa, all farmer populations showed at least one significantly negative value for one of the four neutrality tests (Tajima's (1989) D , Fu and Li's (1993) D and F and Fu's (1997) F_s , **Table I.B-1**), which can be interpreted as a signal of expansion. Conversely, hunter-gatherer populations showed no expansion signal. Aka and Mbuti hunter-gatherers presented at least one significantly positive test, indicating a possible contraction event. Similarly, for HVS-I sequences, we found expansion signals (significantly negative Fu's F_s values) for all farmer populations except the Ewondo, but no expansion signal for

hunter-gatherers (Table I.B-S2). Kola hunter-gatherers showed a significantly positive Tajima's D, indicating a possible contraction event.

Table I.B-1. Summary statistics and neutrality tests computed from the whole autosomal sequences.

Population	Area	Life style	S	K	Tajima's D	Fu&Li's D	Fu & Li's F	Fu's F _s
Akele	Africa	sedentary farmers	6.95	6.45	-0.35	-0.55	<u>-0.57*</u>	<u>-1.12*</u>
Chagga	Africa	sedentary farmers	8.65	7.95	-0.48	<u>-0.70*</u>	<u>-0.74*</u>	<u>-1.27*</u>
Mozambicans	Africa	sedentary farmers	8.80	9.55	<u>-0.62*</u>	<u>-1.15**</u>	<u>-1.15 **</u>	<u>-3.33*</u>
Ngumba	Africa	sedentary farmers	7.05	6.20	-0.20	<u>-0.41*</u>	-0.41	-0.68
Yoruba	Africa	sedentary farmers	7.50	7.15	-0.14	-0.03	-0.03	-0.73
Aka	Africa	nomadic HG	6.95	6.60	0.12	0.34*	0.32	-0.30
G. Baka	Africa	nomadic HG	6.30	6.00	0.008	0.17	0.14	-0.33
S. Baka	Africa	nomadic HG	6.10	5.6	0.17	0.05	0.10	-0.03
Kola	Africa	nomadic HG	6.55	6.25	-0.14	-0.03	-0.08	-0.75
Mbuti	Africa	nomadic HG	6.60	6.10	0.25	0.35*	0.37*	0.16
Danes	Eurasia	sedentary farmers	5.50	4.85	0.30*	0.16	0.24*	0.73*
Han	Eurasia	sedentary farmers	5.20	4.70	-0.03	-0.01	-0.02	0.21
Japanese	Eurasia	sedentary farmers	4.20	3.85	0.45*	0.22	0.34*	1.06*
Chuvash	Eurasia	nomadic herders	5.70	5.05	0.09	0.11	0.12	0.34
Tajiks (TAB)	C. Asia	sedentary farmers	9.00	9.00	0.19	0.03	0.10	0.24
Kyrgyz (KIB)	C. Asia	nomadic herders	10.4	10.4	0.11	0.08	0.11	0.23

S: Number of polymorphisms, K: Number of haplotypes. We report the means over the 20 regions. Values significantly higher than expected on a constant population size model are indicated in bold, whereas significantly lower values are underlined. HG= Hunter-gatherers
Significance levels: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ after FDR correction for multiple testing (Benjamini and Hochberg, 1995).

Similar analyses for autosomes in Europe and East Asia revealed no significant expansion signals, neither in sedentary nor in nomadic populations (**Table I.B-1**). We even observed contraction signals in two sedentary populations, one East Asian and one European. Indeed, we found significantly positive values for two neutrality tests for the Japanese and three neutrality tests for the Danes. Conversely, for HVS-I sequences from Eurasia (**Table I.B-S2**), we obtained significant signals of expansion for at least one test (Fu's F_s) for all populations (including Japanese and Danes). All sedentary populations except Koreans also showed significant signals of expansion for the three other tests, while the Koreans and all nomadic populations showed a significant expansion signal only for Fu's F_s .

Focusing on Central Asia, no neutrality test was significant for the autosomes in either Tajik sedentary farmers (TAB) or Kyrgyz nomadic herders (KIB) (**Table I.B-1**). Conversely, for HVS-I sequences (**Table I.B-S2**), all farmers and herders (KIB and TAB included) presented a significantly negative value (*i.e.* expansion signal) for at least one test, except one farmer population (TDS).

Coalescent-based inferences of demographic history

Africa: Pre-Neolithic demographic expansions in sedentary farmer populations.

Considering first the autosomal data, models consistent with an increase in population size best fitted the data for all African farmer populations (**Table I.B-S3**). The “expansion model” best fitted the data for the two East-African farmer populations (namely Chagga and Mozambicans), while the “exponential model” best fitted the data for all West-African farmer populations (Akele, Ngumba and Yoruba), with positive growth rates in all cases (**Table I.B-S4**). Conversely, no signals of expansion were found for the hunter-gatherers, as the “constant model” always best fitted the data (**Table I.B-S3, I.B-S4**). Consistently, Extended Bayesian Skyline Plots (EBSPs) showed signals of expansions for farmer populations (**Figure I.B-1A**). HPD (Highest Probability Density) intervals for the estimated number of demographic changes did not include 0 for farmer populations, indicating at least one significant change in population size (**Table I.B-S5**). Conversely, we found no evidence of population size changes for hunter-gatherers (**Figure I.B-1B, Table I.B-S5**). Using the population growth curves generated from BEAST, we further dated the onset of farmer expansions from at least 62,275 YBP (assuming $\mu = 2.5 \times 10^{-8}$ /generation /site) or 124,550 YBP (assuming $\mu = 1.2 \times 10^{-8}$ /generation /site) for Mozambicans to 7,975 YBP or 15,950 YBP for Yoruba. Visual

examination of the HPD intervals showed that the expansion event inferred for the Mozambican population was significantly older than those inferred for the other populations. Dating of expansion onsets for each population are detailed in **Table I.B-S6**.

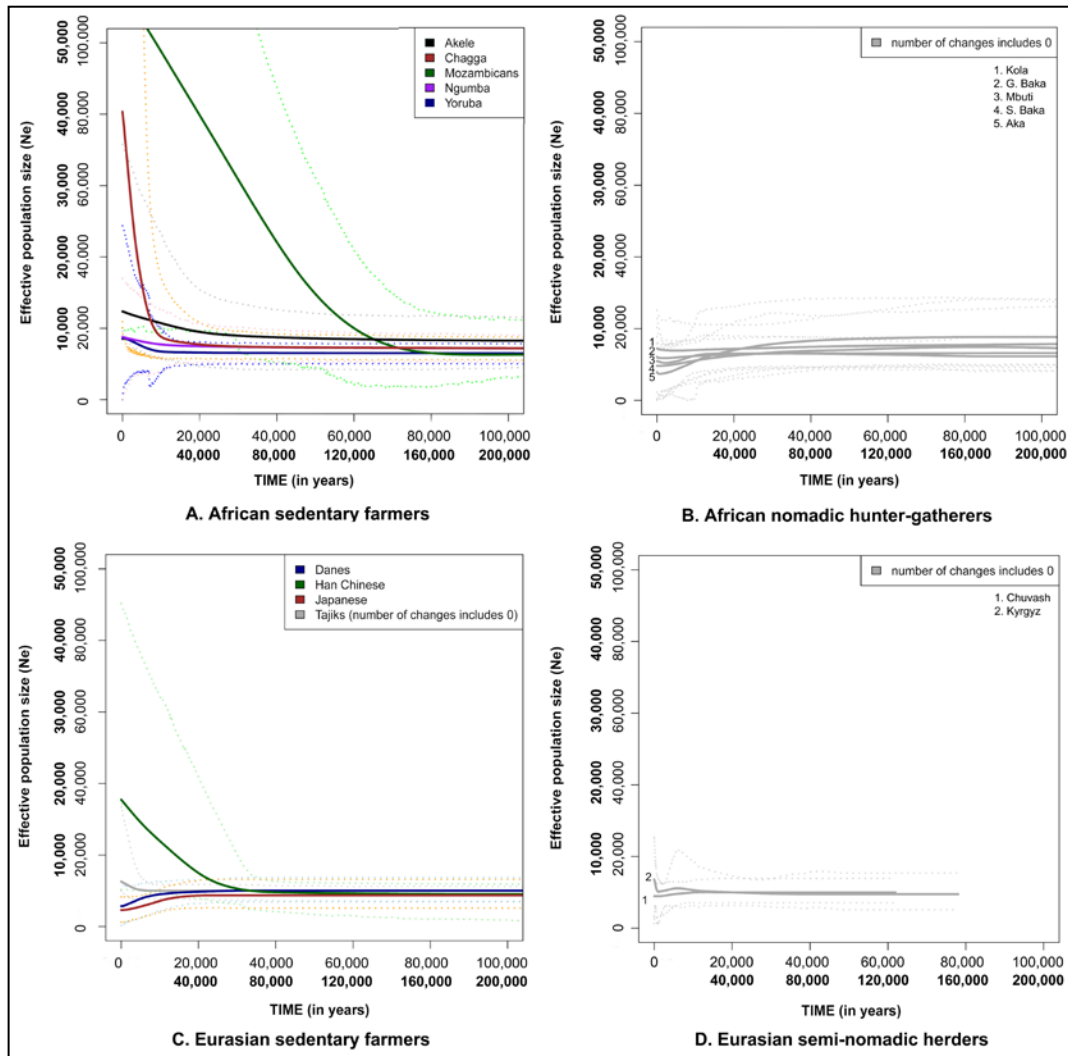


Figure I.B-1. Extended Bayesian Skyline Plots inferred from autosomal sequences in African sedentary farmers (A), African nomadic hunter-gatherers (B), Eurasian sedentary farmers (C) and Eurasian nomadic herders (D). The values indicated in bold on the axes are obtained assuming a mutation rate of $\mu = 1.2 \times 10^{-8}$ / generation / site (measured from trios parents-children by Conrad et al., 2011), and the other values correspond to $\mu = 2.5 \times 10^{-8}$ / generation / site (derived from the sequence divergence human-chimpanzee by Pluzhnikov et al., 2002). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time is represented backward on the X axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines. Populations for which the estimated number of demographic changes include 0 (i.e. no significant signal of expansion or decline) are represented in light grey colour.

We found similar results for the HVS-I sequences from Central Africa (**Table I.B-S3, I.B-S7**). Indeed, the “exponential model” with positive growth rates best fitted the data for all farmer populations, indicating expansion events. Conversely, the “exponential model” with negative modal values for growth rate (*i.e.* contraction event) provided the best fit for all hunter-gatherer populations. However, as the HPD intervals for growth rates included 0, we cannot conclude any significant contraction events for these populations. Similarly, EBSPs indicate a significant expansion event for all farmer populations (**Figure I.B-2A, Table I.B-S5**), while we found no evidence of population size changes for hunter-gatherers (**Figure I.B-2B, Table I.B-S5**).

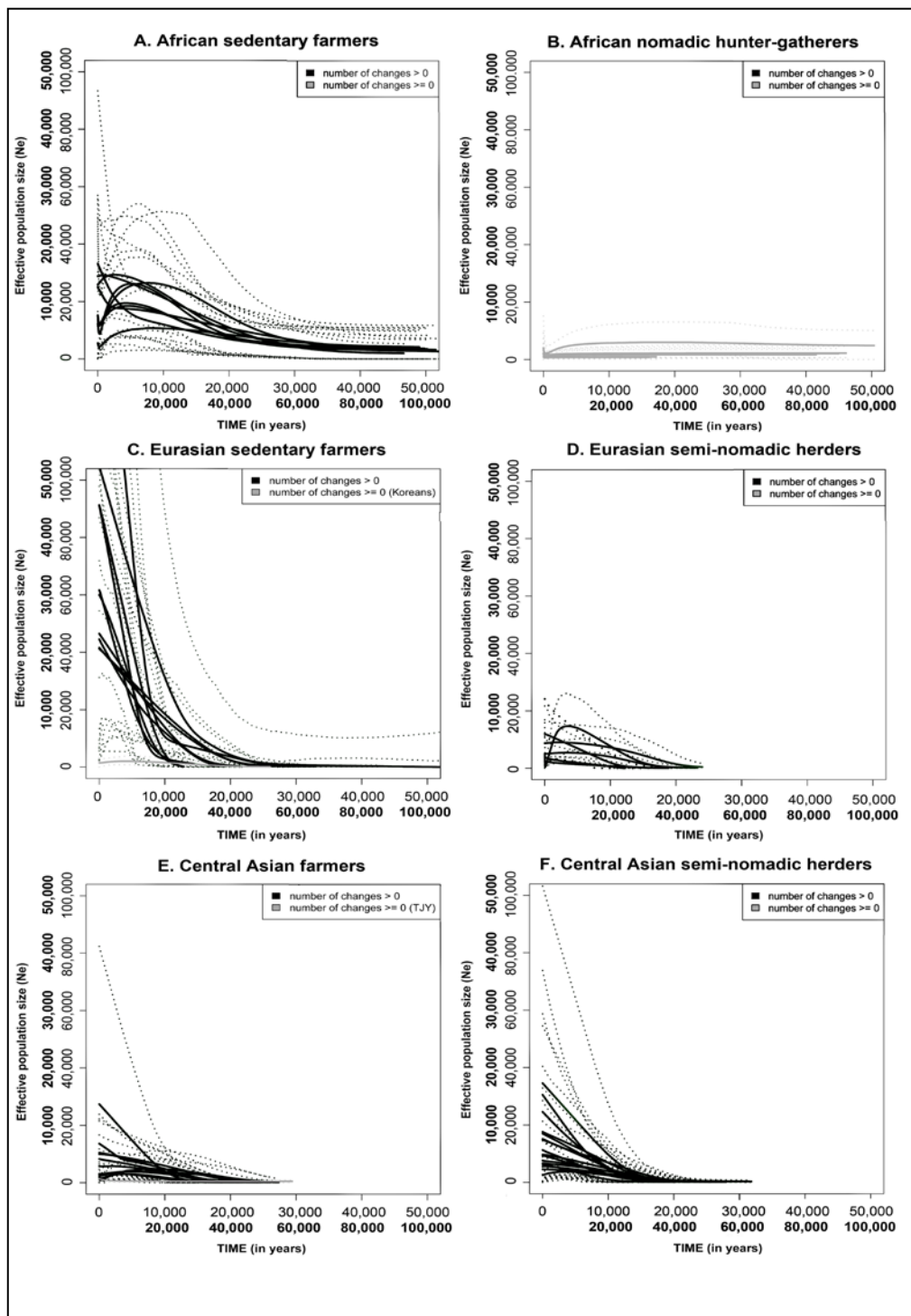


Figure I.B-2. Extended Bayesian Skyline Plots inferred from HVS-I sequences in African sedentary farmers (A), African nomadic hunter-gatherers (B), Eurasian sedentary farmers (C), Eurasian nomadic herders (D), Central Asian sedentary farmers (E) and Central Asian nomadic herders (F). The values indicated in bold are obtained assuming a mutation rate of $\mu = 5 \times 10^{-6}$ / generation /site (transitional changes rate, Forster et al., 1996), and the others correspond to $\mu = 10^{-5}$ / generation /site (pedigree-based, Howell et al., 1996; Heyer et al., 2001). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time is represented backward on the X axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines. Populations for which the estimated number of demographic changes include 0 (i.e. no significant signal of expansion or decline) are represented in light grey colour, and the others in black colour.

We dated the beginning of farmer populations expansions from 31,350 YBP or 62,700 YBP (assuming $\mu = 10^{-5}$ or 5×10^{-6} /generation /site, respectively) to 45,319 YBP or 90,638 YBP (Table I.B-S6).

Finally, note that considering both autosomes and HVS-I, all hunter-gatherer populations had lower current effective population size (N_0) values than farmer populations (Table I.B-S4, I.B-S7). Furthermore, both with autosomes and with HVS-I, the inferred expansion onsets for all farmer populations largely predated the emergence of farming in Central-Africa, between 5,000 YBP and 4,000 YBP (Bocquet-Appel and Bar-Yosef, 2008) (Figure I.B-3, I.B-4).

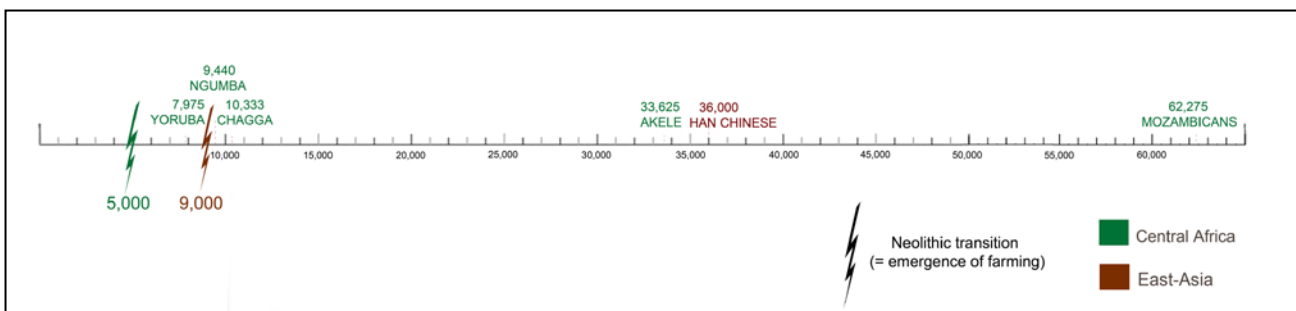


Figure I.B-3. Comparison of estimated times for expansion onsets using autosomes and dating of the first archeological traces of farming in Africa and China. Time is represented backward, in years before present (YBP). Only populations for which the EBSF analysis showed a significant expansion event are represented. We reported the time values estimated with the highest mutation rate that we used for the autosomes ($\mu = 2.5 \times 10^{-8}$ / generation /site). Thus, these time values can be considered as a lower bound for the expansion onsets. The dates for the emergence of farming come from the review by Bocquet-Appel and Bar-Yosef (2008). They are based on archeological remains.

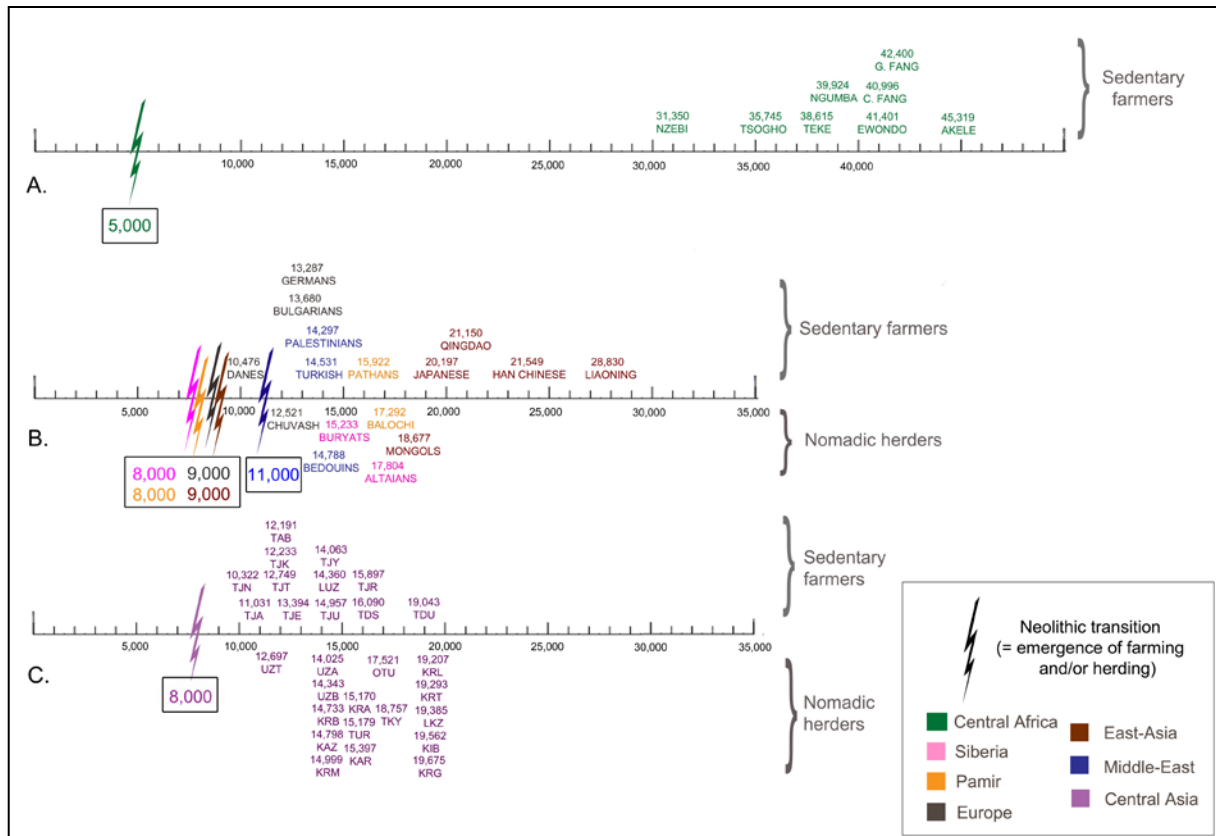


Figure I.B-4. Comparison of estimated times for expansion onsets using HVS-I and dating of the first archeological traces of farming or herding in Central Africa (A), Eurasia (B) and Central Asia (C). Time is represented backward, in years before present (YBP). Only populations for which the EBSA analysis showed a significant expansion event are represented. We reported the time values estimated with the highest mutation rate that we used for the HVS-I sequences ($\mu=10^{-5}$ / generation /site). Thus, these time values can be considered as a lower bound for the expansion onsets. The dates for the emergence of farming come from the review by Bocquet-Appel and Bar-Yosef (2008). They are based on archeological remains.

Eurasia: Contrasting demographic patterns for farmer populations with autosomes and stronger Pre-Neolithic expansions for farmers than herders with HVS-I.

The coalescent-based analyses of autosomes in East-Asian and European populations showed contrasting demographic patterns across sedentary populations (Table I.B-S3, I.B-S4). Using the parametric BEAST analysis, the “expansion model” with positive growth rate was inferred for Han Chinese, indicating an expansion event. Conversely, we inferred that Japanese and Danes either underwent a contraction event or remained at constant size. Indeed, the “exponential model” with negative growth rates best fitted the data for these two populations, but the confidence intervals also included $g=0$. The “constant model” best fitted the data for the Chuvash, a traditionally nomadic population. EBSAs showed a significant

expansion event for the Han population: the value of 0 was not included in the HPD interval of the number of demographic changes (**Table I.B-S5**). These expansion events started at least 36,025 or 72,050 YBP (**Figure I.B-1C, Table I.B-S6**), clearly predating the emergence of farming in East Asia, about 9,000 YBP (Bocquet-Appel and Bar-Yosef, 2008) (**Figure I.B-3**). Japanese showed a significant contraction event (*i.e.* the value of 0 was not included in the HPD interval of the number of demographic changes; **Table I.B-S5**) starting at least 21,350 YBP or 42,700 YBP. Danes also showed a significant contraction event, starting at least 26,440 YBP or 52,880 YBP (**Figure I.B-1C, Table I.B-S5, I.B-S6**). EBSP analyses showed no significant demographic changes for the Chuvash (**Figure I.B-1D, Table I.B-S5**).

For the HVS-I sequences from Eurasia, Central-Asia excluded (**Table I.B-S7, I.B-S8**), the parametric BEAST analyses showed that models consistent with an increase in population size (“expansion model” or “exponential model” with positive growth rates) best fitted the data for all sedentary populations except Koreans. Conversely, the “constant model” best fitted the data for all nomadic populations and for Koreans. EBSPs showed, however, significant expansion events for both farmers and herders, except for Koreans (**Figure I.B-2C, I.B-2D, Table I.B-S5**). Nevertheless, we noted a tendency to stronger expansion rates and higher N_e values in sedentary than in nomadic populations (**Figures I.B-2C, I.B-2D**), although the HPD intervals for N_e were quite large for sedentary populations. Finally, note that the estimated expansion onset times inferred from the EBSPs (**Table I.B-S6**) followed an east-to-west gradient: the expansion events appeared more ancient in Eastern populations, in both sedentary and nomadic populations (**Figure I.B-S1**). The inferred expansion onsets for all populations clearly predated the Neolithic transition in their respective geographic areas (**Figure I.B-4**).

The Central-Asian exception: similar demographic patterns in farmers and herders

For autosomes, the “constant model” best fitted the data for both sedentary farmers (TAB) and traditionally nomadic herders (KIB) (**Table I.B-S3, I.B-S4**). EBSPs showed also no significant demographic changes for these populations (**Figures I.B-1C, I.B-1D, Table I.B-S5**).

For HVS-I (**Table I.B-S3, I.B-S7**), the “exponential model” best fitted the data for six of the 12 sedentary farmer populations (including TAB), whereas the “constant model” was preferred for the other farmers. Unlike the rest of Eurasia, a model indicating expansion (the

“exponential model” with positive growth rates) was also selected for all nomadic herders. Moreover, EBSPs showed significant expansion signals for both herder and farmer populations except TJY, since at least 13,860 YBP (or 27,720 YBP) for farmers and 16,546 (or 33,092 YBP) for herders, on average (**Figure I.B-2E, I.B-2F, Table I.B-S5, I.B-S6**). Again, these inferred expansion onsets predated the emergence of farming in the area, about 8,000 YBP (Bocquet-Appel and Bar-Yosef, 2008) (**Figure I.B-4**). Inferred expansions for Central Asian sedentary farmers seemed overall weaker (*i.e.* lower growth rate and lower N_e) than those observed for other sedentary populations in Eurasia, although we observed important variations in growth rates and N_e among populations and large HPD intervals for some of them (**Figure I.B-2C, I.B-2E**).

Degrees of isolation and migration patterns.

African farmer populations appeared less isolated and received more migrants than hunter-gatherer populations. Indeed, the population-specific F_{ST} values (**Table I.B-S9**) were, on average, significantly lower for farmers than for hunter-gatherers ($\text{mean}_{(farmers)} = 0.058$; $\text{mean}_{(HG)} = 0.192$; Wilcoxon two-sided test p -value=0.0002). Moreover, the estimated number of immigrants was significantly higher for sedentary farmers than for nomadic hunter-gatherers ($\text{mean}_{(farmers)} = 31.4$; $\text{mean}_{(HG)} = 2.21$; p -value=0.0001) (**Table I.B-S10**). For hunter-gatherers, the F_{ST} values were negatively correlated with the negative growth rates that we inferred from the parametric method ($\rho = -0.893$; p -value= 0.012) (**Figure I.B-5B**), meaning that less isolated populations showed weaker contraction events (*i.e.* less negative growth rates). Conversely, there was no significant correlation between the F_{ST} values and the inferred growth rates for sedentary farmers ($\rho = 0.433$; p -value= 0.249) (**Figure I.B-5A**). However, we found a significant positive correlation between the number of immigrants and the inferred growth rates (**Figure I.B-S2**) among sedentary farmer populations ($\rho = 0.867$; p -value= 0.004) but not among nomadic hunter-gatherers ($\rho = 0.536$; p -value= 0.235).

Figure I.B-5. Correlations between population specific F_{ST} values and inferred growth rates in African farmer (A) and hunter-gatherer (B) populations, Eurasian farmer populations (C) and Central-Asian farmer (D) and herder populations (E). Population specific F_{ST} -values were computed with ARLEQUIN v3 (Excoffier et al., 2005). We used the growth rates inferred under the best-fitting model from the parametric method using BEAST (Drummond and Rambaut, 2007). When the best-fitting model was the “Constant model”, we assumed a growth rate of 0. Note that we did not represent Eurasian herder populations as the Constant model best-fitted the data for all of them. Plots and correlation tests were performed using R v2.14.1 (R Development Core Team, 2011).

For Eurasia, we found no significant difference in F_{ST} values between farmers and herders ($\text{mean}_{(\text{farmers})} = 0.039$; $\text{mean}_{(\text{herders})} = 0.043$; $p\text{-value} = 0.77$, **Table I.B-S9**), except in Central Asia, for which we found significantly lower F_{ST} values for nomadic herders than for sedentary farmers ($\text{mean}_{(\text{farmers})} = 0.018$; $\text{mean}_{(\text{herders})} = 0.008$; $p\text{-value} = 0.017$). We report a significant negative correlation between F_{ST} values and inferred growth rates for sedentary farmers in Eurasia ($\rho = -0.673$; $p\text{-value} = 0.028$) (**Figure I.B-5C**) and Central Asia ($\rho = -0.773$; $p\text{-value} = 0.003$) (**Figure I.B-5D**), thus meaning that less isolated populations showed higher inferred growth rates. There was no significant correlation for Central Asian herders ($\rho = -0.092$; $p\text{-value} = 0.736$) (**Figure I.B-5E**). Note that this analysis cannot be performed for the other Eurasian herder populations, as the “constant model” best-fitted the data with the parametric method.

The estimation of the proportion of immigrant did not converge for 11 Eurasian populations (Han Chinese, Liaoning, Qingdao, Palestinians, Pathans, Mongols, as well as three Central-Asian farmer populations and two Central-Asian herder populations). Regarding the other populations, we showed no significant difference in the proportion of immigrant between farmers and herders, both in Central Asia ($\text{mean}_{(\text{farmers})} = 82.868$; $\text{mean}_{(\text{herders})} = 260.265$; $p\text{-value} = 0.12$) and in the rest of Eurasia ($\text{mean}_{(\text{farmers})} = 576,757$; $\text{mean}_{(\text{herders})} = 201,311$; $p\text{-value} = 0.51$) (**Table I.B-S10**). We found also no significant correlation between the proportion of migrant and inferred growth rates for Eurasian farmers ($\rho = -0.238$; $p\text{-value} = 0.48$), Central Asians farmers ($\rho = 0.386$; $p\text{-value} = 0.30$) and Central Asian herders ($\rho = -0.42$; $p\text{-value} = 0.139$) (**Figure I.B-S2**).

I.B.iv. Discussion.

In this study, using a large set of populations from distant geographic areas, we report contrasted demographic histories that correlate well with life-style. Moreover, the expansion signals that we inferred in both African and Eurasian farmer and herder populations predate the Neolithic transition and the sedentarization of these populations. The use of Bayesian skyline plots allowed us to infer the demographic history of populations with no prior assumption of a particular demographic model. This is especially useful for analyzing populations for which previous knowledge about demographic history is scarce, which was often the case here (especially Eurasian nomadic herders). Note that the effective population sizes inferred using BEAST correspond to the N_e of the populations during their recent history, rather than a value of N_e averaged over the history of the population. It explains the finding that, for most populations, we inferred N_e estimates much higher than generally assumed for humans by population geneticists (about 10,000).

Contrasted demographic histories in sedentary and nomadic populations.

For Africa, both mtDNA and autosomal data revealed expansion patterns in most sedentary farmer populations, as indicated by neutrality tests and the parametric and non-parametric BEAST methods. Conversely, we found constant effective population sizes (or possibly contraction events) for all hunter-gatherer populations. Among the farmers, the Yoruba and the Ewondo populations yielded less clear results, as no neutrality test was significant for these populations while they showed evidence of expansion events when analyzed with BEAST. This indicates that these populations may have undergone weaker expansion dynamics (*i.e.* lower growth rates and N_e) than the others. These remarkable results are of particular importance for the Yoruba, as it is a reference population in many databases (HapMap, 1000 genomes). This also demonstrates the higher sensitivity of MCMC methods such as BEAST to detect expansions, in comparison to neutrality tests.

The contrasted patterns inferred between sedentary and nomadic populations in Africa suggest a strong difference in demographic dynamics between these two groups of populations. The question is whether this pattern results mostly from differences in local expansion dynamics or whether spatial expansion processes at a larger scale were also involved. As shown by Ray *et al.* (2003), negative values for the neutrality tests will be observed in a spatial expansion process if the rate of migrants (Nm) is high enough (at least

20), but not otherwise. As in previous studies (*e.g.* Verdu *et al.*, 2013), we report a higher degree of isolation (higher population-specific F_{ST} values) in hunter-gatherer populations than in farmer populations. Using the spatial expansion model of Excoffier (2004) also lead to higher estimates of the number of immigrants into the farmer populations. Thus, both farmers and hunter-gatherers may have been subject to a spatial expansion process, but the limited number of migrants among hunter-gatherers may have resulted in an absence of expansion signals for them. This would be consistent with the positive correlation that we observe between the growth rates estimated with BEAST and the inferred number of immigrants in the sedentary farmer populations. However, this spatial expansion process seems unlikely to completely explain the strong association that we observed between life-style and expansion patterns, as some farmer populations (Teke, Gabonese Fang) displayed F_{ST} values similar to those of hunter-gatherers but a clear signal of expansion with relatively high growth rates. This suggests that even rather isolated farmer populations show substantial level of expansions. Moreover, F_{ST} values and inferred growth rates in farmer populations were not significantly correlated. Therefore, our results suggest that the expansion patterns observed in sedentary populations result not only from a spatial expansion pattern, but also from local dynamics connected with the higher capacity of food production by farmers, hence explaining their much stronger expansion signatures, relative to neighboring hunter-gatherer populations.

For Eurasia, when considering the mtDNA data, all three methods (neutrality tests, parametric BEAST analyses and EBSPs) yielded expansion signals for all sedentary farmer populations except Koreans. Conversely only EBSPs and Fu's F_s test showed expansion signals for nomadic herders, but not the parametric BEAST method and the other neutrality tests. This result points towards weaker expansion dynamics in herders than in farmers, as supported also by the tendency for lower growth rates and N_e in herder populations than in farmer populations on the EBSP graphs (**Figure I.B-2C, I.B-2D**). It thus seems that the flexibility and non-parametric nature of EBSP analyses could permit to detect weaker expansion events than the parametric method. Moreover, Fu's F_s is known to be more sensitive than the other neutrality tests to detect expansions (Ramos-Onsins and Rozas, 2002). Again, these inferred expansions may result at least in part from spatial expansion processes. The population-specific F_{ST} values are indeed rather low in Eurasia. Moreover, we found a significant negative correlation between F_{ST} values and the inferred growth rates for the sedentary farmers, indicating that less isolated populations show a stronger expansion signal.

However, while we inferred much stronger expansion patterns for the farmers than for the herders, we did not observe any differences in Eurasia between the farmers and the herders in the population-specific F_{ST} values or in the estimated number of immigrants, suggesting that spatial processes alone cannot explain the strong difference that we observed between the expansion patterns of these two groups of populations. This indicates that the intrinsic demographic growth patterns are different between these two kinds of populations, the farmers showing much higher growth rates than the herders.

To our knowledge, while other studies have found different patterns between hunter-gatherers and farmers (*e.g.* Verdu *et al.*, 2009), our study is the first to show differences between farmers and herders, the two major post-Neolithic human groups. A plausible explanation could be that nomadic herders and hunter-gatherers share several of the constraints of a nomadic way of life. For instance, birth intervals are generally longer (at least four years) in nomadic populations than in sedentary populations (*e.g.* Short, 1982). According to Bocquet-Appel (2011), these longer birth intervals may be mainly determined by diet differences. Indeed, Vallengia and Ellison (2009) demonstrated that birth interval is mainly determined by the rapidity of post-partum energy recovery, which may be increased by high-carbohydrate food (like cereals) consumption. Moreover the nomadic herder way of life may offer less food security than sedentary farming, the latter facilitating efficient long-term food storage.

However, unlike in Africa, we did not find systematically consistent patterns between the autosomal and mtDNA data in Eurasia. The possible contraction events that our results suggest for two sedentary populations (Japanese and Danes) with autosomes appeared concomitant with historical events that could have led to bottleneck processes. For the Japanese population, this contraction signal could indeed result from a founder effect due to the Paleolithic colonization of Japan by a subset of the Northern Asiatic people (especially from Korea; Nei, 1995). Similarly, a bottleneck process may also have occurred in the Danish population, linked with the last glacial maximum occurring between 26,500 and 19,500 YBP (Clark *et al.*, 2009). The reasons for which these processes impacted the autosomes but not the mtDNA data remain to be determined, for instance through simulation studies. In any case, our study clearly emphasizes the utility of combining mtDNA and autosomal sequences, as they allow access to different aspects of human history. A recent study on harbor porpoises has similarly shown that nuclear markers were sensitive to a recent contraction event, while

the polymorphism of mtDNA markers reflected a more ancient expansion (Fontaine *et al.*, 2012).

Interestingly, Central Asia displayed a pattern distinct from the rest of Eurasia. Indeed, we did not infer higher expansion rates for sedentary farmers than for nomadic herders in that area. It could result from harsh local environmental conditions due to the arid continental climate in this area. Indeed, using pollen records, Dirksen and Van Geel (2004) showed that the paleoclimate in Central Asia was very arid from at least 12,000 YBP to 3,000 YBP, which could have limited the amount of suitable areas for farming and impacted human demography. Spatial expansion processes may also have played a role in this difference, as population-specific F_{ST} values were higher for the farmers than for the herders. This may indicate that more migrants were involved in the spatial expansion process for the herders than for the farmers, yielding a weaker expansion signal (*i.e.* lower inferred growth rate) for the latter (Ray *et al.*, 2003). This is supported by the negative correlation between the F_{ST} values and the inferred growth rates in the farmer populations.

The Korean population also stood out as an exception in Eurasia. Even though it is a population of sedentary farmers, it showed no significant expansion signal with both the parametric and non-parametric methods with HVS-I. This could be explained by a later sedentarisation of this population. Indeed, recent events are often difficult to detect with coalescent-based methods, as they can be masked by more ancient events (Lavery *et al.*, 1996). The Korean Neolithic is notably defined by the introduction of Jeulmun ware ceramics about 8,000 YBP, but the people of the Jeulmun period were still predominantly semi-nomadic fishers and hunter-gatherers. In fact, little is known about early Korean agricultural history, except that more recently, about 3,000 YBP, Koreans practiced intensive crop production implying a sedentary life style (Nelson, 1993).

Inferred expansion signals predate the emergence of farming

EBSP analyses led to an interesting observation: the inferred expansion events in farmers and herder populations appeared to be more ancient than the emergence of farming and herding. Therefore, the differences in demographic patterns between farmers and herders seem to predate their divergence in life-style, which raises the question of the chronology of demographic expansions and the Neolithic transition. These findings appear to be quite robust to the choice of the scaling parameters. We used here both the lower and the higher mutation

rate estimates in humans for autosomes (Pluzhnikov *et al.*, 2002; Conrad *et al.*, 2011) and for the HVS-I sequence (Forster *et al.*, 1996; Howell *et al.*, 1996). Despite this uncertainty in mutation rates, which lead to a two-fold uncertainty in our time estimates, the inferred expansion signals predated the emergence of agriculture in both cases for all populations. Similarly, using a generation time of 29 years (Tremblay and Vezina, 2000) instead of 25 years lead to slightly more ancient estimates, thus do not change our conclusions (data not shown).

However, note that for HVS-I, using the higher bound of the credibility interval for the highest estimated mutation rate (2.75×10^{-5} /generation/site, Heyer *et al.*, 2001) instead of the mean value (*i.e.* 10^{-5} /generation/site) lead to expansion time estimates that are consistent with the Neolithic transition in Eurasian populations (**Table I.B-S11**). Nevertheless, these estimates still clearly predated the Neolithic for the African populations. However, 10^{-5} /generation/site is by far the highest estimation of mutation rate in the literature (Howell *et al.*, 1996). To infer Neolithic expansions in most Eurasian populations, one needs to assume a mutation rate of at least 2×10^{-5} /generation/site, much higher than other estimates from the literature, and is thus probably unrealistic. Moreover, our method for determining the expansion onset time using EBSP graph is very conservative and also tends to favor the lower bound of expansion onset times. Finally, for autosomes, using similarly 4.74×10^{-8} /generation/site instead of 2.5×10^{-8} /generation/site (Pluzhnikov *et al.*, 2002) lead to an inferred expansion onset time that is not compatible with the Neolithic transition for all Eurasian and African populations, except for one African population, the Yoruba (**Table I.B-S12**). Consequently, it seems very likely that the expansions inferred in this study correspond to Paleolithic rather than Neolithic demographic events, in agreement also with most previous studies, as detailed below.

In Africa, the emergence of agriculture has been dated between 5,000 YBP and 4,000 YBP in the Western part of Central Africa and subsequently rapidly expanded to the rest of Sub-Saharan Africa (Phillipson, 1993). However, using HVS-I, we showed expansion events in farmer populations since about 30,000 or 60,000 YBP, thus largely pre-dating the emergence of agriculture in the area. Similarly, using autosomes, especially in Eastern African populations, we inferred expansion signals that were clearly too ancient to be a consequence of Neolithic changes in life style. Notably, we inferred an expansion signal for Mozambicans since at least 80,000 YBP. Several genetic studies have already highlighted that

expansion events occurred in African farmers before the Neolithic transition (*e.g.* Atkinson *et al.*, 2009; Laval *et al.*, 2010; Batini *et al.*, 2011). This finding is also consistent with paleoanthropological data (*i.e.* radiocarbon dating), which suggest an ancient expansion event in Africa 60,000 - 80,000 YBP (Mellars, 2006a). This Paleolithic demographic expansion could be linked to a rapid environmental change towards a dryer climate (Partridge *et al.*, 1997) and/or to the emergence of new hunting technologies (Mellars, 2006a).

According to Mellars (2006a), this period corresponds to a major increase in the complexity of the technological, economic, social, and cognitive behavior of certain African groups. It corresponds in particular to the emergence of projectile technologies (Shea, 2009), which was probably part of a broader pattern of ecological diversification of early *Homo sapiens* populations. These changes could have been decisive for the human spread “Out of Africa” during the same period, and could have ultimately also led to the sedentarization of the remaining populations. This inference is consistent with the demographic theory of Sauer (1952), who stated that late Paleolithic demographic expansions could have favored the sedentarization and the emergence of agriculture in some human populations. In the case of Central Africa, the period of 60,000 YBP corresponds to the separation between hunter-gatherers and farmers ancestors (Patin *et al.*, 2009; Verdu *et al.*, 2009). Thus, these two groups may have presented contrasting demographic patterns since their divergence. Much later, higher expansion rates and larger population sizes among farmers’ ancestors may have induced the emergence of agriculture and sedentarization.

With respect to Eurasia, the expansion profiles inferred with HVS-I for all populations and with autosomes for the Han Chinese population also seem to have begun during the Paleolithic, thus before the Neolithic transition. Some genetic studies already reported pre-Neolithic expansions in Asia and Europe (*e.g.* Chaix *et al.*, 2008). Notably, using mismatch and intermatch distributions, Chaix *et al.* (2008) showed an east-to-west Paleolithic expansion wave in Eurasia. We found a similar pattern here, as the inferred expansions of East Asian populations were earlier than the expansions of Central Asian populations, themselves earlier than those of European populations. Moreover, we found this pattern in both sedentary farmer and nomadic herder populations. Thus the ancestors of currently nomadic herder populations also experienced these Paleolithic expansions. However, Paleolithic expansion signals in nomadic populations seem lower than in sedentary populations. This is again compatible with the demographic theory of the Neolithic sedentarisation (Sauer, 1952): some populations may

have experienced more intense Paleolithic expansions, which may have led ultimately to their sedentarisation.

The inferred Paleolithic expansion signals might result partly from spatial expansions out of some refuge areas after the LGM (26,500 - 19,500 YBP, Clark *et al.*, 2009), as this time interval matches with our inferred dating for expansion onsets in East-Asia with HVS-I using the pedigree-based mutation rate, and in Europe and Middle East using the transitional mutation rate. Some of the earlier date estimates might also be consistent with the out-of-Africa expansion of *Homo sapiens*. However, the time estimates of the spread of *Homo sapiens* in Eurasia from radiocarbon datation of archeological remains are generally more ancient than our inferred expansion onset timings. For instance, Mellars (2006b) dated the colonization of the Middle East by *Homo sapiens* 47,000 – 49,000 YBP and the colonization of Europe around 41,000 – 42,000 YBP. Pavlov *et al.* (2001) report traces of modern human occupation nearly 40,000 years old in Siberia. Finally, Liu *et al.* (2010) described modern human fossils from South China, dated to at least 60 000 YBP. Moreover out-of-Africa or post-LGM expansions would not explain our finding of an east-to-west gradient of expansion onsets timing, which rather supports the hypothesis of a demographic expansion diffused from east to west in Eurasia in a demic (*i.e.* migrations of individuals) or cultural (favored by the diffusion of new technologies).

Possible confounding factors

Our approach makes the assumption that populations are isolated and panmictic, which is questionable for human populations. However, we analyzed a large set of populations sampled in very distant geographical regions (*i.e.* Central Africa, East Africa, Europe, Middle-East, Central Asia, Pamir, Siberia and East-Asia). The main conclusions of this study rely on consistent patterns between most of these areas, and it seems unlikely that processes such as admixture could have biased the estimates similarly everywhere. Moreover, in Central Africa, several studies have shown that hunter-gatherer populations show signals of admixture, while it is not the case for farmer populations (Patin *et al.*, 2009; Verdu *et al.*, 2009, 2013). If this introgression had been strong enough, this may have yielded a spurious expansion signal in the hunter-gatherer populations, which is not what we observed here. In Europe, spatial expansion processes during the Neolithic may have led to admixture with Paleolithic populations. As pointed out by a simulation study (Arenas *et al.*, 2013), this may

lead to a predominance of the Paleolithic gene pool. This may be one of the factors explaining why we observed mostly Paleolithic expansions here.

Similarly, potential selection occurring on the whole mitochondrial genome (*e.g.* Pakendorf and Stoneking, 2005) seems unlikely to have impacted in the same way all the studied populations within each group (*e.g.* stronger positive selection on sedentary than on nomadic populations), as we analyzed different nomadic and sedentary populations living near each other, in several geographically distant areas.

Regarding the potential effects of recombination on the inferences from autosomal data, we found that neutrality tests gave similar results on the whole sequences, when using a simulation procedure that was taking the known recombination rate of each sequence into account (**Table I.B-1**), and on the largest non-recombining blocks as inferred with IMgc, without taking recombination into account in the simulation process (**Table I.B-S12**). It thus appears unlikely that the BEAST analyses that can only handle the largest inferred non-recombining blocks are biased because of this.

Further work

It will be interesting to expand our analyses to include other types of genetic markers. In particular, paternally-inherited markers (Y-chromosome) could provide more information about the paternal history of populations. Simulation studies will also be planned to better understand the specific responses of each type of genetic markers (*e.g.* autosomes, mitochondrial DNA, Y-chromosome) to past demographical events. Finally, genome-wide and next-generation sequencing data also open up new interesting perspectives, as the large number of markers that they offer should allow more accurate and precise inferences. For instance, it is now possible to infer N_e through time from the distribution of polymorphic sites in complete genome sequences from one individual, using the coalescent model with recombination suggested by Li and Durbin (2011).

I.B.v. Material and methods

Genetic markers

- *Autosomal sequences*

The 20 non-coding, a priori neutral and unlinked autosomal regions have been selected by Patin *et al.* (2009) to be at least 200 kb away from any known or predicted gene, to not be in linkage disequilibrium (LD) neither with each other nor with any known or predicted gene, and to have a region of homology with the chimpanzee genome. These 20 regions are on average 1253 bp long, for a total length of 25 060 bp. Using the four-gamete test (Hudson and Kaplan, 1985) as implemented in IMgc online (Woerner *et al.*, 2007), we identified recombination events for six among these 20 regions. As some methods used in this study cannot handle recombination, for these six sequences we retained the largest non-recombining block within each of them (with IMgc). Because of this reduction, the 20 regions used were on average 1 228bp long (i.e. a total length of 24 560bp). To identify potential bias related to this method (*e.g.* some recombination events may not be detected using the four-gamete test; larger blocks of non-recombining sequence may select for gene trees that are shorter than expected), we computed the summary statistics and performed neutrality tests (see below) both on the whole sequences (**Table I.B-1**) and on the largest non-recombining blocks (**Table I.B-S12**).

- *Mitochondrial sequences*

We used the first hyper-variable segment of the mitochondrial control region (HVS-I), sequenced between positions 16067 and 16383, excluding the hypervariable poly-C region (sites 16179 to 16195). The total length of the sequence was thus of 300 bp per population.

Population panel

For Africa, we used the autosomal sequences dataset of Patin *et al.* (2009), which consists of five farmer populations, with a total of 118 individuals, and five Pygmy hunter-gatherer populations with a total of 95 individuals. In addition, we used the HVS-I dataset from Quintana-Murci *et al.* (2008), which consists of 486 individuals from nine Central-African farmer populations and 318 individuals from seven Central-African hunter-gatherer populations (**Table I.B-S1**).

For Eurasia, we used the autosomal sequences dataset of Laval *et al.* (2010), consisting of 48 individuals from two East-Asian populations (Han Chinese and Japanese) and 47 individuals from two European populations (Chuvash and Danes). We also used the data from 48 individuals from one sedentary Central-Asian population (Tajik farmers) and 48

individuals from one nomadic Central-Asian population (Kyrgyz herders), sequenced by Ségurel *et al.* (2013). For HVS-I, we analyzed data from 17 Eurasian populations located from Eastern to Western Eurasia, belonging to several published data sets (**Table I.B-S1**). It represents a total of 494 individuals.

For our detailed study of Central-Asia, we used HVS-I sequences from 28 populations: 12 farmer populations (408 individuals in total) and 16 herder populations (567 individuals in total). Among these populations, the data come from the studies by Chaix *et al.* (2007) and Heyer *et al.* (2009) for 25 populations (**Table I.B-S1**). The other three populations (KIB, TAB, and TKY) were sequenced for this study. As in Chaix *et al.* (2007) and Heyer *et al.* (2009), DNA was extracted from blood samples using standard protocols, and the sequence quality was ensured as follows: each base pair was determined once with a forward and once with a reverse primer; any ambiguous base call was checked by additional and independent PCR and sequencing reactions; all sequences were examined by two independent investigators. All sampled individuals were healthy donors from whom informed consent was obtained. The study was approved by appropriate Ethic Committees and scientific organizations in all countries where samples have been collected.

Demographic inferences from sequences analysis

- Summary statistics and neutrality tests

We computed classical summary statistics (number of polymorphic sites (S), number of haplotypes (K)) and four neutrality tests (Tajima's (1989) D , Fu and Li's (1993) D^* and F^* and Fu's (1997) F_s) on both mitochondrial and autosomal sequences. Although neutrality tests were originally designed to detect selective events, they also give information about demographic processes, especially when applied to neutral markers. Indeed, the excess of singleton mutations expected in case of expansion lead to more negative values than expected in absence of selective and demographic processes. Conversely, contraction events lead to default of singletons and more positive values of the neutrality tests. For HVS-I sequences, we both computed summary statistics and neutrality tests and tested their departure from neutrality using the coalescent-based tests provided in DnaSP (Librado and Rozas, 2009).

For the autosomal sequences, we used the procedure developed in Laval *et al.* (2010), which combines all autosomal sequences into a single test. This procedure consists in

computing the mean value of each summary statistics across the 20 loci, and in testing whether this mean value departs significantly from its expectation under neutrality in a constant-size population model, using a simulation procedure. For a given population with sample size n , we produced 10^5 simulated samples of the same size n , under a constant population size model, using the generation-per-generation coalescent-based algorithm implemented in SIMCOAL v2 (Laval and Excoffier, 2004). Each simulated individual was constituted by 20 independent sequences of 1253pb (same as the average length of the experimental sequences; see above). Then, we used ARLEQUIN v3 (Excoffier *et al.*, 2005) as modified by Laval *et al.* (2010) to compute the summary statistics on these simulated samples. We assessed whether the observed statistics differed significantly from the constant population model under neutrality by comparing these statistics with their null distribution obtained from the simulated data. We used gamma distributed mutation rates with a mean value of 2.5×10^{-8} / generation / site [95% CI: 1.476×10^{-8} - 4.036×10^{-8}], in agreement with previous studies (Pluzhnikov *et al.*, 2002; Voight *et al.*, 2005). This procedure yielded a p -value for the significance of the departure from a constant size model. We performed this procedure both on the whole sequences and on the largest non-recombining blocks. For the whole sequences (*i.e.* including recombination), we performed the data simulation under a coalescent model with recombination, using for each locus the recombination rate provided by the HapMap build GRCh37 genetic map (The International HapMap Consortium, 2003) (**Table I.B-S8**), while for the largest non-recombining blocks we used a coalescent model without recombination

Both for autosomes and for HVS-I sequences, we adjusted the obtained p -values for each neutrality test using a FDR correction (Benjamini and Hochberg, 1995) in R v2.14.1 (R Development Core Team, 2011), in order to take in account the increased error probability in the case of multiple testing.

- *MCMC Estimations of demographic parameters*

We used the MCMC algorithm implemented in BEAST v1.6 (Drummond and Rambaut, 2007). We tested the four demographic models implemented in this software: constant effective population size (N_e) (“Constant model”), population expansion with an increasing rate of growth (g) (“Exponential model”), population expansion with an decreasing rate of growth (g) (“Logistic model”) and expansion starting at a given moment in the past from an

ancestral population of constant size (“Expansion model”). In fact, BEAST estimates composite parameters for each model, namely $N_0\mu$ and g/μ , where N_0 is the current effective population size, g the growth rate and μ the mutation rate. In addition, for the “Expansion model”, the ratio between the current (N_0) and ancestral (N_1) effective population size is also estimated. To infer N_0 and g from these composite parameters, we then need to assume a value for the mutation rate μ . However, there is no consensus for mutation rates in humans in the literature, as different methods lead to different estimations. For autosomes, the most commonly-used value is the phylogenetic rate of $\mu = 2.5 \times 10^{-8}$ /generation /site, estimated by Pluzhnikov *et al.*, 2002. However, recent studies based on the 1,000 genome project (The 1000 Genomes Project Consortium, 2012) have found a two-fold lower rate ($\mu = 1.2 \times 10^{-8}$ /generation /site) by directly comparing genome-wide sequences from children and their parents (Conrad *et al.*, 2011; Scally and Durbin, 2012). We used here both mutation rates. Similarly, for HVS-I, estimated mutation rates are highly dependent on methodologies and modes of calibration that are used (*e.g.* Endicott *et al.* 2009). We used both the lower and the higher estimated mutation rates: the transitional changes mutation rate of $\mu = 5 \times 10^{-6}$ /generation/site (Forster *et al.*, 1996) and the pedigree-based rate of $\mu = 10^{-5}$ /generation /site (Howell *et al.*, 1996; Heyer *et al.*, 2001). We used a general time-reversible (GTR) substitution model (Rodriguez *et al.*, 1990). We assumed a generation time of 25 years, permitting the comparison with previous human population genetics studies (*e.g.* Chaix *et al.*, 2008; Patin *et al.*, 2009; Laval *et al.*, 2010). As BEAST cannot handle recombination events, we used the largest non-recombining block within each sequence (see above).

We performed three runs of 10^7 steps per population and per demographic model for the HVS-I sequence, and three runs of 2×10^8 steps (which corresponded to three runs of 10^7 steps per locus) for the autosomal sequences. We recorded one tree every 1000 steps, which thus implied a total of 10^5 trees per locus and per run. We then removed the first 10% steps of each run (burn-in period) and combined the runs to obtain acceptable effective sample sizes (ESS of 100 or above). The convergence of these runs was assessed using two methods: visual inspection of traces using Tracer v1.5 (Rambaut and Drummond, 2007) to check for concordance between runs, and computation of the Gelman and Rubin's (1992) convergence diagnostic using R v2.14.1 (R Development Core Team, 2011) with the function “gelman.diag” available in the add-on package coda (Plummer *et al.*, 2006).

In order to facilitate a large exploration of the parameter space, for the autosomal sequences, we chose uniform priors between 0 and 0.05 for $2N_0\mu$ and between -10^9 and 10^9 for g/μ . For HVS-I sequences, we chose uniform priors for $N_0\mu$ between 0 to 10 and for g/μ between -2.5×10^6 to 2.5×10^6 , resulting for the same priors on N_0 and g than for autosomal sequences if we assumed $\mu = 10^{-5}$ /generation /site (*i.e.* N_0 constrained between 0 and 10^6 and g constrained between -1 and 1 per year). Conversely, if we assumed $\mu = 5 \times 10^{-6}$ /generation /site, it means that N_0 was constrained between 0 and 2×10^6 and g was constrained between -0.5 and 0.5 per year.

For each population and model, we obtained the mode and the 95% HPD of N_0 and g , inferred from their posterior distributions (**Table I.B-S4, I.B-S7**) using the add-on package Locfit (Loader, 1999) in R v2.14.1. We selected the best-fitting model among the four tested demographic models by estimating marginal likelihoods using both two methods: path-sampling and stepping stone sampling, both implemented and described by Baele *et al.* (2012). The model with the greater marginal likelihood (**Table I.B-S3**) is considered as the best fitting model.

Extended Bayesian Skyline Plots (EBSP)

EBSPs (Heled and Drummond, 2010), also implemented in BEAST, estimate demographic changes occurring continuously through time in a population, using the time intervals between successive coalescent events. This method allows a visualization of the evolution of N_e through time. As above, we combined three runs of 10^7 steps for mitochondrial sequences and three runs of 2×10^8 steps for autosomal sequences to obtain acceptable ESS values. We assumed the same mutation rates as above, and a generation time of 25 years. Outputs were analyzed with Tracer v1.5 to visually check for convergence and ESS, also to obtain the HPD interval for the number of demographic changes that occurred in the population (**Table I.B-S5**). A constant population size can be rejected since the HPD of the number of change points exclude 0 but contained 1 (Heled and Drummond, 2010). Then, we used R v2.14.1 to compute Gelman and Rubin's (1992) convergence diagnostic as above, as well as to compute skyline plots. Finally, we used the population growth curves generated from BEAST to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 – 20 generation intervals. We consider that the population increase (or decrease) since both the median and HPD values for N_e increase (or

decrease) between more than 2 successive data points. Although this method does not allow providing a confidence interval for inferred expansion timings, this conservative approach ensure that we consider only relevant expansion signals.

Correlation tests of inferred growth rates and isolation/ immigration patterns.

In order to test how isolation degrees and migration patterns differences could have impacted our demographic inferences, we used ARLEQUIN v3.11 (Excoffier *et al.*, 2005) to (1) calculate population-specific F_{ST} values (Weir and Hill, 2002) from HVS-I data for Central Africa, Eurasia and Central Asia (**Table I.B-S9**), (2) compute immigration rates from mismatch distributions under a spatially explicit model (Excoffier, 2004) (**Table I.B-S10**). We then performed spearman tests using R v2.14.1 to investigate how inferred growth rates with the parametric method for each region were correlated to those F_{ST} values and immigration rates. We considered a value of 0 for growth rate when the constant model best fitted the data.

I.B.vi. Supplementary Figures.

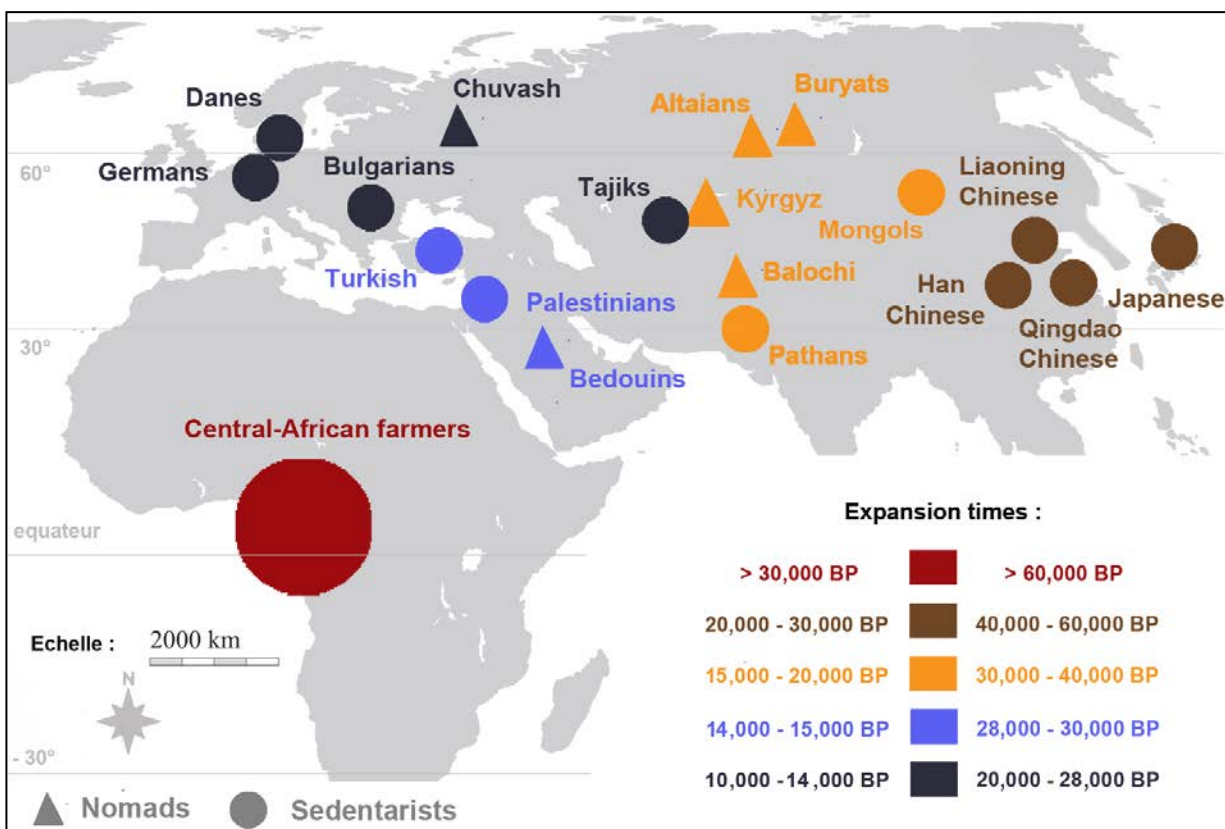
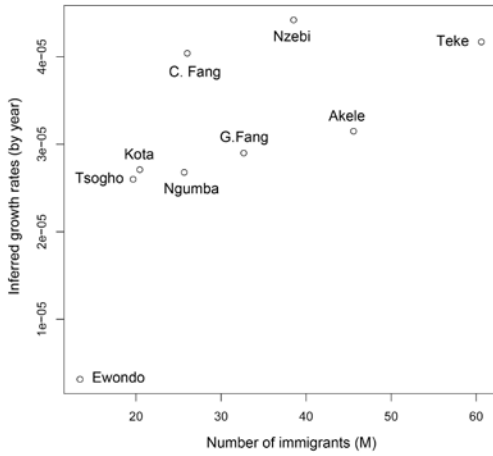


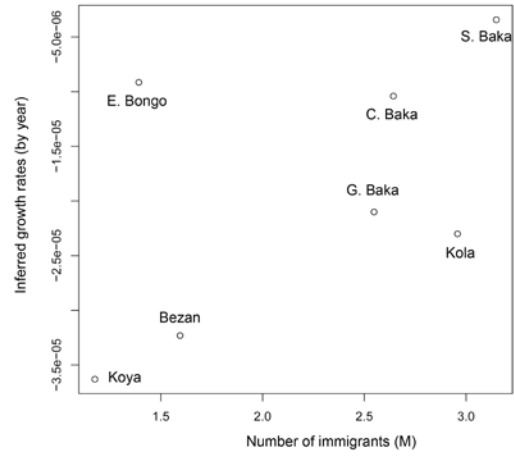
Figure I.B-S1. Distribution of expansion ages in Eurasia inferred from HVS-I sequences using EBSPs. *Populations are colored according to their expansion time, as indicated by the scale, with the times obtained with the transitional changes rate on the left and the pedigree-based mutation rate on the right. Sedentary populations are represented by circles and nomadic populations are represented by triangles. We represented only the populations for which the EBSPs showed at least one significant expansion event (we did thus not represent the Koreans and African hunter-gatherer populations). We grouped all Central African farmer populations because of their geographical proximity and similar expansion onset times.*

A. African sedentary farmers



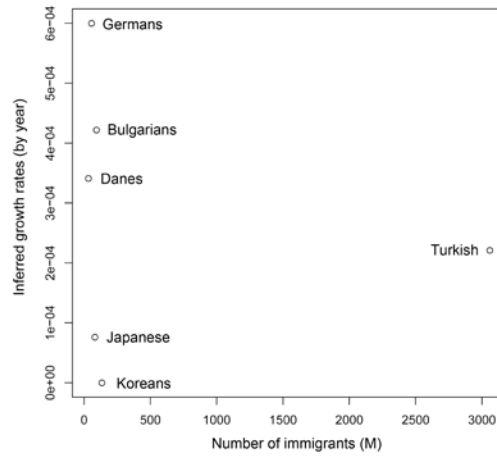
Correlation coefficient (Spearman) = 0.867, p-value = 0.004

B. African nomadic hunter-gatherers



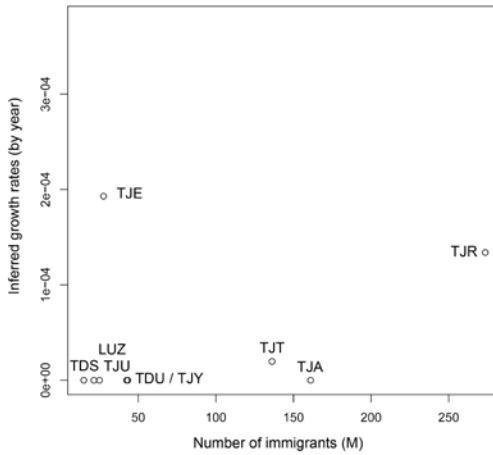
Correlation coefficient (Spearman) = 0.536, p-value = 0.235

C. Eurasian sedentary farmers



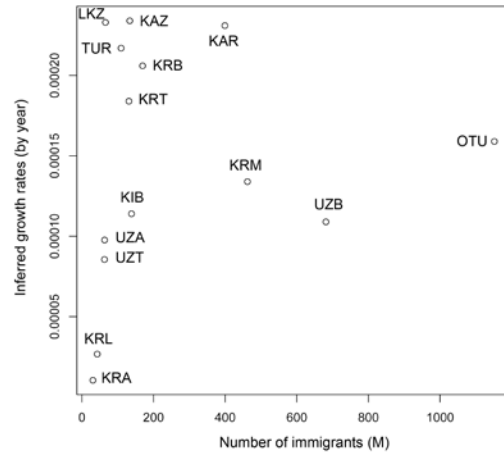
Correlation coefficient (Spearman) = -0.238, p-value = 0.48

D. Central Asian sedentary farmers



Correlation coefficient (Spearman) = 0.386, p-value = 0.30

E. Central Asian nomadic herders



Correlation coefficient (Spearman) = 0.42, p-value = 0.139

Figure I.B-S2. Correlations between the number of immigrant and inferred growth rates in African farmer (A) and hunter-gatherer (B) populations, Eurasian farmer populations (C), and Central-Asian farmer (D) and herder populations (E). *We obtained the number of immigrants from mismatch distributions under a spatially explicit model, using ARLEQUIN v3 (Excoffier et al., 2005). We used the growth rates inferred under the best-fitting model from the parametric method using BEAST (Drummond and Rambaut, 2007). When the best-fitting model was the “Constant model”, we set the growth rate at 0. Note that we did not represent Eurasian herder populations as the Constant model was preferred for all of these. Plots and correlation tests was performed using R v2.14.1 (R Development Core Team, 2011).*

I.B.vii. Supplementary Tables.

Table I.B-S1. Description and location of studied populations.

Population	Code	area	Country	lifestyle	n _{AUTO}	n _{HVS-I}	Reference (autosomes)	Reference (HVS-I)
Akele	–	Central Africa	Gabon	sedentary farmers	16	46	Patin <i>et al.</i> , 2009	Quintana-Murci <i>et al.</i> , 2008
Chagga	–	East Africa	Tanzania	sedentary farmers	32	–	Patin <i>et al.</i> , 2009	–
Ewondo	–	Central Africa	Cameroon	sedentary farmers	–	25	–	Quintana-Murci <i>et al.</i> , 2008
C. Fang	–	Central Africa	Cameroon	sedentary farmers	–	39	–	Quintana-Murci <i>et al.</i> , 2008
G. Fang	–	Central Africa	Gabon	sedentary farmers	–	65	–	Quintana-Murci <i>et al.</i> , 2008
Kota	–	Central Africa	Gabon	sedentary farmers	–	56	–	Quintana-Murci <i>et al.</i> , 2008
Mozambicans	–	East Africa	Mozambique	sedentary farmers	23	–	Patin <i>et al.</i> , 2009	–
Ngumba	–	Central Africa	Cameroon	sedentary farmers	16	74	Patin <i>et al.</i> , 2009	Quintana-Murci <i>et al.</i> , 2008
Nzebi	–	Central Africa	Gabon	sedentary farmers	–	62	–	Quintana-Murci <i>et al.</i> , 2008
Teke	–	Central Africa	Gabon	sedentary farmers	–	54	–	Quintana-Murci <i>et al.</i> , 2008
Tsogho	–	Central Africa	Gabon	sedentary farmers	–	64	–	Quintana-Murci <i>et al.</i> , 2008
Yoruba	–	Central Africa	Nigeria	sedentary farmers	31	–	Patin <i>et al.</i> , 2009	–
Aka	–	Central Africa	CAR	nomadic HG	24	–	Patin <i>et al.</i> , 2009	–
C. Baka	–	Central Africa	Cameroon	nomadic HG	–	30	–	Quintana-Murci <i>et al.</i> , 2008
G. Baka	–	Central Africa	Gabon	nomadic HG	16	39	Patin <i>et al.</i> , 2009	Quintana-Murci <i>et al.</i> , 2008
S. Baka	–	Central Africa	Cameroon	nomadic HG	15	58	Patin <i>et al.</i> , 2009	Quintana-Murci <i>et al.</i> , 2008
Bezan	–	Central Africa	Cameroon	nomadic HG	–	35	–	Quintana-Murci <i>et al.</i> , 2008

Bongo	–	Central Africa	Gabon	nomadic HG	–	37	–	Quintana-Murci <i>et al.</i> , 2008
Kola	–	Central Africa	Cameroon	nomadic HG	16	88	Patin <i>et al.</i> , 2009	Quintana-Murci <i>et al.</i> , 2008 Quintana-Murci <i>et al.</i> , 2008
Koya	–	Central Africa	Gabon	nomadic HG	–	31	–	Quintana-Murci <i>et al.</i> , 2008
Mbuti	–	Central Africa	DRC	nomadic HG	24	–	Patin <i>et al.</i> , 2009	–
Bulgarians	–	Eurasia	Bulgaria	sedentary farmers	–	141	–	Richards <i>et al.</i> , 2000
Danes	–	Eurasia	Danemark	sedentary farmers	23	38	Laval <i>et al.</i> , 2010	Richards <i>et al.</i> , 2000
Germans	–	Eurasia	Germany	sedentary farmers	–	174	–	Richards <i>et al.</i> , 2000
Han	–	Eurasia	China	sedentary farmers	24	256	Laval <i>et al.</i> , 2010	Yao <i>et al.</i> , 2002
Japanese	–	Eurasia	Japan	sedentary farmers	24	159	Laval <i>et al.</i> , 2010	Imaizumi <i>et al.</i> , 2002
Koreans	–	Eurasia	Korea	sedentary farmers	–	48	–	Kong <i>et al.</i> , 2003
Liaoning	–	Eurasia	China	sedentary farmers	–	51	–	Yao <i>et al.</i> , 2002
Palestinians	–	Eurasia	Israel	sedentary farmers	–	117	–	Richards <i>et al.</i> , 2000
Pathans	–	Eurasia	Pakistan	sedentary farmers	–	38	–	Quintana-Murci <i>et al.</i> , 2004
Qingdao	–	Eurasia	China	sedentary farmers	–	49	–	Yao <i>et al.</i> , 2002
Turkish	–	Eurasia	Turkey	sedentary farmers	–	218	–	Richards <i>et al.</i> , 2000
Altaians	–	Eurasia	Russia	nomadic herders	–	202	–	Derenko <i>et al.</i> , 2000
Balochi	–	Eurasia	Pakistan	nomadic herders	–	39	–	Quintana-Murci <i>et al.</i> , 2004
Bedouins	–	Eurasia	Saoudi Arabia	nomadic herders	–	29	–	Richards <i>et al.</i> , 2000
Buryats	–	Eurasia	Russia	nomadic herders	–	92	–	Derenko <i>et al.</i> , 2000
Chuvash	–	Eurasia	Russia	nomadic herders	24	55	Laval <i>et al.</i> , 2010	Bermisheva <i>et al.</i> , 2002
Mongols	–	Eurasia	Mongolia	nomadic herders	–	48	–	Kong <i>et al.</i> , 2003

Tajiko-uzbeks (Bukhara)	LUZ	Central Asia	Uzbekistan	sedentary farmers	–	46	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Boukhara)	TAB	Central Asia	Uzbekistan	sedentary farmers	48	24	Ségurel <i>et al.</i> , 2013	This study
Tajiks (Ferghana)	TJK	Central Asia	Uzbekistan / Kyrgyzstan border	sedentary farmers	–	40	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Ferghana)	TJR	Central Asia	Uzbekistan / Kyrgyzstan border	sedentary farmers	–	29	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJE	Central Asia	Tajikistan	sedentary farmers	–	32	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJN	Central Asia	Tajikistan	sedentary farmers	–	34	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJT	Central Asia	Tajikistan	sedentary farmers	–	32	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Penjikent)	TDU	Central Asia	Tajikistan/ Uzbekistan border	sedentary farmers	–	40	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Penjikent)	TDS	Central Asia	Tajikistan/ Uzbekistan border	sedentary farmers	–	31	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Samarkand)	TJA	Central Asia	Uzbekistan / Tajikistan border	sedentary farmers	–	32	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Samarkand)	TJU	Central Asia	Uzbekistan / Tajikistan border	sedentary farmers	–	29	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Tajiks (Yagnobs from Dushanbe)	TJY	Central Asia	Tajikistan	sedentary farmers	–	40	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Karakalpaks (Karakalpakia)	KAR	Central Asia	Uzbekistan	nomadic herders	–	55	–	Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Karakalpaks	OTU	Central Asia	Uzbekistan	nomadic herders	–	53	–	Chaix <i>et al.</i> , 2007;

(Karakalpakia)									Heyer <i>et al.</i> , 2009
Kazaks (Bukhara)	LKZ	Central Asia	Uzbekistan	nomadic herders	–	31	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kazaks (Karakalpakia)	KAZ	Central Asia	Uzbekistan	nomadic herders	–	50	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Andijan)	KRA	Central Asia	Uzbekistan / Kyrgyzstan border	nomadic herders	–	48	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Bichkek)	KIB	Central Asia	Kyrgyzstan	nomadic herders	48	29	Ségurel <i>et al.</i> , 2013		This study
Kyrgyz (Issyk Kul)	KRT	Central Asia	Kyrgyzstan	nomadic herders	–	29	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRB	Central Asia	Kyrgyzstan	nomadic herders	–	30	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRG	Central Asia	Kyrgyzstan	nomadic herders	–	20	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRL	Central Asia	Kyrgyzstan	nomadic herders	–	24	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRM	Central Asia	Kyrgyzstan	nomadic herders	–	26	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Turkmen (Urgench)	TKY	Central Asia	Uzbekistan	nomadic herders	–	35	–		This study
Turkmen (Karakalpakia)	TUR	Central Asia	Uzbekistan	nomadic herders	–	51	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Uzbeks (Ferghana)	UZA	Central Asia	Uzbekistan / Kyrgyzstan border	nomadic herders	–	36	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Uzbeks (Karakalpakia)	UZB	Central Asia	Uzbekistan	nomadic herders	–	40	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009
Uzbeks (Penjikent)	UZT	Central Asia	Tajikistan/ Uzbekistan border	nomadic herders	–	39	–		Chaix <i>et al.</i> , 2007; Heyer <i>et al.</i> , 2009

Table I.B-S2. Summary statistics and neutrality tests computed for the HVS-I sequences.

Population	Area	Life style	S	K	Tajima's <i>D</i>	Fu&Li's <i>D</i>	Fu & Li's <i>F</i>	Fu's <i>FS</i>
Akele	Africa	sedentary farmers	54	33	-0.79	-1.21	-1.26	<u>-14.39*</u>
C. Fang	Africa	sedentary farmers	45	45	-0.51	-1.49	-1.37	<u>-9.46*</u>
G. Fang	Africa	sedentary farmers	55	36	-0.88	-0.14	-0.50	<u>-13.2**</u>
Ewondo	Africa	sedentary farmers	37	12	0.05	0.54	0.46	0.95
Kota	Africa	sedentary farmers	59	32	-0.70	-0.84	-0.94	<u>-8.28*</u>
Ngumba	Africa	sedentary farmers	51	40	-0.45	-0.61	-0.66	<u>-13.99*</u>
Nzebi	Africa	sedentary farmers	64	21	-1.31	-1.24	-1.52	<u>-22.22**</u>
Teke	Africa	sedentary farmers	53	39	-0.91	-1.21	-1.31	<u>-22.11**</u>
Tsogo	Africa	sedentary farmers	56	32	-1.05	-0.14	-0.59	<u>-8.67*</u>
C. Baka	Africa	nomadic HG	20	7	0.26	1.005	0.90	3.34
G. Baka	Africa	nomadic HG	17	7	0.08	-1.51	-1.16	2.76
S. Baka	Africa	nomadic HG	34	12	-0.75	-0.13	-0.43	1.52
Bezan	Africa	nomadic HG	8	4	1.46	-0.0002	0.53	4.38
Bongo	Africa	nomadic HG	23	9	-0.11	-1.60	-1.30	2.06
Kola	Africa	nomadic HG	10	6	2.01*	0.69	1.36	5.05
Koya	Africa	nomadic HG	17	4	-0.99	-2.02	-1.99	4.26
Bulgarians	Eurasia	sedentary farmers	67	83	<u>-2.23*</u>	<u>-3.83*</u>	<u>-3.76**</u>	<u>-129.67**</u>
Danes	Eurasia	sedentary farmers	28	21	<u>-1.91*</u>	<u>-2.17*</u>	<u>-2.47*</u>	<u>-37.24*</u>
Germans	Eurasia	sedentary farmers	73	102	<u>-2.30*</u>	<u>-4.09**</u>	<u>-3.95**</u>	<u>-183.20***</u>
Han	Eurasia	sedentary farmers	121	218	<u>-2.02*</u>	<u>-3.78*</u>	<u>-3.49*</u>	<u>-469.60***</u>
Japanese	Eurasia	sedentary farmers	82	110	<u>-1.79*</u>	<u>-2.87*</u>	<u>-2.87*</u>	<u>-158.51**</u>
Koreans	Eurasia	sedentary farmers	38	28	-1.57	-1.14	-1.54	<u>-17.24*</u>

Liaoning	Eurasia	sedentary farmers	64	48	<u>-2.07*</u>	<u>-3.01*</u>	<u>-3.18*</u>	<u>-61.01**</u>
Palestinians	Eurasia	sedentary farmers	81	95	<u>-2.24*</u>	<u>-3.59**</u>	<u>-3.62**</u>	<u>-161.50***</u>
Pathans	Eurasia	sedentary farmers	46	38	<u>-1.96*</u>	<u>-2.38*</u>	<u>-2.65*</u>	<u>-53.08**</u>
Qingdao	Eurasia	sedentary farmers	55	43	<u>-1.89*</u>	<u>-2.6*</u>	<u>-2.79*</u>	<u>-48.92*</u>
Turks	Eurasia	sedentary farmers	88	144	<u>-2.22*</u>	<u>-2.87*</u>	<u>-3.07*</u>	<u>-282.68***</u>
Altaians	Eurasia	nomadic herders	65	53	-2.13	-1.59	-2.20	<u>-45.72*</u>
Balochi	Eurasia	nomadic herders	37	26	-1.79	-1.64	-2.00	<u>-18.40*</u>
Bedouins	Eurasia	nomadic herders	44	27	-1.54	-2.07	-2.24	<u>-22.68**</u>
Buryats	Eurasia	nomadic herders	62	60	-1.10	-0.50	-0.82	<u>-18.59*</u>
Chuvash	Eurasia	nomadic herders	44	38	-1.71	-1.07	-1.52	<u>-12.80*</u>
Mongols	Eurasia	nomadic herders	51	38	-1.75	-1.16	-1.64	<u>-34.09**</u>
Tajiko-uzbeks (LUZ)	C. Asia	sedentary farmers	43	26	<u>-1.65*</u>	-1.08	-1.53	<u>-11.76**</u>
Tajiks (TAB)	C. Asia	sedentary farmers	31	23	-1.81	-2.02	-2.29	<u>-24.11***</u>
Tajiks (TDS)	C. Asia	sedentary farmers	33	14	-1.06	-0.09	-0.49	-1.61
Tajiks (TDU)	C. Asia	sedentary farmers	32	22	-1.37	-1.38	-1.63	<u>-9.86**</u>
Tajiks (TJA)	C. Asia	sedentary farmers	29	22	-1.54	-1.66	-1.91	<u>-14.19**</u>
Tajiks (TJE)	C. Asia	sedentary farmers	42	26	-1.91	<u>-2.47*</u>	<u>-2.70*</u>	<u>-20.35**</u>
Tajiks (TJK)	C. Asia	sedentary farmers	41	30	-1.83	<u>-2.2*</u>	<u>-2.46*</u>	<u>-24.94**</u>
Tajiks (TJN)	C. Asia	sedentary farmers	41	28	<u>-2.06*</u>	-1.97	<u>-2.37*</u>	<u>-26.06***</u>
Tajiks (TJR)	C. Asia	sedentary farmers	48	24	-1.91	<u>-2.12*</u>	<u>-2.42*</u>	<u>-15.68*</u>
Tajiks (TJT)	C. Asia	sedentary farmers	42	26	-1.91	<u>-2.47*</u>	<u>-2.70*</u>	<u>-20.35***</u>
Tajiks (TJU)	C. Asia	sedentary farmers	39	19	<u>-1.65*</u>	-1.68	-1.97	<u>-7.68*</u>
Tajiks (TJY)	C. Asia	sedentary farmers	22	17	-0.62	-0.22	-0.42	<u>-4.75*</u>
Karakalpaks	C. Asia	nomadic herders	59	48	<u>-2.03*</u>	<u>-2.86*</u>	<u>-3.05*</u>	<u>-59.58***</u>

(KAR)								
Karakalpaks (OTU)	C. Asia	nomadic herders	61	44	-1.95	<u>-3.10*</u>	<u>-3.19*</u>	<u>-34.33**</u>
Kazaks (KAZ)	C. Asia	nomadic herders	54	45	-1.97	<u>-2.22*</u>	<u>-2.54*</u>	<u>-57.01***</u>
Kazaks (LKZ)	C. Asia	nomadic herders	36	24	-1.59	-2.04	-2.23	<u>-16.55*</u>
Kyrgyz (KIB)	C. Asia	nomadic herders	45	26	-1.56	-1,32	-1,65	<u>-19,17**</u>
Kyrgyz (KRA)	C. Asia	nomadic herders	52	32	-1.98	<u>-3.28*</u>	<u>-3.35*</u>	<u>-22.67***</u>
Kyrgyz (KRB)	C. Asia	nomadic herders	40	24	-1.82	<u>-3.01*</u>	<u>-3.09*</u>	<u>-17.26**</u>
Kyrgyz (KRG)	C. Asia	nomadic herders	32	19	-1.35	-1.49	-1.69	<u>-14.03*</u>
Kyrgyz (KRL)	C. Asia	nomadic herders	39	21	-1.73	-1.99	-2.24	<u>-11.76*</u>
Kyrgyz (KRM)	C. Asia	nomadic herders	34	22	-1.28	-0.88	-1.18	<u>-14.46**</u>
Kyrgyz (KRT)	C. Asia	nomadic herders	39	24	-1.73	-2.01	-2.26	<u>-16.92**</u>
Turkmen (TKY)	C. Asia	nomadic herders	46	31	-1.73	-1,94	<u>-2,21*</u>	<u>-27,94***</u>
Turkmen (TUR)	C. Asia	nomadic herders	46	36	-1.63	-1.91	-2.16	<u>-28.73***</u>
Uzbeks (UZA)	C. Asia	nomadic herders	41	26	-1.62	-1.52	-1.84	<u>-16.17**</u>
Uzbeks (UZB)	C. Asia	nomadic herders	54	33	<u>-2.06*</u>	<u>-3.16*</u>	<u>-3.30*</u>	<u>-29.65***</u>
Uzbeks (UZT)	C. Asia	nomadic herders	43	24	-1.77	-2.25	-2.47	<u>-12.28*</u>

S: Number of polymorphisms, *K*: Number of haplotypes. Values significantly higher than expected on a constant population size model are indicated in bold, whereas significantly lower values are underlined. HG= Hunter-gatherers. Significance levels: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ after FDR correction for multiple testing (Benjamini and Hochberg, 1995).

Table I.B-S3. Estimated marginal likelihoods using path-sampling (PS) and stepping-stone (SS) methods (Baele *et al.*, 2012) for model selection.

Genetic marker	Population	Area	Life-style	Constant model (PS)	Constant model (SS)	Exponential model (PS)	Exponential model (SS)	Logistic model (PS)	Logistic model (SS)	Expansion model (PS)	Expansion model (SS)
Autosomes	Aka	Africa	Nomadic HG	-36426	-36426	-36445	-36447	-36429	-36430	-36443	-36445
Autosomes	G. Baka	Africa	Nomadic HG	-35783	-35783	-35788	-35789	-35787	-35787	-37778	-37779
Autosomes	Kola	Africa	Nomadic HG	-35629	-35630	-37782	-37783	-37231	-37231	-36452	-36452
Autosomes	Mbuti	Africa	Nomadic HG	-35326	-35326	-36545	-36546	-36319	-36319	-36478	-36478
Autosomes	S. Baka	Africa	Nomadic HG	-35618	-35619	-35630	-35631	-35629	-35629	-35627	-35628
Autosomes	Akele	Africa	Sedentary farmers	-35184	-35184	-35162	-35163	-35168	-35167	-35179	-35179
Autosomes	Chagga	Africa	Sedentary farmers	-37152	-37153	-37097	-37100	-37973	-37973	-35627	-35628
Autosomes	Mozambicans	Africa	Sedentary farmers	-36927	-36927	-36893	-36893	-36942	-36943	-36748	-36748
Autosomes	Ngumba	Africa	Sedentary farmers	-38562	-38562	-38493	-38493	-38511	-38512	-38538	-38539
Autosomes	Yoruba	Africa	Sedentary farmers	-37472	-37473	-37412	-37413	-37463	-37463	-37459	-37459
Autosomes	Chuvash	Eurasia	Nomadic herders	-33515	-33515	-33728	-33729	-33703	-33703	-33689	-33689
Autosomes	Danes	Eurasia	Sedentary farmers	-33615	-33616	-33001	-33001	-33607	-33609	-33584	-33584
Autosomes	Han Chinese	Eurasia	Sedentary farmers	-33587	-33587	-33551	-33551	-33591	-33591	-33538	-33539
Autosomes	Japanese	Eurasia	Sedentary farmers	-33398	-33398	-33322	-33322	-33354	-33354	-33382	-33384
Autosomes	Kyrgyz (KIB)	Central Asia	Nomadic herders	-35172	-35172	-35204	-35205	-35178	-35178	-35184	-35185
Autosomes	Tajiks (TAB)	Central Asia	Sedentary farmers	-35397	-35397	-35427	-35427	-34517	-34517	-35401	-35401
HVS-I	Bezan	Africa	Nomadic HG	-613	-613	-591	-591	-598	-598	-597	-597
HVS-I	C. Baka	Africa	Nomadic HG	-707	-707	-687	-687	-693	-693	-692	-692
HVS-I	E. Bongo	Africa	Nomadic HG	-715	-715	-695	-695	-701	-701	-704	-704
HVS-I	G. Baka	Africa	Nomadic HG	-687	-687	-672	-672	-678	-678	-677	-677

HVS-I	Kola	Africa	Nomadic HG	-697	-697	-683	-683	-690	-690	-691	-691
HVS-I	Kota	Africa	Nomadic HG	-1322	-1322	-1311	-1311	-1318	-1318	-1316	-1316
HVS-I	Koya	Africa	Nomadic HG	-633	-633	-619	-619	-626	-626	-625	-625
HVS-I	S. Baka	Africa	Nomadic HG	-887	-887	-875	-875	-884	-884	-877	-877
HVS-I	Akele	Africa	Sedentary farmers	-1212	-1212	-1199	-1199	-1206	-1206	-1200	-1200
HVS-I	C. Fang	Africa	Sedentary farmers	-1089	-1089	-1068	-1068	-1071	-1071	-1079	-1079
HVS-I	Ewondo	Africa	Sedentary farmers	-849	-849	-829	-829	-843	-843	-843	-843
HVS-I	G. Fang	Africa	Sedentary farmers	-1312	-1313	-1302	-1302	-1306	-1306	-1304	-1304
HVS-I	Ngumba	Africa	Sedentary farmers	-1396	-1396	-1389	-1389	-1391	-1391	-1391	-1391
HVS-I	Nzebi	Africa	Sedentary farmers	-1459	-1459	-1441	-1441	-1454	-1454	-1443	-1443
HVS-I	Teke	Africa	Sedentary farmers	-1377	-1378	-1357	-1358	-1368	-1368	-1359	-1359
HVS-I	Tsogho	Africa	Sedentary farmers	-1277	-1277	-1270	-1270	-1273	-1273	1271	1271
HVS-I	Altaians	Eurasia	Nomadic herders	-1644	-1644	-1655	-1656	-1646	-1647	-1645	-1646
HVS-I	Balochi	Eurasia	Nomadic herders	-824	-824	-835	-835	-827	-827	-833	-833
HVS-I	Bedouins	Eurasia	Nomadic herders	-857	-8757	-874	-874	-860	-860	-862	-862
HVS-I	Buryats	Eurasia	Nomadic herders	-1339	-1339	-1362	-1363	-1344	-1345	-1346	-1346
HVS-I	Chuvash	Eurasia	Nomadic herders	-1234	-1234	-1250	-1251	-1237	-1237	-1238	-1238
HVS-I	Mongols	Eurasia	Nomadic herders	-1047	-1047	-1065	-1065	-1057	-1057	-1048	-1048
HVS-I	Bulgarians	Eurasia	Sedentary farmers	-1711	-1711	-1664	-1664	-1669	-1669	-1668	-1669
HVS-I	Danes	Eurasia	Sedentary farmers	-729	-729	-700	-700	-718	-719	-726	-726
HVS-I	Germans	Eurasia	Sedentary farmers	-1853	-1853	-1853	-1854	-1913	-1914	-1852	-1852
HVS-I	Han Chinese	Eurasia	Sedentary farmers	-4120	-4121	-3994	-3995	-4006	-4007	-4004	-4005
HVS-I	Japanese	Eurasia	Sedentary farmers	-2356	-2356	-2312	-2312	-2319	-2319	-2319	-2320

HVS-I	Koreans	Eurasia	Sedentary farmers	-891	-891	-906	-906	-893	-893	-899	-900
HVS-I	Liaoning	Eurasia	Sedentary farmers	-1260	-1260	-1223	-1223	-1281	-1281	-1219	-1219
HVS-I	Palestinians	Eurasia	Sedentary farmers	-1889	-1889	-1871	-1871	-1884	-1884	-1884	-1884
HVS-I	Pathans	Eurasia	Sedentary farmers	-942	-942	-932	-932	-938	-938	-937	-937
HVS-I	Qingdao	Eurasia	Sedentary farmers	-1128	-1128	-1123	-1123	-1132	-1133	-1093	-1093
HVS-I	Turkish	Eurasia	Sedentary farmers	-2714	-2715	-2636	-2636	-2648	-2648	-2646	-2646
HVS-I	KAR	Central Asia	Nomadic herders	-1253	-1253	-1227	-1227	-1231	-1231	-1232	-1232
HVS-I	KAZ	Central Asia	Nomadic herders	-1168	-1168	-1139	-1139	-1143	-1143	-1144	-1144
HVS-I	KRA	Central Asia	Nomadic herders	-1102	-1103	-1085	-1085	-1087	-1087	-1095	-1095
HVS-I	KRB	Central Asia	Nomadic herders	-851	-852	-836	-836	-839	-839	-846	-846
HVS-I	KRG	Central Asia	Nomadic herders	-813	-813	-796	-796	-798	-798	-804	-804
HVS-I	KRL	Central Asia	Nomadic herders	-862	-862	-848	-848	-850	-850	-855	-855
HVS-I	KRM	Central Asia	Nomadic herders	-846	-846	-832	-832	-835	-835	-842	-842
HVS-I	KRT	Central Asia	Nomadic herders	-886	-886	-867	-867	-870	-870	-875	-875
HVS-I	Kyrgyz (KIB)	Central Asia	Nomadic herders	-941	-941	-917	-917	-923	-923	-929	-929
HVS-I	LKZ	Central Asia	Nomadic herders	-875	-876	-859	-859	-861	-861	-865	-865
HVS-I	OTU	Central Asia	Nomadic herders	-1262	-1262	-1236	-1236	-1239	-1239	-1244	-1244
HVS-I	TKY	Central Asia	Nomadic herders	-1007	-1007	-986	-986	-1000	-1001	-992	-993
HVS-I	TUR	Central Asia	Nomadic herders	-1099	-1099	-1084	-1084	-1085	-1085	-1092	-1092
HVS-I	UZA	Central Asia	Nomadic herders	-927	-928	-916	-916	-923	-923	-925	-925
HVS-I	UZB	Central Asia	Nomadic herders	-1077	-1077	-1055	-1055	-1061	-1061	-1062	-1062
HVS-I	UZT	Central Asia	Nomadic herders	-914	-914	-900	-900	-904	-904	-909	-910
HVS-I	LUZ	Central Asia	Sedentary farmers	-981	-981	-1009	-1009	-990	-990	-990	-990

HVS-I	Tajiks (TAB)	Central Asia	Sedentary farmers	-798	-798	-780	-780	-792	-792	-787	-787
HVS-I	TDS	Central Asia	Sedentary farmers	-803	-803	-814	-814	-806	-806	-811	-811
HVS-I	TDU	Central Asia	Sedentary farmers	-851	-851	-863	-864	-852	-852	-859	-859
HVS-I	TJA	Central Asia	Sedentary farmers	-801	-801	-812	-812	-805	-805	-809	-809
HVS-I	TJE	Central Asia	Sedentary farmers	-933	-933	-911	-911	-925	-925	-928	-928
HVS-I	TJK	Central Asia	Sedentary farmers	-975	-975	-951	-951	-959	-959	-966	-966
HVS-I	TJN	Central Asia	Sedentary farmers	-888	-888	-867	-867	-893	-893	-879	-879
HVS-I	TJR	Central Asia	Sedentary farmers	-959	-959	-924	-924	-928	-928	-940	-940
HVS-I	TJT	Central Asia	Sedentary farmers	-941	-941	-911	-911	-916	-916	-929	-929
HVS-I	TJU	Central Asia	Sedentary farmers	-848	-848	-856	-856	-851	-851	-850	-850
HVS-I	TJY	Central Asia	Sedentary farmers	-748	-748	-763	-763	-755	-755	-752	-752

The model with the greatest marginal likelihood is considered as the best fitted model.

Table I.B-S4. Mode and 95% HPD of current effective population size (N_0) and per year growth rate (g) estimated under the best-fitting (BF) model for autosomal sequences.

Population	Area	Life-style	BF model	N_0 95% lower	N_0 mode	N_0 95% upper	g 95% lower	g mode	g 95% upper
Akele	Africa	sedentary farmers	exponential	7525/ 15652	9774/ 20331	12892/ 26817	2.15x10⁻⁰⁷/ 4.47x10 ⁻⁰⁷	5.62x10⁻⁰⁷/ 1.17x10 ⁻⁰⁶	1.03x10⁻⁰⁶/ 2.15x10 ⁻⁰⁶
Chagga	Africa	sedentary farmers	expansion	5093/ 10595 (48/ 101)	42750/ 88922 (9861/ 20512)	77746/ 161712 (25376/ 52784)	1.25x10⁻⁰⁵/ 2.59x10 ⁻⁰⁵	1.02x10⁻⁰⁴/ 2.13x10 ⁻⁰⁴	7.2x10⁻⁰⁵/ 3.58x10 ⁻⁰⁴
Mozambicans	Africa	sedentary farmers	expansion	7321/ 15229 (0/ 1)	473549/ 984982 (38618/ 80326)	480768/ 999999 (68728/ 142955)	1.36x10⁻⁰⁵/ 2.83x10 ⁻⁰⁵	6.6x10⁻⁰⁵/ 1.38x10 ⁻⁰⁴	1.55x10⁻⁰⁴/ 3.23x10 ⁻⁰⁴
Ngumba	Africa	sedentary farmers	exponential	6490/ 13501	8295/ 17254	10801/ 22467	6.25x10⁻⁰⁸/ 1.30x10 ⁻⁰⁷	3.43x10⁻⁰⁷/ 7.13x10 ⁻⁰⁷	7.4x10⁻⁰⁷/ 1.53x10 ⁻⁰⁶
Yoruba	Africa	sedentary farmers	exponential	6699/ 13935	8996/ 18713	11759/ 24460	1.68x10⁻⁰⁸/ 3.49x10 ⁻⁰⁸	2.96x10⁻⁰⁷/ 6.16x10 ⁻⁰⁷	0.68x10⁻⁰⁶/ 1.41x10 ⁻⁰⁶
Aka	Africa	nomadic HG	constant	5012/ 10426	5941/ 12358	7025/ 14613	—	—	—
G. Baka	Africa	nomadic HG	constant	4996/ 10392	5971/ 12420	7130/ 14831	—	—	—
S. Baka	Africa	nomadic HG	constant	4617/ 9604	5553/ 11551	6708/ 13953	—	—	—
Kola	Africa	nomadic HG	constant	5392/ 11217	6421/ 13357	7662/ 15938	—	—	—
Mbuti	Africa	nomadic HG	constant	4714/ 9806	5602/ 11653	6687/ 13911	—	—	—
Danes	Eurasia	sedentary farmers	exponential	3135/ 6522	4000/ 8322	5086/ 10579	-3.41x10⁻⁰⁷/ -7.11x10 ⁻⁰⁷	-0.84x10⁻⁰⁷/ -1.74x10 ⁻⁰⁷	2.40x10⁻⁰⁷/ 5.00x10 ⁻⁰⁷
Han	Eurasia	sedentary farmers	expansion	5134/ 10679 (116/ 242)	17127/ 35625 (5335/ 11097)	38751/ 80603 (15778/ 32819)	3.15x10⁻⁰⁵/ 6.56x10 ⁻⁰⁵	1.35x10⁻⁰⁴/ 2.81x10 ⁻⁰⁴	2.28x10⁻⁰⁴/ 4.75x10 ⁻⁰⁴
Japanese	Eurasia	sedentary farmers	exponential	2313/ 4813	2994/ 6229	3918/ 8150	-5x10⁻⁰⁷/ -1.04x10 ⁻⁰⁶	-1.85x10⁻⁰⁷/ -3.84x10 ⁻⁰⁷	4.08x10⁻⁰⁷/ 4.08x10 ⁻⁰⁷
Chuvash	Eurasia	nomadic herders	constant	4125 /8250	4904/ 9808	5472/ 11382	—	—	—
Tajiks (TAB)	C. Asia	sedentary farmers	constant	3959/ 8236	4746/ 9873	5826/ 11653	—	—	—
Kyrgyz (KIB)	C. Asia	nomadic herders	constant	3578/ 7443	4286/ 8916	5111/ 10631	—	—	—

^aFor the cases where the "Expansion model" was selected, the estimated ancestral effective population size (N_1) is indicated in parentheses. Values estimated assuming a mutation rate of 1.2×10^{-8} /generation /site (Conrad et al., 2011) are indicated in bold. The others are estimated assuming a mutation rate of 2.5×10^{-8} /generation /site (Pluzhnikov et al., 2002). ^bHG= hunter-gatherers.

Table I.B-S5. Mode and HPD intervals of the number of demographic changes inferred with EBSP.

Population	Area	Life style	95% Lower (autosomes)	Mode (autosomes)	95% upper (autosomes)	95% Lower (HVS-I)	Mode (HVS-I)	95% upper (HVS-I)
Akele	Africa	Sedentary farmers	1	2	3	1	1	3
Chagga	Africa	Sedentary farmers	1	2	3	–	–	–
C. Fang	Africa	Sedentary farmers	–	–	–	1	1	3
G.Fang	Africa	Sedentary farmers	–	–	–	1	2	4
Ewondo	Africa	Sedentary farmers	–	–	–	1	2	4
Mozambicans	Africa	Sedentary farmers	2	2	4			
Ngumba	Africa	Sedentary farmers	1	2	3	1	2	4
Nzebi	Africa	Sedentary farmers	–	–	–	1	1	3
Teke	Africa	Sedentary farmers	–	–	–	1	1	3
Tsogo	Africa	Sedentary farmers	–	–	–	1	2	4
Yoruba	Africa	Sedentary farmers	1	1	2	–	–	–
Aka	Africa	Nomadic HG	0	0	3	–	–	–
Bezan	Africa	Nomadic HG	–	–	–	0	0	3
C. Baka	Africa	Nomadic HG	–	–	–	0	0	3
E.Bongo	Africa	Nomadic HG	–	–	–	0	0	3
G. Baka	Africa	Nomadic HG	0	0	3	0	0	3
Kola	Africa	Nomadic HG	0	0	3	–	–	–
Kota	Africa	Nomadic HG	–	–	–	0	1	3
Koya	Africa	Nomadic HG	–	–	–	0	2	3
S. Baka	Africa	Nomadic HG	0	0	3	0	1	3

Mbuti	Africa	Nomadic HG	0	0	3	–	–	–
Danes	Eurasia	Sedentary farmers	1	1	2	1	1	3
Germans	Eurasia	Sedentary farmers	–	–	–	1	2	3
Han Chinese	Eurasia	Sedentary farmers	1	2	3	2	2	3
Japanese	Eurasia	Sedentary farmers	1	1	2	1	2	4
Koreans	Eurasia	Sedentary farmers	–	–	–	0	1	3
Liaoning	Eurasia	Sedentary farmers	–	–	–	1	1	3
Mongols	Eurasia	Sedentary farmers	–	–	–	1	1	3
Palestinians	Eurasia	Sedentary farmers	–	–	–	1	1	3
Pathans	Eurasia	Sedentary farmers	–	–	–	1	1	3
Qingdao	Eurasia	Sedentary farmers	–	–	–	1	1	3
Turkish	Eurasia	Sedentary farmers	–	–	–	1	2	3
Altaians	Eurasia	Nomadic herders	–	–	–	2	3	4
Balochi	Eurasia	Nomadic herders	–	–	–	1	1	3
Bedouins	Eurasia	Nomadic herders	–	–	–	1	1	3
Bulgarians	Eurasia	Nomadic herders	–	–	–	1	2	3
Buryats	Eurasia	Nomadic herders	–	–	–	1	1	3
Chuvash	Eurasia	Nomadic herders	0	1	2	1	1	3
Tajiks (TAB)	C. Asia	Sedentary farmers	0	1	2	1	1	2
LUZ	C. Asia	Sedentary farmers	–	–	–	1	2	4
TAB	C. Asia	Sedentary farmers	–	–	–	1	1	2
TDS	C. Asia	Sedentary farmers	–	–	–	2	3	4

TDU	C. Asia	Sedentary farmers	-	-	-	1	2	4
TJA	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJE	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJK	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJN	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJR	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJT	C. Asia	Sedentary farmers	-	-	-	1	1	3
TJU	C. Asia	Sedentary farmers	-	-	-	1	2	4
TJY	C. Asia	Nomadic herders	-	-	-	0	0	2
Kyrgyz (KIB)	C. Asia	Nomadic herders	0	1	2	1	1	3
KAR	C. Asia	Nomadic herders	-	-	-	1	1	3
KAZ	C. Asia	Nomadic herders	-	-	-	1	1	3
KRA	C. Asia	Nomadic herders	-	-	-	1	1	3
KRB	C. Asia	Nomadic herders	-	-	-	1	1	3
KRG	C. Asia	Nomadic herders	-	-	-	1	1	2
KRL	C. Asia	Nomadic herders	-	-	-	1	1	3
KRM	C. Asia	Nomadic herders	-	-	-	1	1	3
KRT	C. Asia	Nomadic herders	-	-	-	1	1	3
LKZ	C. Asia	Nomadic herders	-	-	-	1	1	3
OTU	C. Asia	Nomadic herders	-	-	-	1	1	3
TKY	C. Asia	Nomadic herders	-	-	-	1	1	3
TUR	C. Asia	Nomadic herders	-	-	-	1	1	3

UZA	C. Asia	Nomadic herders	–	–	–	1	1	3
UZB	C. Asia	Nomadic herders	–	–	–	1	1	3
UZT	C. Asia	Nomadic herders	–	–	–	1	2	4

This indicates how many significant changes in N_e occurred during the period of time covered by the EBSP graph. A constant population size can therefore be rejected since the HPD interval of the number of change points exclude 0 but contained 1 (Heled and Drummond, 2010).

Table I.B-S6. Inferred dating for the expansion onsets using EBSP.

Population	Area	Life style	Autosomal data	HVS-I data
Akele	Africa	Sedentary farmers	33,627/ 67,254	45,319/ 90,638
Chagga	Africa	Sedentary farmers	10,333/ 20,666	40,996/ 81,992
C. Fang	Africa	Sedentary farmers	—	40,996/ 81,992
G. Fang	Africa	Sedentary farmers	—	42,400/ 84,800
Ewondo	Africa	Sedentary farmers	—	41,401/ 82,802
Mozambicans	Africa	Sedentary farmers	62,275/ 124,550	—
Ngumba	Africa	Sedentary farmers	9,450/ 18,900	39,024/ 78,048
Nzebi	Africa	Sedentary farmers	—	31,350/ 62,700
Teke	Africa	Sedentary farmers	—	38,615/ 77,230
Tsogo	Africa	Sedentary farmers	—	35,745/ 71,490
Yoruba	Africa	Sedentary farmers	7,975/ 15,950	—
Bulgarians	Eurasia	Sedentary farmers	—	13,680/ 27,360
Danes	Eurasia	Sedentary farmers	—	10,476/ 20,952
Germans	Eurasia	Sedentary farmers	—	13,287/ 26,574
Han Chinese	Eurasia	Sedentary farmers	36,025/ 72,050	21,549/ 43,098
Japanese	Eurasia	Sedentary farmers	—	20,197/ 40,394
Liaoning	Eurasia	Sedentary farmers	—	26,830/ 53,660
Palestinians	Eurasia	Sedentary farmers	—	14,297/ 28,594
Pathans	Eurasia	Sedentary farmers	—	15,922/ 31,844
Qingdao	Eurasia	Sedentary farmers	—	21,150/ 42,300
Turkish	Eurasia	Sedentary farmers	—	14,531/ 29,062
Altaians	Eurasia	Nomadic herders	—	17,804/ 35,608
Balochi	Eurasia	Nomadic herders	—	17,292/ 34,584
Bedouins	Eurasia	Nomadic herders	—	14,788/ 29,576
Buryats	Eurasia	Nomadic herders	—	15,233/ 30,466
Chuvash	Eurasia	Nomadic herders	—	12,521/ 25,042
Mongols	Eurasia	Nomadic herders	—	18,667/ 37,334
Tajiks (TAB)	Central Asia	Sedentary farmers	—	12,191/ 24,382
LUZ	Central Asia	Sedentary farmers	—	14,360/ 28,720
TAB	Central Asia	Sedentary farmers	—	12,191/ 24,382
TDS	Central Asia	Sedentary farmers	—	16,090/ 32,180

TDU	Central Asia	Sedentary farmers	–	19,043/ 38,086
TJA	Central Asia	Sedentary farmers	–	11,031/ 22,062
TJE	Central Asia	Sedentary farmers	–	13,394/ 26,788
TJK	Central Asia	Sedentary farmers	–	12,233/ 24,466
TJN	Central Asia	Sedentary farmers	–	10,322/ 20,644
TJR	Central Asia	Sedentary farmers	–	15,897/ 31,794
TJT	Central Asia	Sedentary farmers	–	12,749/ 25,498
TJU	Central Asia	Sedentary farmers	–	14,957/ 29,914
TJY	Central Asia	Sedentary farmers	–	14,063/ 28,126
Kyrgyz (KIB)	Central Asia	Nomadic herders	–	19,562/ 39,124
KAR	Central Asia	Nomadic herders	–	15,397/ 30,794
KAZ	Central Asia	Nomadic herders	–	14,798/ 29,596
KIB	Central Asia	Nomadic herders	–	19,562/ 39,124
KRA	Central Asia	Nomadic herders	–	15,170/ 30,340
KRB	Central Asia	Nomadic herders	–	14,733/ 29,466
KRG	Central Asia	Nomadic herders	–	19,675/ 39,350
KRL	Central Asia	Nomadic herders	–	19,207/ 38,414
KRM	Central Asia	Nomadic herders	–	14,999/ 29,998
KRT	Central Asia	Nomadic herders	–	19,293/ 38,586
LKZ	Central Asia	Nomadic herders	–	19,385/ 38,770
OTU	Central Asia	Nomadic herders	–	17,521/ 35,042
TKY	Central Asia	Nomadic herders	–	18,757/ 37,514
TUR	Central Asia	Nomadic herders	–	15,179/ 30,358
UZA	Central Asia	Nomadic herders	–	14,025/ 28,050
UZB	Central Asia	Nomadic herders	–	14,343/ 28,686
UZT	Central Asia	Nomadic herders	–	12,697/ 25,394

We used the population growth curves generated from BEAST to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 – 20 generation intervals. We considered that the population increased when both the median and HPD values for N_e increased gradually between more than 3 successive data points. Although this estimation method does not allow providing a confidence interval for inferred expansion timings, this conservative approach ensure that we consider only strong and significant expansion signals. Values estimated assuming a mutation rate of 1.2×10^{-8} /generation /site for autosomes (Conrad et al., 2011) and 5.10^{-6} /generation/site for HVS-I (Forster et al., 1996) are indicated in bold. The others are estimated assuming a mutation rate of 2.5×10^{-8} /generation /site for autosomes (Pluzhnikov et al., 2002) and 10^{-5} /generation/site for HVS-I (Howell et al., 1996; Heyer et al., 2001).

Table I.B-S7. Modes and 95% HPD of current effective population size (N_0) and per year growth rate (g) estimated under the best-fitting (BF) model for HVS-I sequences.

Population	Area	Life-style	BF model	N_0 95% lower	N_0 mode	N_0 95% upper	g 95% lower	g mode	g 95% upper
Akele	Africa	sedentary farmers	exponential	5523 / 11047	9788 / 19576	17512 / 35024	1.49x10 ⁻⁰⁵ / 2.98x10 ⁻⁰⁵	3.15x10 ⁻⁰⁵ / 6.29x10 ⁻⁰⁵	5.15x10 ⁻⁰⁵ / 1.03x10 ⁻⁰⁴
Ewondo	Africa	sedentary farmers	exponential	1484 / 2968	3296/ 6592	6370/ 12740	6.29x10 ⁻⁰⁶ / 1.2x10 ⁻⁰⁵	3.15x10 ⁻⁰⁵ / 6.30x10 ⁻⁰⁵	7.55x10 ⁻⁰⁵ / 1.51x10 ⁻⁰⁴
C. Fang	Africa	sedentary farmers	exponential	4725 / 9451	10303 / 20606	19903 / 39806	1.87x10 ⁻⁰⁵ / 3.74x10 ⁻⁰⁵	4.04x10 ⁻⁰⁵ / 8.07x10 ⁻⁰⁵	7.00x10 ⁻⁰⁵ / 1.40x10 ⁻⁰⁴
G. Fang	Africa	sedentary farmers	exponential	4624 / 9249	7228 / 14456	11623 / 23246	1.28x10 ⁻⁰⁵ / 2.56x10 ⁻⁰⁵	2.90x10 ⁻⁰⁵ / 5.80x10 ⁻⁰⁵	4.86x10 ⁻⁰⁵ / 9.72x10 ⁻⁰⁵
Kota	Africa	sedentary farmers	exponential	5985 / 11970	9695 / 19390	15678 / 31356	1.29x10 ⁻⁰⁵ / 2.58x10 ⁻⁰⁵	2.71x10 ⁻⁰⁵ / 5.42x10 ⁻⁰⁵	4.38x10 ⁻⁰⁵ / 8.76x10 ⁻⁰⁵
Ngumba	Africa	sedentary farmers	exponential	4672 / 9344	7065 / 14130	10794 / 21588	1.19x10 ⁻⁰⁵ / 2.38x10 ⁻⁰⁵	2.68x10 ⁻⁰⁵ / 5.35x10 ⁻⁰⁵	4.51x10 ⁻⁰⁵ / 9.01x10 ⁻⁰⁵
Nzebi	Africa	sedentary farmers	exponential	9635 / 19271	16876 / 33752	28267 / 56534	2.60x10 ⁻⁰⁵ / 5.19x10 ⁻⁰⁵	4.42x10 ⁻⁰⁵ / 8.84x10 ⁻⁰⁵	6.60x10 ⁻⁰⁵ / 1.32x10 ⁻⁰⁴
Teke	Africa	sedentary farmers	exponential	8873 / 17747	15462 / 30924	28068 / 56136	2.39x10 ⁻⁰⁵ / 4.77x10 ⁻⁰⁵	4.17x10 ⁻⁰⁵ / 8.34x10 ⁻⁰⁵	6.35x10 ⁻⁰⁵ / 1.27x10 ⁻⁰⁴
Tsogo	Africa	sedentary farmers	exponential	4244 / 8488	6682 / 13364	10450 / 20900	1.02x10 ⁻⁰⁵ / 2.03x10 ⁻⁰⁵	2.60x10 ⁻⁰⁵ / 5.20x10 ⁻⁰⁵	4.45x10 ⁻⁰⁵ / 8.90x10 ⁻⁰⁵
C. Baka	Africa	nomadic HG	exponential	269 / 538	576 / 1152	1143 / 1286	-7.2x10 ⁻⁰⁵ / -1.44x10 ⁻⁰⁴	-1.04x10 ⁻⁰⁵ / -2.08x10 ⁻⁰⁵	4.34x10 ⁻⁰⁵ / 8.68x10 ⁻⁰⁵
G. Baka	Africa	nomadic HG	exponential	187 / 374	403 / 806	805 / 1610	-9.15x10 ⁻⁰⁵ / -1.83x10 ⁻⁰⁴	-2.10x10 ⁻⁰⁵ / -4.2x10 ⁻⁰⁵	3.58x10 ⁻⁰⁵ / 7.16x10 ⁻⁰⁵
S. Baka	Africa	nomadic HG	exponential	686 / 1372	1140 / 2280	1898 / 3796	-2.37x10 ⁻⁰⁵ /	-3.42x10 ⁻⁰⁶ /	-1.60x10 ⁻⁰⁷ /

Bezan	Africa	nomadic HG	exponential	58 / 116	226 / 452	519 / 1038	-4.75x10 ⁻⁰⁵ -2.26x10 ⁻⁰⁴ / -4.52x10 ⁻⁰⁴	-6.84x10 ⁻⁰⁶ -3.23x10 ⁻⁰⁵ / -6.46x10 ⁻⁰⁵	-3.20x10 ⁻⁰⁷ -1.76x10 ⁻⁰⁴ / -3.52x10 ⁻⁰⁴
Bongo	Africa	nomadic HG	exponential	296 / 592	576 / 1152	1143 / 5586	-6.87x10 ⁻⁰⁵ / -1.37x10 ⁻⁰⁴	-9.14x10 ⁻⁰⁶ / -1.82x10 ⁻⁰⁵	-4.45x10 ⁻⁰⁵ / -8.9x10 ⁻⁰⁵
Kola	Africa	nomadic HG	exponential	83 / 166	217 / 434	448 / 896	-2.17x10 ⁻⁰⁴ / -4.34x10 ⁻⁰⁴	-2.6x10 ⁻⁰⁵ / -5.2x10 ⁻⁰⁵	1.80x10 ⁻⁰⁴ / 3.6x10 ⁻⁰⁴
Koya	Africa	nomadic HG	exponential	213 / 426	278 / 556	603 / 1206	-1.23x10 ⁻⁰⁴ / -2.46x10 ⁻⁰⁴	-3.63x10 ⁻⁰⁵ / -7.26x10 ⁻⁰⁵	2.66x10 ⁻⁰⁵ / 5.32x10 ⁻⁰⁵
Bulgarians	Eurasia	sedentary farmers	exponential	12500 / 25000	138827 / 277654	384155 / 768310	2.71x10 ⁻⁰⁴ / 3.42x10 ⁻⁰⁴	4.22x10 ⁻⁰⁴ / 8.44x10 ⁻⁰⁴	6.34x10 ⁻⁰⁴ / 1.26x10 ⁻⁰³
Danes	Eurasia	sedentary farmers	exponential	1100 / 2200	13990 / 27981	282836 / 565672	1.8x10 ⁻⁴ / 3.6x10 ⁻⁴	3.41x10 ⁻⁰⁴ / 6.82x10 ⁻⁰⁴	7.69x10 ⁻⁰⁴ / 1.54x10 ⁻⁰³
Germans	Eurasia	sedentary farmers	expansion	1106 (0) / 2212 (0)	71675 (13875) / 34335 (27750)	86542 (59071) / 173084 (118142)	2.00x10 ⁻⁰⁴ / 3.99x10 ⁻⁰⁴	3.41x10 ⁻⁰⁴ / 6.81x10 ⁻⁰⁴	6.00x10 ⁻⁰⁴ / 1.20x10 ⁻⁰³
Han	Eurasia	sedentary farmers	exponential	50878 / 101756	119967 / 239934	286200 / 572400	9.45x10 ⁻⁰⁵ / 1.89x10 ⁻⁰⁴	1.17x10 ⁻⁰⁴ / 2.34x10 ⁻⁰⁴	1.45x10 ⁻⁰⁴ / 2.90x10 ⁻⁰⁴
Japanese	Eurasia	sedentary farmers	exponential	36605 / 73210	58511 / 117022	98333 / 196666	5.50x10 ⁻⁰⁵ / 1.10x10 ⁻⁰⁴	7.60x10 ⁻⁰⁵ / 1.52x10 ⁻⁰⁴	1.06x10 ⁻⁰⁴ / 2.12x10 ⁻⁰⁴
Koreans	Eurasia	sedentary farmers	constant	2553 / 5106	13707 / 27414	25289 / 50578	—	—	—
Liaoning	Eurasia	sedentary farmers	expansion	21000 (0) / 42000 (0)	58938 (560) / 117876 (1120)	94851 (9351) / 189703 (18702)	1.16x10 ⁻⁰⁴ / 2.31x10 ⁻⁰⁴	1.80x10 ⁻⁰⁴ / 3.60x10 ⁻⁰⁴	2.80x10 ⁻⁰⁴ / 5.60x10 ⁻⁰⁴
Palestinians	Eurasia	sedentary farmers	exponential	5593 / 11186	157364 / 314728	365855 / 731710	1.16x10 ⁻⁰⁴ / 2.31x10 ⁻⁰⁴	1.64x10 ⁻⁰⁴ / 3.28x10 ⁻⁰⁴	2.31x10 ⁻⁰⁴ / 4.61x10 ⁻⁰⁴
Pathans	Eurasia	sedentary farmers	exponential	16932 / 33865	53420 / 106841	143191 / 286383	3.72x10 ⁻⁰⁴ / 7.44x10 ⁻⁰⁴	5.54x10 ⁻⁰⁴ / 1.10x10 ⁻⁰³	7.21x10 ⁻⁰⁴ / 1.44x10 ⁻⁰³

Qingdao	Eurasia	sedentary farmers	expansion	19000 (0) / 38000 (0)	23064 (630) / 46129 (1260)	93214 (5373) / 186429 (10746)	8.20x10⁻⁰⁵ / 1.64x10 ⁻⁰⁴	1.40x10⁻⁰⁴ / 2.79x10 ⁻⁰⁴	2.34x10⁻⁰⁴ / 4.67x10 ⁻⁰⁴
Turks	Eurasia	sedentary farmers	exponential	45104 / 70208	86393 / 172781	194272 / 388545	1.59x10⁻⁰⁴ / 3.17x10 ⁻⁰⁴	2.21x10⁻⁰⁴ / 4.42x10 ⁻⁰⁴	3.04x10⁻⁰⁴ / 6.08x10 ⁻⁰⁴
Altaians	Eurasia	nomadic herders	constant	412 / 824	519 / 1038	651 / 1302	—	—	—
Balochi	Eurasia	nomadic herders	constant	1261 / 2522	6860 / 13720	17815 / 35630	—	—	—
Bedouins	Eurasia	nomadic herders	constant	5793 / 11586	9157 / 18304	20123 / 40246	—	—	—
Buryats	Eurasia	nomadic herders	constant	513 / 1027	3671 / 7342	6876 / 13753	—	—	—
Chuvash	Eurasia	nomadic herders	constant	3862 / 7724	5136 / 10272	6826 / 13652	—	—	—
Mongols	Eurasia	nomadic herders	constant	4736 / 9472	16526 / 33052	25160 / 50320	—	—	—
Tajiko-uzbeks (LUZ)	C. Asia	sedentary farmers	constant	2363 / 4726	3395 / 6790	4920 / 9840	—	—	—
Tajiks (TAB)	C. Asia	sedentary farmers	exponential	4761 / 9522	20300 / 40600	79343 / 158687	2.39x10⁻⁰⁴ / 4.77x10 ⁻⁰⁴	3.77x10⁻⁰⁴ / 7.54x10 ⁻⁰⁴	5.32x10⁻⁰⁴ / 1.06x10 ⁻⁰³
Tajiks (TDS)	C. Asia	sedentary farmers	constant	1373 / 2746	2139 / 4278	3373 / 6746	—	—	—
Tajiks (TDU)	C. Asia	sedentary farmers	constant	1519 / 3038	2329 / 4658	3460 / 6920	—	—	—
Tajiks (TJA)	C. Asia	sedentary farmers	constant	1603 / 3206	2525 / 5050	3868 / 7736	—	—	—
Tajiks (TJE)	C. Asia	sedentary farmers	exponential	1502 / 3004	5386 / 10771	21742 / 43484	9.58x10⁻⁰⁵ / 1.92x10 ⁻⁰⁴	1.93x10⁻⁰⁴ / 3.85x10 ⁻⁰⁴	3.63x10⁻⁰⁴ / 7.26x10 ⁻⁰⁴
Tajiks (TJK)	C. Asia	sedentary farmers	exponential	3556 / 7111	26111 / 52222	159450 / 318900	6.76x10⁻⁰⁵ / 1.35x10 ⁻⁰⁴	1.29x10⁻⁰⁵ / 2.57x10 ⁻⁰⁴	2.32x10⁻⁰⁴ / 4.64x10 ⁻⁰⁴
Tajiks (TJN)	C. Asia	sedentary farmers	exponential	1300 / 2600	14127 / 28254	64918 / 129835	1.06x10⁻⁰⁴ / 2.13x10 ⁻⁰⁴	2.17x10⁻⁰⁴ / 4.35x10 ⁻⁰⁴	4.31x10⁻⁰⁴ / 8.61x10 ⁻⁰⁴
Tajiks (TJR)	C. Asia	sedentary farmers	exponential	1800 / 3600	17803 / 35605	117085 / 234171	7.02x10⁻⁰⁵ / 1.40x10 ⁻⁰⁴	1.34x10⁻⁰⁴ / 2.68x10 ⁻⁰⁴	2.36x10⁻⁰⁴ / 4.72x10 ⁻⁰⁴

Tajiks (TJT)	C. Asia	sedentary farmers	exponential	1500 / 3000	20757 / 41515	69495 / 138991	9.74x10⁻⁰⁵ / 1.95x10 ⁻⁰⁴	1.97x10⁻⁰⁵ / 3.95x10 ⁻⁰⁴	3.69x10⁻⁰⁴ / 7.37x10 ⁻⁰⁴
Tajiks (TJU)	C. Asia	sedentary farmers	constant	2135 / 4270	3300 / 6600	5065 / 10130	–	–	–
Tajiks (TJY)	C. Asia	sedentary farmers	constant	654 / 1308	3828 / 7656	32354 / 64708	–	–	–
Karakalpaks (KAR)	C. Asia	nomadic herders	exponential	100 / 200	68103 / 136205	103356 / 206710	1.43x10⁻⁰⁴ / 2.86x10 ⁻⁰⁴	2.31x10⁻⁰⁴ / 4.62x10 ⁻⁰⁴	3.71x10⁻⁰⁴ / 7.42x10 ⁻⁰⁴
Karakalpaks (OTU)	C. Asia	nomadic herders	exponential	1800 / 3600	28088 / 56176	105320 / 210640	8.48x10⁻⁰⁵ / 1.70x10 ⁻⁰⁴	1.59x10⁻⁰⁴ / 3.18x10 ⁻⁰⁴	2.74x10⁻⁰⁴ / 5.49x10 ⁻⁰³
Kazaks (KAZ)	C. Asia	nomadic herders	exponential	1500 / 3000	91948 / 183896	126038 / 252076	1.40x10⁻⁰⁴ / 2.80x10 ⁻⁰⁴	2.34x10⁻⁰⁴ / 4.68x10 ⁻⁰⁴	3.79x10⁻⁰⁴ / 7.59x10 ⁻⁰⁴
Kazaks (LKZ)	C. Asia	nomadic herders	exponential	1500 / 3000	15589 / 31178	78712 / 157424	9.24x10⁻⁰⁵ / 1.85x10 ⁻⁰⁴	2.33x10⁻⁰⁴ / 4.65x10 ⁻⁰⁴	4.53x10⁻⁰⁴ / 9.05x10 ⁻⁰⁴
Kyrgyz (KIB)	C. Asia	nomadic herders	exponential	5634 / 11267	24319 / 48639	54504 / 109008	6.45x10⁻⁰⁵ / 1.29x10 ⁻⁰⁴	1.14x10⁻⁰⁴ / 2.27x10 ⁻⁰⁴	1.93x10⁻⁰⁴ / 3.86x10 ⁻⁰⁴
Kyrgyz (KRA)	C. Asia	nomadic herders	exponential	1500 / 3000	41441 / 24044	132673 / 120556	6.18x10⁻⁰⁵ / 1.24x10 ⁻⁰⁴	1.21x10⁻⁰⁴ / 2.42x10 ⁻⁰⁴	2.13x10⁻⁰⁴ / 4.27x10 ⁻⁰⁴
Kyrgyz (KRB)	C. Asia	nomadic herders	exponential	1512 / 3024	23570 / 47139	134962 / 269923	9.25x10⁻⁰⁵ / 1.85x10 ⁻⁰⁴	2.06x10⁻⁰⁴ / 4.11x10 ⁻⁰⁴	4.11x10⁻⁰⁴ / 8.22x10 ⁻⁰⁴
Kyrgyz (KRG)	C. Asia	nomadic herders	exponential	1700 / 3400	22833 / 45666	68346 / 136692	1.59x10⁻⁰⁴ / 3.18x10 ⁻⁰⁴	2.59x10⁻⁰⁴ / 5.18x10 ⁻⁰⁴	3.74x10⁻⁰⁴ / 7.48x10 ⁻⁰⁴
Kyrgyz (KRL)	C. Asia	nomadic herders	exponential	1716 / 3432	23230 / 46460	51887 / 103775	1.34x10⁻⁰⁴ / 2.69x10 ⁻⁰⁴	2.68x10⁻⁰⁵ / 5.36x10 ⁻⁰⁴	3.83x10⁻⁰⁴ / 7.66x10 ⁻⁰⁴
Kyrgyz (KRM)	C. Asia	nomadic herders	exponential	1700 / 3400	14554 / 29108	92462 / 184923	5.60x10⁻⁰⁵ / 1.12x10 ⁻⁰⁴	1.27x10⁻⁰⁴ / 2.55x10 ⁻⁰⁴	2.41x10⁻⁰⁵ / 4.82x10 ⁻⁰⁴
Kyrgyz (KRT)	C. Asia	nomadic herders	exponential	1642 / 3284	19004 / 38007	86568 / 173136	7.84x10⁻⁰⁵ / 1.57x10 ⁻⁰⁴	1.84x10⁻⁰⁴ / 3.66x10 ⁻⁰⁴	3.44x10⁻⁰⁴ / 6.89x10 ⁻⁰⁴
Turkmen (TKY)	C. Asia	nomadic herders	exponential	2468 / 4936	12795 / 25590	73195 / 146390	1.06x10⁻⁰⁵ /	1.89x10⁻⁰⁴ /	3.31x10⁻⁰⁴ /

Turkmen (TUR)	C. Asia	nomadic herders	exponential	6794 / 13588	13834 / 27667	55245 / 110490	2.12x10 ⁻⁰⁴ 5.41x10⁻⁰⁵ / 1.08x10 ⁻⁰⁴	3.78x10 ⁻⁰⁴ 9.65x10⁻⁰⁵ / 1.93x10 ⁻⁰⁴	6.51x10 ⁻⁰⁴ 1.53x10⁻⁰⁴ / 3.06x10 ⁻⁰⁴
Uzbeks (UZA)	C. Asia	nomadic herders	exponential	1614 / 3229	10380 / 20759	57666 / 115331	4.90x10⁻⁰⁵ / 9.81x10 ⁻⁰⁵	9.77x10⁻⁰⁵ / 1.95x10 ⁻⁰⁴	1.68x10⁻⁰⁵ / 3.36x10 ⁻⁰⁴
Uzbeks (UZB)	C. Asia	nomadic herders	exponential	1601 / 3202	22072 / 44144	50340 / 100681	1.22x10⁻⁰⁴ / 2.44x10 ⁻⁰⁴	2.37x10⁻⁰⁴ / 4.75x10 ⁻⁰⁴	3.73x10⁻⁰⁴ / 7.46x10 ⁻⁰⁴
Uzbeks (UZT)	C. Asia	nomadic herders	exponential	257 / 514	1927 / 3855	7496 / 14992	4.09x10⁻⁵ / 8.19x10 ⁻⁵	8.56x10⁻⁵ / 1.71x10 ⁻⁴	1.48x10⁻⁴ / 2.96x10 ⁻⁴

For the cases where the "Expansion model" was selected, the estimated ancestral effective population size (N_1) is indicated in parentheses. Values estimated assuming the transitional changes rate of 5×10^{-6} /generation /site (Forster et al., 1996) are indicated in bold. The others are estimated assuming the pedigree-based mutation rate of 10^{-5} /generation /site (Howell et al., 1996; Heyer et al., 2001).^bHG= Hunter-gatherers.

Table I.B-S8. Average recombination rate for each studied autosomal region.

Marker	Length (in pb)	Chromosome	Position	Rate
R04	1328	1	106787624-106788951	2.54×10^{-09}
R05	1271	2	76163330-76164600	1.19×10^{-09}
R06	1300	2	117090864-117092163	1.12×10^{-09}
R07	1373	3	20665087-20666459	9.94×10^{-10}
R08	1371	3	146671295-146672665	9.25×10^{-10}
R09	1256	4	29547438-29548693	6.75×10^{-09}
R10	1326	4	179693433-179694758	2.23×10^{-09}
R14	1300	6	91757485-91758784	4.50×10^{-09}
R16	1401	7	13052349-13053749	6.15×10^{-09}
R17	1350	7	118479068-118480417	1.47×10^{-11}
R18	1256	8	5139032-5140287	1.08×10^{-08}
R20	1282	8	137174739-137176020	1.07×10^{-09}
R21	1390	9	11567765-11569154	2.43×10^{-06}
R25	1340	11	80750645-80751984	2.83×10^{-08}
R26	1330	11	96959343-96960672	6.23×10^{-09}
R30	1345	14	82997591-82998935	2.24×10^{-09}
R32	1301	15	84958820-84960120	2.05×10^{-10}
R33	1350	16	58084866-58086215	4.15×10^{-09}
R36	1200	18	26002122-26003321	2.02×10^{-09}
R42	1347	20	53069630-53070976	2.90×10^{-09}

These recombination rates are from the HapMap build GRCh37 genetic map (The International HapMap Consortium, 2003).

Table I.B-S9. Population-specific F_{ST} values.

Area	Life-style	Population	Population specific F_{ST}
Africa	Sedentary farmers	Ewondo	0.017
Africa	Sedentary farmers	C.Fang	0.019
Africa	Sedentary farmers	Akele	0.022
Africa	Sedentary farmers	Kota	0.042
Africa	Sedentary farmers	Ngumba	0.048
Africa	Sedentary farmers	Tsogho	0.084
Africa	Sedentary farmers	Nzebi	0.091
Africa	Sedentary farmers	G.Fang	0.097
Africa	Sedentary farmers	Teke	0.104
Africa	Nomadic HG	S.Baka	0.092
Africa	Nomadic HG	C.Baka	0.128
Africa	Nomadic HG	G.Baka	0.145
Africa	Nomadic HG	E.Bongo	0.181
Africa	Nomadic HG	Kola	0.225
Africa	Nomadic HG	Bezan	0.256
Africa	Nomadic HG	Koya	0.317
Eurasia	Sedentary farmers	Bulgarians	0.036
Eurasia	Sedentary farmers	Danes	0.043
Eurasia	Sedentary farmers	Germans	0.030
Eurasia	Sedentary farmers	Han Chinese	0.049
Eurasia	Sedentary farmers	Japanese	0.074
Eurasia	Sedentary farmers	Koreans	0.066
Eurasia	Sedentary farmers	Liaoning	0.025
Eurasia	Sedentary farmers	Palestinians	0.021
Eurasia	Sedentary farmers	Qingdao	0.039
Eurasia	Sedentary farmers	Turkish	0.031
Eurasia	Sedentary farmers	Pathans	0.014
Eurasia	Nomadic herders	Altaians	0.016
Eurasia	Nomadic herders	Balochi	0.017
Eurasia	Nomadic herders	Bedouins	0.034
Eurasia	Nomadic herders	Buryats	0.120

Eurasia	Nomadic herders	Chuvash	0.028
Eurasia	Nomadic herders	Mongols	0.043
Central Asia	Sedentary farmers	TJA	0.043
Central Asia	Sedentary farmers	TJU	0.022
Central Asia	Sedentary farmers	TJK	0.007
Central Asia	Sedentary farmers	TJN	-0.001
Central Asia	Sedentary farmers	TJE	0.022
Central Asia	Sedentary farmers	TJT	0.007
Central Asia	Sedentary farmers	TJR	-0.004
Central Asia	Sedentary farmers	TDS	0.027
Central Asia	Sedentary farmers	TDU	0.020
Central Asia	Sedentary farmers	TJY	0.049
Central Asia	Sedentary farmers	LUZ	0.026
Central Asia	Sedentary farmers	TAB	0.003
Central Asia	Nomadic herders	OTU	0.015
Central Asia	Nomadic herders	KAR	0.003
Central Asia	Nomadic herders	KAZ	-0.0006
Central Asia	Nomadic herders	TUR	0.012
Central Asia	Nomadic herders	KRA	0.005
Central Asia	Nomadic herders	KRG	0.016
Central Asia	Nomadic herders	KRM	0.014
Central Asia	Nomadic herders	LKZ	0.013
Central Asia	Nomadic herders	UZB	0.002
Central Asia	Nomadic herders	UZT	0.016
Central Asia	Nomadic herders	TKY	0.012
Central Asia	Nomadic herders	KIB	0.002
Central Asia	Nomadic herders	KRT	0.014
Central Asia	Nomadic herders	KRL	-0.003
Central Asia	Nomadic herders	KRB	0.001
Central Asia	Nomadic herders	UZA	0.009

We used ARLEQUIN v3.11 (Excoffier et al., 2005) to compute population-specific F_{ST} values for each population within each area (Africa, Eurasia and Central Asia).

Table I.B-S10. Number of immigrants computed from mismatch distributions under a spatially explicit model.

Area	Life-style	Population	M
Africa	Sedentary farmers	Ewondo	13.41
Africa	Sedentary farmers	C.Fang	26.03
Africa	Sedentary farmers	Akele	45.58
Africa	Sedentary farmers	Kota	20.44
Africa	Sedentary farmers	Ngumba	25.67
Africa	Sedentary farmers	Tsogho	19.66
Africa	Sedentary farmers	Nzebi	38.53
Africa	Sedentary farmers	G.Fang	32.67
Africa	Sedentary farmers	Teke	60.61
Africa	Nomadic HG	S.Baka	3.149
Africa	Nomadic HG	C.Baka	2.643
Africa	Nomadic HG	G.Baka	2.548
Africa	Nomadic HG	E.Bongo	1.392
Africa	Nomadic HG	Kola	2.958
Africa	Nomadic HG	Bezan	1.595
Africa	Nomadic HG	Koya	1.176
Eurasia	Sedentary farmers	Bulgarians	93.987
Eurasia	Sedentary farmers	Danes	33.014
Eurasia	Sedentary farmers	Germans	57.067
Eurasia	Sedentary farmers	Han Chinese	NC ⁺
Eurasia	Sedentary farmers	Japanese	82.425
Eurasia	Sedentary farmers	Koreans	135.051
Eurasia	Sedentary farmers	Liaoning	NC ⁺
Eurasia	Sedentary farmers	Palestinians	NC ⁺
Eurasia	Sedentary farmers	Qingdao	NC ⁺
Eurasia	Sedentary farmers	Turkish	3059
Eurasia	Sedentary farmers	Pathans	NC ⁺
Eurasia	Nomadic herders	Altaians	3.402
Eurasia	Nomadic herders	Balochi	93.998
Eurasia	Nomadic herders	Bedouins	359.016

Eurasia	Nomadic herders	Buryats	496.541
Eurasia	Nomadic herders	Chuvash	53.6
Eurasia	Nomadic herders	Mongols	NC ⁺
Central Asia	Sedentary farmers	TJA	161.077
Central Asia	Sedentary farmers	TJU	21.482
Central Asia	Sedentary farmers	TJK	NC ⁺
Central Asia	Sedentary farmers	TJN	NC ⁺
Central Asia	Sedentary farmers	TJE	27.666
Central Asia	Sedentary farmers	TJT	136.194
Central Asia	Sedentary farmers	TJR	273.58
Central Asia	Sedentary farmers	TDS	14.934
Central Asia	Sedentary farmers	TDU	43.272
Central Asia	Sedentary farmers	TJY	42.528
Central Asia	Sedentary farmers	LUZ	25.08
Central Asia	Sedentary farmers	TAB	NC ⁺
Central Asia	Nomadic herders	OTU	1151.657
Central Asia	Nomadic herders	KAR	399.327
Central Asia	Nomadic herders	KAZ	133.693
Central Asia	Nomadic herders	TUR	109.657
Central Asia	Nomadic herders	KRA	30.8
Central Asia	Nomadic herders	KRG	NC ⁺
Central Asia	Nomadic herders	KRM	462.238
Central Asia	Nomadic herders	LKZ	65.905
Central Asia	Nomadic herders	UZB	681.37
Central Asia	Nomadic herders	UZT	63.143
Central Asia	Nomadic herders	TKY	NC ⁺
Central Asia	Nomadic herders	KIB	138.565
Central Asia	Nomadic herders	KRT	131.383
Central Asia	Nomadic herders	KRL	42.979
Central Asia	Nomadic herders	KRB	169.503
Central Asia	Nomadic herders	UZA	63.486

We used ARLEQUIN v3 (Excoffier et al., 2005) to compute M as $M=2Nm$, where $2N$ is the effective population size and m is the migration rate. *NC = the algorithm did not converge.

Table I.B-S11. Inferred dating for the expansion onsets using the upper bound of the higher mutation rate estimate.

Population	Area	Life style	Autosomal data	HVS-I data
Akele	Africa	Sedentary farmers	18,682	16,785
Chagga	Africa	Sedentary farmers	5,741	15,184
C. Fang	Africa	Sedentary farmers	–	15,184
G. Fang	Africa	Sedentary farmers	–	15,704
Ewondo	Africa	Sedentary farmers	–	15,334
Mozambicans	Africa	Sedentary farmers	34,597	–
Ngumba	Africa	Sedentary farmers	5,250	14,453
Nzebi	Africa	Sedentary farmers	–	11,611
Teke	Africa	Sedentary farmers	–	14,302
Tsogo	Africa	Sedentary farmers	–	13,239
Yoruba	Africa	Sedentary farmers	4,431	–
Bulgarians	Eurasia	Sedentary farmers	–	5,067
Danes	Eurasia	Sedentary farmers	–	3,880
Germans	Eurasia	Sedentary farmers	–	4,921
Han Chinese	Eurasia	Sedentary farmers	20,014	7,981
Japanese	Eurasia	Sedentary farmers	–	7,480
Liaoning	Eurasia	Sedentary farmers	–	9,937
Palestinians	Eurasia	Sedentary farmers	–	5,295
Pathans	Eurasia	Sedentary farmers	–	5,897
Qingdao	Eurasia	Sedentary farmers	–	7,833
Turkish	Eurasia	Sedentary farmers	–	5,382
Altaians	Eurasia	Nomadic herders	–	6,594
Balochi	Eurasia	Nomadic herders	–	6,404
Bedouins	Eurasia	Nomadic herders	–	5,477
Buryats	Eurasia	Nomadic herders	–	5,642
Chuvash	Eurasia	Nomadic herders	–	4,637
Mongols	Eurasia	Nomadic herders	–	6,914
Tajiks (TAB)	Central Asia	Sedentary farmers	–	4,515
LUZ	Central Asia	Sedentary farmers	–	5,319
TDS	Central Asia	Sedentary farmers	–	5,959

TDU	Central Asia	Sedentary farmers	–	7,053
TJA	Central Asia	Sedentary farmers	–	4,086
TJE	Central Asia	Sedentary farmers	–	4,961
TJK	Central Asia	Sedentary farmers	–	4,531
TJN	Central Asia	Sedentary farmers	–	3,823
TJR	Central Asia	Sedentary farmers	–	5,888
TJT	Central Asia	Sedentary farmers	–	4,722
TJU	Central Asia	Sedentary farmers	–	5,540
TJY	Central Asia	Sedentary farmers	–	5,209
Kyrgyz (KIB)	Central Asia	Nomadic herders	–	7,245
KAR	Central Asia	Nomadic herders	–	5,703
KAZ	Central Asia	Nomadic herders	–	5,481
KRA	Central Asia	Nomadic herders	–	5,619
KRB	Central Asia	Nomadic herders	–	5,457
KRG	Central Asia	Nomadic herders	–	7,287
KRL	Central Asia	Nomadic herders	–	7,114
KRM	Central Asia	Nomadic herders	–	5,555
KRT	Central Asia	Nomadic herders	–	7,146
LKZ	Central Asia	Nomadic herders	–	7,180
OTU	Central Asia	Nomadic herders	–	6,489
TKY	Central Asia	Nomadic herders	–	6,947
TUR	Central Asia	Nomadic herders	–	5,622
UZA	Central Asia	Nomadic herders	–	5,194
UZB	Central Asia	Nomadic herders	–	5,312
UZT	Central Asia	Nomadic herders	–	4,703

We used the population growth curves generated from BEAST to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 – 20 generation intervals. We considered that the population increased since both the median and HPD values for N_e increased gradually between more than three successive data points. Although this estimation method does not provide a confidence interval for inferred expansion timings, this conservative approach ensure that we consider only strong and significant expansion signals. We used here the upper confidence intervals of mutation rates estimated by Pluzhnikov et al. (2002) for autosomes (4.74×10^{-8} /generation/site) and Heyer et al. (2001) for HVS-I (2.75×10^{-5} /generation /site).

Table I.B-S12. Summary statistics and neutrality tests computed for the largest non-recombining part of autosomal sequences.

Population	Area	Life style	S	K	Tajima's <i>D</i>	Fu&Li's <i>D</i>	Fu & Li's <i>F</i>	Fu's <i>FS</i>
Akele	Africa	sedentary farmers	6.75	6.4	-0.36	-0.57	<u>-0.59*</u>	<u>-1.12*</u>
Chagga	Africa	sedentary farmers	7.95	7.35	-0.38	<u>-0.49*</u>	<u>-0.54*</u>	<u>-1.09*</u>
Mozambicans	Africa	sedentary farmers	8.05	7.95	<u>-0.56*</u>	<u>-1.05**</u>	<u>-1.05**</u>	<u>-2.17**</u>
Ngumba	Africa	sedentary farmers	6.60	5.95	-0.19	<u>-0.38*</u>	-0.37	-0.54
Yoruba	Africa	sedentary farmers	7.10	6.80	-0.12	0.07	-0.01	-0.57
Aka	Africa	nomadic HG ^d	6.40	6.15	0.03	0.22*	0.19	-0.29
G. Baka	Africa	nomadic HG	6.00	5.75	0.07	0.19	0.18	-0.25
S. Baka	Africa	nomadic HG	5.60	5.45	0.19	0.22	0.24	-0.11
Kola	Africa	nomadic HG	6.30	6.05	-0.16	-0.05	-0.10	-0.65
Mbuti	Africa	nomadic HG	6.70	6.25	0.36	0.46*	0.50**	0.25
Danes	Eurasia	sedentary farmers	4.55	4.40	0.53*	0.37	0.49*	0.95*
Han	Eurasia	sedentary farmers	4.45	4.35	-0.07	-0.03	-0.002	0.28
Japanese	Eurasia	sedentary farmers	<u>3.50**</u>	<u>3.60**</u>	0.61**	0.32	0.48*	1.06*
Chuvash	Eurasia	nomadic herders	4.95	4.75	0.09	0.07	0.09	0.28
Tajiks (TAB)	C. Asia	sedentary farmers	5.70	5.45	0.21	0.02	0.13	0.44
Kyrgyz (KIB)	C. Asia	nomadic herders	5.65	5.30	0.21	0.14	0.20	0.37

*Number of polymorphisms, Number of haplotypes. We report the means over the 20 regions. Values significantly higher than expected under a constant population size model are indicated in bold, whereas significantly lower values are underlined. ^dHG= Hunter-gatherers We used the four-gamete test (Hudson and Kaplan, 1985) implemented in IMgc online (Woerner et al., 2007) to identify recombination events and to select the largest non-recombining part of each sequence. Significance levels: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ after FDR correction for multiple testing (Benjamini and Hochberg, 1995).*

I.C. Microsatellite data show recent demographic expansions in sedentary but not in nomadic human populations in Africa and Eurasia.

Carla Aimé, Paul Verdu, Laure Ségurel, Begoña Martínez-Cruz, Evelyne Heyer, Frédéric Austerlitz.

Article soumis en l'état à European Journal of Human genetics.

I.C.i. Abstract.

The transition from hunting and gathering to plant and animal domestication was one of the most important cultural and technological revolutions in human history. According to archeologists and paleoanthropologists, this transition triggered major demographic expansions. However, few genetic studies have found traces of Neolithic expansions in the current repartition of genetic polymorphism, pointing rather toward Paleolithic expansions. Here, we used microsatellite autosomal data to investigate the past demographic history of 60 African and Eurasian human populations with contrasted lifestyles (nomadic hunter-gatherers, semi-nomadic herders and sedentary farmers). Despite the variability in their location and environment, we found consistent expansions for all sedentary farmers, while we inferred constant population sizes for all hunter-gatherers and most herders, which could result from constraints linked to a nomadic or semi-nomadic lifestyle and/or competition for land between herders and farmers. As an exception, we inferred expansions for Central Asian herders. This might be linked with the arid environment of this area, which may have been more favorable to nomadic herders than to sedentary farmers. Alternatively, current Central Asian herders may descent from populations who have first experienced a transition from hunter-gathering to sedentary agropastoralism, and then a second transition to nomadic herding. Finally, likely due to the combination of a higher mutation rate and the possibility to analyze several loci as independent replicates of the coalescent process, the analysis of microsatellite data allowed us to infer more recent expansions than previous genetic studies, potentially resulting from the Neolithic transition.

I.C.ii. Introduction.

Reconstructing the demographic history of human populations is a major goal in many disciplines such as paleoanthropology, archeology and population genetics. The transition from hunting and gathering to plant and animal domestication, which occurred independently in several regions of the World during the Neolithic period (10 000 – 3 000 years before present, YBP) is of particular interest. This transition, which represents one of the most important cultural and technological revolutions in our history, affected many aspects of lifestyle (diet, technologies, social organization) and led to the sedentarization of many populations, which is widely assumed by archeologists and paleoanthropologists to have driven recent human population expansions (Bar-Yosef and Belfer-Cohen, 1991; Mazoyer and Roudart, 2006; Bocquet-Appel and Bar-Yosef, 2008). Bocquet-Appel (2011) notably showed an increase in both the number of enclosures and the proportion of subadults in Eurasian burial sites during the Neolithic, which was interpreted as a proof of a demographic expansion resulting from the Neolithic sedentarization. At the same time, many other populations remained nomadic during this transition. Among these populations, some kept a hunter-gatherer lifestyle, but others developed new means of subsistence like nomadic herding.

Population genetic studies have provided some additional insights on the understanding of human demographic history. Numerical methods based on the coalescent theory (Kingman, 1982) allow inferring demographic parameters from current population genetic data (Kingman, 1982; Emerson *et al.*, 2001; Excoffier and Heckel, 2006; Kuhner *et al.*, 2008). In particular, many studies based on non-recombining sex-specific markers, namely the Y chromosome and the hypervariable control region (HVS-I) of the mitochondrial DNA, have detected strong middle and upper Paleolithic expansions in both African and Eurasian post-Neolithic populations. With HVS-I data, these expansions were dated between 80 000 and 30 000 YBP in Africa (Harpending *et al.*, 1993; Watson *et al.*, 1997), and between 63 000 YBP and 17 000 YBP in Eurasia (Harpending *et al.*, 1993; Chaix *et al.*, 2008). Estimates based on the analysis of the Y-chromosome polymorphism are generally more recent: between 15 000 and 12 000 YBP for Africa, and between 40 000 and 11 000 for Eurasia (Chaix *et al.*, 2008; Pritchard *et al.*, 1999). Nevertheless, these Y-chromosome-based estimates still predate the Neolithic transition.

Moreover, some authors have taken advantage of neutral autosomal DNA sequences by combining inferences from several independent loci. This has allowed refining the dating of Paleolithic expansions in Africa from 50 000 to 25 000 YBP (Voight *et al.*, 2005; Laval *et al.*, 2010). All these studies therefore pointed to expansions starting in the Paleolithic, without evidence for a subsequent Neolithic expansion. To our knowledge, only one study detected Neolithic expansions: By separately analyzing different mitochondrial haplogroups according to their time of origin, Gignoux *et al.* (2011) showed expansions starting about 3 000 YBP in Africa, 8 000 YBP in Europe and 5 000 YBP in Southeastern Asia.

Regarding the relationship between lifestyle and demographic history, a recent study based on complete mitochondrial sequences inferred constant population sizes for hunter-gatherer populations in Africa, but demographic expansions for farmers starting about 60 000 YBP (Batini *et al.*, 2011). In a previous study, we also found contrasted demographic history between sedentary and nomadic populations predating the Neolithic transition in Africa and Eurasia, using resequencing data from both the mitochondrial HVS-I region and neutral autosomal sequences. This suggests that strong Paleolithic expansions may have ultimately favored the emergence of agriculture in some populations (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguérel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz; submitted). However, as ancient expansions could have masked potential signals of more recent expansions, these findings do not preclude the possibility of a second expansion event resulting from the Neolithic transition. Indeed, strong ancient demographic events are more easily detected with coalescent-based methods, and, in some cases, they can mask the signals of more recent events, especially if the population has not yet reached genetic equilibrium (Lavery *et al.*, 1996).

In this paper, we used autosomal microsatellites data, which offer both the advantage of a high mutation rate and the possibility to analyze several loci as independent replicates, to infer recent expansion events. Recent simulation studies and empirical studies on non-human species have shown that these markers can reveal very recent demographic events that more slowly evolving markers such as DNA sequences may fail to detect (Cornuet *et al.*, 2010; Hu *et al.*, 2011; Fontaine *et al.*, 2012). Here we investigated the past demographic history of 60 African and Eurasian populations with contrasted lifestyles (*i.e.* sedentary farmers, nomadic hunter-gatherers and nomadic herders), using a coalescent-based approach, in order to determine (1) whether autosomal microsatellite data allowed us to detect recent expansions

and (2) if such expansions occurred concurrently in populations characterized by different lifestyles.

First, we focused on Central Africa, where we studied a set of populations (commonly called Pygmies) that have retained a nomadic lifestyle based on hunter-gathering, as well as neighboring farmer populations. These two groups of populations are genetically differentiated and their divergence was inferred to have occurred about 60 000 years ago (Patin *et al.*, 2009; Verdu *et al.*, 2009), thus long before the Neolithic sedentarization of farmer populations (about 5 000 YPB in this area; Bocquet-Appel and Bar-Yosef, 2008). Secondly, we analyzed a sample of populations from four distant geographical regions of Eurasia (Middle-East, Pamir, Russia and East Asia), where sedentary farmers coexist with nomadic herders, studying one herder and one farmer population from each region. Finally, we performed a more detailed study on farmer and herder populations from Central-Asia. This area is of peculiar interest as it is characterized by a more arid climate than the rest of Eurasia, which could have impacted the development of farming and the demography of human (Dirksen and Van Geel, 2004). Moreover, although Central Asia is thought to have been a major corridor during the successive Eurasian migration waves (Nei and Roychoudhury, 1993; Cavalli-Sforza *et al.*, 1994), many aspects of the demographic history of this area remain poorly understood. In particular, it is not known whether nomadic herding in Central Asia descended directly from hunting and gathering or if the ancestors of current nomadic herders have experienced two successive life style shifts (*i.e.* the first to agropastoralism, and then to nomadic herding).

I.C.iii. Material and Methods.

Population Sampling and Marker Sets

For Africa, we used the data set of Verdu *et al* (2009, 2013), which consists of 12 farmer populations, with a total of 337 individuals, and ten Pygmy hunter-gatherer populations, with a total of 281 individuals (**Table I.C-S1**). These individuals were genotyped for 28 unlinked autosomal microsatellite locus located on 18 different chromosomes.

For Eurasia, we selected eight populations from the HGDP-CEPH Human Genome Diversity Cell Line Panel (Cann *et al.*, 2002; Rosenberg *et al.*, 2002). To avoid any bias of geographic structure in the sampling, we first identified four distant areas where farmer and

herders coexist (Middle-East, Pamir, Russia and East Asia), and then selected randomly one traditionally nomadic herder population and one neighbouring sedentary farmer population from each of these regions (Han Chinese and Mongols from East Asia, Pathans and Balochi from Pamir, Palestinians and Bedouins from Middle-East, and Russians and Yakut from Russia). This represented a total of 231 individuals (**Table I.C-S1**). For our detailed study of Central Asia, we used data from 30 populations from West Uzbekistan to East Kyrgyzstan: 17 herder populations, for a total of 584 individuals, and 13 farmer populations, for a total of 382 individuals (**Table I.C-S1**). Among these populations, 26 had been genotyped in previous studies (Ségurel *et al.*, 2008; Martinez-Cruz *et al.*, 2011). The four other populations (KIB, TAB, KIM, and TAM) were genotyped for this study, using the same PCR and genotyping procedures as in Ségurel *et al.* (2008). For all Eurasian and Central Asian populations, we used 21 autosomal microsatellites markers from the 27 selected by Ségurel *et al.* (2008) from the HGPD-CEPH panel. We removed 6 loci (GATA22F11, GATA88H02, SE30, GATA28D01, GATA11B12 and GATA12A07) from the original data set as they presented compound repeat motives, and thus might not well fit the assumption of a SSM (single stepwise model).

All sampled individuals were healthy donors from whom informed consent was obtained. The study was approved by appropriate Ethic Committees and scientific organizations in all countries where samples have been collected.

Data analysis

We computed the likelihood distribution of demographic parameters (Kuhner *et al.*, 1995) using the MCMC algorithm implemented in BEAST v1.7.4 (Drummond and Rambaut *et al.*, 2007). We tested four demographic models implemented in this software: constant effective population size (N_0) (“Constant model”), exponential variation of N_0 (“Exponential model”), logistic variation of N_0 (“Logistic model”) and expansion starting at a given moment in the past from an ancestral population of constant size N_1 (“Expansion model”). This allowed us to infer the current effective population sizes (N_0) and growth rates (g) for all analyzed populations. For the “Expansion model”, we also inferred the ancestral population size (N_1). We then inferred the dates of expansion onsets (t) using the following formula: $t = (1/g) \times \ln(N_1/N_0)$, applied to each step of the MCMC algorithm. We used a single step

mutational model (SSM), which takes homoplasy into account. The implementation of microsatellite models in BEAST v1.7.4 is described in Wu and Drummond (2011).

We performed three runs of 2.8×10^8 steps per population and per demographic model for the African populations, and three runs of 2.1×10^8 steps for the Eurasian populations (which corresponded for all populations to three runs of 10^7 steps per locus). We recorded one tree every 1 000 steps, resulting in a total of 10^5 trees per locus and per run. We then removed the first 10% of each run (burn-in period) and combined the runs to obtain reasonably high effective sample sizes (ESS of 100 or above, see *e.g.* Gignoux *et al.*, 2011). The convergence of these runs was assessed by visual inspection of traces using Tracer v1.5 (Rambaut and Drummond, 2007) to check for concordance between runs, and also by the computation of the Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992) using R v2.14.1 (R Development Core Team, 2011) with the function “gelman.diag” available in the add-on package coda (Plummer *et al.*, 2006).

In order to facilitate a large exploration of the parameter space, for the autosomal sequences, we chose uniform priors between 1 and 10^6 for $2N_e$ and between -0.1 and 0.1 per generation for g . In agreement with previous studies, we assumed a uniform prior between 10^{-4} and 10^{-3} per generation for the mean mutation rate (μ) (Verdu *et al.*, 2009), and the mutation rate of each locus was drawn independently from a Gamma distribution (mean= μ and shape=2). As in previous studies (Voight *et al.*, 2005; Chaix *et al.*, 2008; Patin *et al.*, 2009; Laval *et al.*, 2010; Batini *et al.*, 2011), we assumed a generation time of 25 years. For each population and demographic model, we obtained the mode and the 95% Highest Probability Density (HPD) interval of each parameter, inferred from the posterior distributions obtained using the add-on package Locfit (Loader, 1999) in R v2.14.1. Then, we computed the Deviance Information Criteria (DIC) of each model, as the model with lower DIC is considered as the best-fitting model for the observed data (Spiegelhalter, 2002). DIC is computed as $DIC = p_D + \bar{D}$, where p_D represents the effective number of parameters of the model, defined as $\bar{D} - D(\bar{\theta})$. $D(\bar{\theta})$ is the deviance parameter calculated as $-2 \log(p(y|\theta))$, where $p(y|\theta)$ represents the likelihood function, substituted here with the probability of the data (y) given the posterior mean of the parameter values (θ). \bar{D} is the expectation of the likelihood associated with the means of posterior parameters distributions ($\bar{D} = E^\theta [D(\bar{\theta})]$). To obtain \bar{D} , we repeated all analyses using the posterior mean of each parameter as punctual

priors. A difference of five points in DIC was considered as significant (Spiegelhalter *et al.*, 2002; Bolker, 2008).

I.C.iv. Results.

Contrasting demographic histories for sedentary and nomadic population

For Africa, the expansion model had the lowest DIC value (*i.e.* best fitted the data) for all sedentary farmer populations, while the constant model provided the best fit to the data for all nomadic hunter-gatherer populations (**Table I.C-1**). Note that the differences in DIC values between the best-fitting model and the others were higher than five points for all populations except for one farmer population (Ewondo). For this population, the Expansion and the Constant model could therefore not be distinguished (**Table I.C-S2**).

For Eurasia, among the eight populations from the HGDP-CEPH panel, we found that the expansion model best fitted the data for sedentary farmers, whereas this was the constant model for nomadic herders (**Table I.C-1, I.C-S2**). Conversely, when focusing on Central-Asian populations, we found that the expansion model best fitted the data for both farmer and herder populations (**Table I.C-1, I.C-S2**). See **Table I.C-S3** for the modes and 95% HPD of all inferred parameters for each population separately.

Timing of expansions

For sedentary farmer populations from Africa, we inferred expansion signals starting between 3 036 and 6 939 [1 217 – 23,054] YBP, concomitantly with the Neolithic transition in this area, about 5 000 YPB (**Table I.C-2, I.C-S3**). For all Eurasians farmers, as well as for Central Asian nomadic herder populations, we estimated expansion onset times from 5 111 [2 669 – 30 851] YBP for Palestinians to 2 864 [1609 – 12842] YBP for Han Chinese (**Table I.C-2, I.C-S3**). In all cases, the modal estimates of the expansion onsets fell after the Neolithic transition but the 95% HPD included the upper Paleolithic period (**Table I.C-2**).

Table I.C-1. Mean values of the modes and 95% HPD limits for the inferred current effective population size (N_0), ancestral effective population size (N_1) and per year growth rate (g) estimated under the best-fitting (BF) model for each area.

Area	Life style	BF model	N_1 95% lower	N_1 mode	N_1 95% upper	N_0 95% lower	N_0 mode	N_0 95% upper	g 95% lower	g mode	g 95% upper
Africa	Sedentary farmers	Expansion ¹	17	18 429	93 394	39	40 370	186 263	1.39×10^{-03}	3.47×10^{-03}	1.50×10^{-02}
Africa	Nomadic HG	Constant	–	–	–	9 873	15 859	19 026	–	–	–
Eurasia	Sedentary farmers	Expansion	12 181	15 853	22 404	19 099	25 600	37 769	2.47×10^{-04}	1.96×10^{-03}	3.49×10^{-03}
Eurasia	Nomadic herders	Constant	–	–	–	21 969	26 651	32 552	–	–	–
C.Asia	Sedentary farmers	Expansion	11 592	24 988	47 875	23 713	32 662	49 933	2.27×10^{-04}	1.85×10^{-03}	3.28×10^{-03}
C.Asia	Nomadic herders	Expansion	11 042	21 706	41 412	21 752	28 804	42 180	2.41×10^{-04}	1.90×10^{-03}	3.48×10^{-03}

For each area, we report only the mean values over populations. Detailed results for each population separately are presented in **Table I.C-S3**.

¹Note that for one African population (Ewondo), the Expansion model and a Constant model could not be distinguished from their DIC values (**Table I.C-S2**).

Table I.C-2. Modes and 95% HPD of expansion onset times (t) estimated under the expansion model for expanding populations, compared with datations of the emergence of farming.

Area	Analyzed populations	Life-style	Emergence of farming or herding	t 95% lower	t mode	t 95% upper
Africa	Central-African farmers ¹	Sedentary farmers	5 000 YBP	2 905	5 339	17 801
East-Asia	Han Chinese	Sedentary farmers	9 000 YBP	1 608	2 864	12 842
Middle-East	Palestinians	Sedentary farmers	11 000 YBP	2 668	5 111	30 851
Pamir	Pathans	Sedentary farmers	8 000 YBP	2 153	4 588	22 691
Russia	Russians	Sedentary farmers	7 500 YBP	2 415	4 489	21 896
C.Asia	indo-iranian farmers ¹	Sedentary farmers	9 000 YBP	2 399	4 760	23 406
C.Asia	Turko-mongol herders ¹	Nomadic herders	9 000 YBP	2 160	4 280	21 163

¹For these area, we report the mean values over populations. Detailed results for each population separately are presented in **Table I.C-S3**.

I.C.v. Discussion.

In this study, we have found genetic signals of demographic expansions for all sedentary farmer populations analyzed, while we inferred constant population sizes for all nomadic hunter-gatherer populations in Africa and most nomadic herder populations in Eurasia. As an exception, the nomadic herders from Central Asia displayed an expansion signal similar to that of the farmer populations in the same area. For both African and Eurasian populations, the modal estimate of expansion times were consistent with an onset of these expansions during or after the Neolithic expansion, even though the 95% HPD could not exclude late Paleolithic expansions (up to ~30 000 years ago).

It is important to note that the method used in this study makes the assumption that populations are isolated and panmictic, which is questionable for human populations. Nevertheless, we analyzed a large set of populations sampled in very distant geographical regions (*i.e.* Central Africa, Middle-East, Central Asia, Pamir, Russia and East Asia), and the main conclusions of this study are consistent across these different regions. It therefore seems unlikely that local processes, such as admixture, would have similarly biased estimates in each of these regions, as these admixture processes are likely to vary in nature and intensity across regions. Since we used many independent loci, we can also be confident to have avoided confounding effects of selection.

Contrasted demographic histories associated with different life-styles.

Despite cultural and environmental variability, we inferred similar patterns of expansion for all studied sedentary farmer populations in both Africa and Eurasia. Moreover, except for Central Asia, we found constant population sizes for both nomadic hunter-gatherers and nomadic herders. These findings are consistent with paleo-anthropological and archeological records, which suggest that the growth of farmer populations after the Neolithic resulted from the processes of sedentarization (Bar-Yosef and Belfer-Cohen, 1991; Bocquet-Appel and Bar-Yosef, 2008). Conversely, both herders and hunter-gatherer populations may have remained at constant size due to constraints associated with their nomadic life-style. For instance, birth intervals are generally longer (at least by four years) in nomadic populations than in sedentary populations (Short, 1982). In addition, competition between farmers and nomadic populations for land may have led to a reduction in, and fragmentation of, available

habitat and resources for nomads, thereby limiting their possibilities of expansion (Diamond, 2002).

It is interesting to note that we detected expansion signals for both sedentary farmers and nomadic herders in Central Asia. This could be due to the arid continental climate in this area, which has strongly impacted the cultural development and limited the spread of farmer civilizations up to 3 000 YBP (Dirksen and Van Geel, 1994). The limitation of opportunities for farmers to expand in this area might thus have allowed herders population size to increase more than elsewhere. Alternatively, as suggested by Renfrew (1996) from linguistic data, current Central Asian herders may descend from sedentary agropastoralist populations (Jeitun culture). In this case, these populations may have experienced two successive transitions, the first occurring from hunter-gathering to agropastoralism and having potentially led to demographic expansions, and the second occurring from agropastoralism to nomadic herding.

Autosomal microsatellites allow inferring more recent expansions as compared to other markers.

Previous population genetic studies also found demographic expansion events in farmers but not in hunter-gatherers in Africa using autosomal DNA sequences (Patin *et al.*, 2009; Laval *et al.*, 2010) or complete mitochondrial DNA sequences (Batini *et al.*, 2011). Similarly, our previous work on HVS-1 data has demonstrated strong signals of expansion in both African and Eurasian sedentary farmer populations, weaker expansion events in Eurasian herders and no expansion in African hunter-gatherers (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz; submitted). However, these studies reported expansions that started during the Paleolithic. Here, using autosomal microsatellite data, although confidence intervals were quite large, we inferred modal estimates for the expansion onsets during or after the Neolithic transition for all farmer populations in Africa and Eurasia, as well as for herder populations in Central Asia. This finding appears to be robust to the uncertainty of generation time in humans. Indeed, using a generation time of 29 years (Tremblay and Vezina, 2000) instead of 25 years (*e.g.* Chaix *et al.*, 2008; Patin *et al.*, 2009; Laval *et al.*, 2010) yields estimates that are only 1.2 times more ancient, thus still after the Neolithic transition (data not shown).

This result is consistent with Gignoux *et al.* (2011) who detected Neolithic expansions using mtDNA data. Indeed, by separately analyzing different haplogroups according to their

time of origin, they reported expansions starting around 3 000 YBP in Africa, 8 000 YBP in Europe and 5 000 YBP in Southeastern Asia for the most recent haplogroups. They inferred demographic patterns at the continental scale by simultaneously analyzing individuals from multiple populations with various lifestyles. By studying a large set of populations, we were able to show here that (1) different expansion patterns are associated with different lifestyles, (2) demographic patterns (and their timing) are consistent between populations of the same group (*i.e.* nomadic or sedentary) within each area and (3) peculiar regions such as Central Asia display atypical patterns. Moreover, as we pointed out above, mtDNA data can be affected by selection. The fact that we obtained similar expansion patterns from autosomal microsatellite data strongly suggests that these inferred expansions truly reflect demographic processes and not selective sweeps.

Otherwise, most previous population genetics studies focusing on the past demographic history of African and Eurasian populations have inferred upper or middle Paleolithic expansions, but no Neolithic expansions. For Africa, using mitochondrial or autosomal sequences data, many authors have inferred expansion times ranging from 80 000 to 25 000 YBP (Harpending *et al.*, 1993; Watson *et al.*, 1997; Voight *et al.*, 2005; Laval *et al.*, 2010; Batini *et al.*, 2011). Using Y-chromosome microsatellites, Pritchard *et al.* (1999) estimated more recent but still Paleolithic expansion times, between 15 000 and 12 000 YBP. Finally, analyzing separately different mitochondrial haplogroups, Atkinson *et al.* (2009) found several successive Paleolithic expansion waves occurring at different times depending upon the population, but earlier than 8 000 YBP in all cases. For Eurasia, Chaix *et al.* (2008) showed estimations of expansion times ranging from 63 000 and 17 000 YBP using HVS-I data, and between 40 000 and 11 000 YBP using Y-chromosome microsatellites (see also Pritchard *et al.*, 1999). Accordingly, Harpending *et al.* (1993) inferred an expansion event starting around 40 000 YBP in Europe using HVS-I data. Finally, one study showed expansion events in early Europeans farmers, starting at least 12 000 YBP according to an unbiased sample of complete mitochondrial DNA sequences (Fu *et al.*, 2012). Altogether, these results suggest that different type of markers can be informative at different time-scales. Furthermore, we found more recent expansion events in East-Asia than in the rest of Eurasia, unlike a previous study showing a Eurasian East-to-West gradient of Paleolithic expansions onsets (Chaix *et al.*, 2008). It suggests that we may have detected another set of expansion

events, potentially resulting from the Neolithic transition. These events therefore do not appear to have followed the same East-to-West process as the Paleolithic expansions.

The fact that microsatellites may allow detecting more recent events than other type of markers with coalescent-based methods had previously been demonstrated with simulation-based analyses (Cornuet *et al.*, 2008) and empirical data analyses on non-human organisms (Cornuet *et al.*, 2008; Hu *et al.*, 2011). This aspect of microsatellites could be related to several peculiar properties of microsatellite markers, in particular their high mutation rate. However, Y-chromosome microsatellite-based studies (Pritchard *et al.*, 1999; Chaix, 2008) generally point toward Paleolithic expansion events, more ancient than those that we found here with autosomal microsatellites. Thus, the higher sensitivity to recent events of autosomal microsatellites cannot only be explained by their relatively high mutation rates, as then similar results would be expected for analyses of Y-chromosome microsatellites. Instead, we suggest that the increased sensitivity to recent demography reflects the combination of high mutation rates and of the fact that, unlike Y-chromosome microsatellites, autosomal microsatellites are independent loci and provide therefore independent replicates of the coalescent process. Indeed, mutations can accumulate on the different Y chromosome haplotypes, and even if homoplasmy may punctually reduce the differences between haplotypes, these differences are expected to increase with time. Conversely the autosomal microsatellites evolve independently and therefore this process of accumulation does not occur. Altogether, high mutability and independence between loci may thus render autosomal microsatellites more sensitive to recent demographic events and hence make them suitable for their analysis.

I.C.vi. Conclusion and further work.

Using microsatellite autosomal data for a large set of populations from several distant geographic areas, we have shown contrasted demographic histories that correlate with lifestyle differences. Benefiting from the high mutation rate of microsatellites and the simultaneous analysis of multiple independent loci, we detected more recent expansion events than some previous studies, which may result from the sedentarization of farmer populations during the Neolithic transition. Finally, comparing our results with previous studies, we have shown that different types of markers can be informative at different time scales. It will be therefore interesting to extend those analyses to other types of genetic markers, such as

genome-wide data. Simulation studies may also help to better understand the specific responses of each type of genetic markers to past demographical events.

I.C.vii. Supplementary Tables.

Table I.C-S1. Description and location of studied populations.

Population	code	Area	Country	Life-style	number of markers	number of samples
Akele	—	Central Africa	Gabon	Sedentary farmers	28	13
Bangando	—	Central Africa	Cameroon	Sedentary farmers	28	30
C. Fang	—	Central Africa	Cameroon	Sedentary farmers	28	30
Ewondo	—	Central Africa	Cameroon	Sedentary farmers	28	24
G. Fang	—	Central Africa	Gabon	Sedentary farmers	28	30
Kota	—	Central Africa	Gabon	Sedentary farmers	28	30
Ngumba	—	Central Africa	Cameroon	Sedentary farmers	28	30
Nzebi	—	Central Africa	Gabon	Sedentary farmers	28	30
Nzime	—	Central Africa	Cameroon	Sedentary farmers	28	31
Teke	—	Central Africa	Gabon	Sedentary farmers	28	30
Tikar	—	Central Africa	Cameroon	Sedentary farmers	28	30
Tsogho	—	Central Africa	Gabon	Sedentary farmers	28	30
Bezan	—	Central Africa	Cameroon	Nomadic HG	28	30
C. Baka	—	Central Africa	Cameroon	Nomadic HG	28	30
C.Bongo	—	Central Africa	Gabon	Nomadic HG	28	13
E. Bongo	—	Central Africa	Gabon	Nomadic HG	28	30
E. Baka	—	Central Africa	Cameroon	Nomadic HG	28	29
G.Baka	—	Central Africa	Gabon	Nomadic HG	28	30
Kola	—	Central Africa	Cameroon	Nomadic HG	28	31
Koya	—	Central Africa	Gabon	Nomadic HG	28	29
S.Bongo	—	Central Africa	Gabon	Nomadic HG	28	30
S.Baka	—	Central Africa	Cameroon	Nomadic HG	28	30
H.Chinese	—	Eurasia	China	Sedentary farmers	27	44
Palestinians	—	Eurasia	Israël	Sedentary farmers	27	46
Pathans	—	Eurasia	Pakistan	Sedentary farmers	27	24
Russians	—	Eurasia	Russia	Sedentary farmers	27	25

Balochi	–	Eurasia	Pakistan	Nomadic herders	27	25
Bedouins	–	Eurasia	Saudi Arabia	Nomadic herders	27	46
Mongols	–	Eurasia	Mongolia	Nomadic herders	27	10
Yakut	–	Eurasia	Siberia	Nomadic herders	27	26
Tajiko-uzbeks (Bukhara)	LUZ	Central-Asia	Uzbekistan	Sedentary farmers	27	40
Tajiks (Boukhara)	TAB	Central-Asia	Uzbekistan	Sedentary farmers	27	46
Tajiks (Muchaus)	TAM	Central-Asia	Uzbekistan	Sedentary farmers	27	32
Tajiks (Penjikent)	TDS	Central-Asia	Tajikistan/ Uzbekistan border	Sedentary farmers	27	25
Tajiks (Penjikent)	TDU	Central-Asia	Tajikistan/ Uzbekistan border	Sedentary farmers	27	25
Tajiks (Samarkand)	TJA	Central-Asia	Uzbekistan/ Tajikistan border	Sedentary farmers	27	31
Tajiks (Gharm)	TJE	Central-Asia	Tajikistan	Sedentary farmers	27	26
Tajiks (Ferghana)	TJK	Central-Asia	Uzbekistan/ Kyrgyzstan border	Sedentary farmers	27	26
Tajiks (Gharm)	TJN	Central-Asia	Tajikistan	Sedentary farmers	27	24
Tajiks (Ferghana)	TJR	Central-Asia	Uzbekistan/ Kyrgyzstan border	Sedentary farmers	27	29
Tajiks (Gharm)	TJT	Central-Asia	Tajikistan	Sedentary farmers	27	25
Tajiks (Samarkand)	TJU	Central-Asia	Uzbekistan/ Tajikistan border	Sedentary farmers	27	29
Tajiks (Yagnobs from Dushanbe)	TJY	Central-Asia	Tajikistan	Sedentary farmers	27	25
Karakalpaks (Karakalpakia)	KAR	Central-Asia	Uzbekistan	Nomadic herders	27	45
Kazaks (Karakalpakia)	KAZ	Central-Asia	Uzbekistan	Nomadic herders	27	49
Kyrgyz (Bichkek)	KIB	Central-Asia	Kyrgyzstan	Nomadic herders	27	48
Kyrgyz (At-Bashy)	KIM	Central-Asia	Kyrgyzstan	Nomadic herders	27	32
Kyrgyz (Andijan)	KRA	Central-Asia	Uzbekistan/ Kyrgyzstan border	Nomadic herders	27	48
Kyrgyz (Narin)	KRB	Central-Asia	Kyrgyzstan	Nomadic herders	27	25

Kyrgyz (Narin)	KRG	Central-Asia	Kyrgyzstan	Nomadic herders	27	18
Kyrgyz (Narin)	KRL	Central-Asia	Kyrgyzstan	Nomadic herders	27	22
Kyrgyz (Narin)	KRM	Central-Asia	Kyrgyzstan	Nomadic herders	27	19
Kyrgyz (Issyk Kul)	KRT	Central-Asia	Kyrgyzstan	Nomadic herders	27	37
Kazaks (Bukhara)	LKZ	Central-Asia	Uzbekistan	Nomadic herders	27	25
Karakalpaks (Karakalpakia)	OTU	Central-Asia	Uzbekistan	Nomadic herders	27	45
Turkmen (Urgench)	TKY	Central-Asia	Kyrgyzstan	Nomadic herders	27	40
Turkmen (Karakalpakia)	TUR	Central-Asia	Kyrgyzstan	Nomadic herders	27	47
Uzbeks (Ferghana)	UZA	Central-Asia	Uzbekistan/ Kyrgyzstan border	Nomadic herders	27	25
Uzbeks (Karakalpakia)	UZB	Central-Asia	Uzbekistan	Nomadic herders	27	35
Uzbeks (Penjikent)	UZT	Central-Asia	Tajikistan/ Uzbekistan border	Nomadic herders	27	25

Table I.C-S2. Deviance Information Criterion (DIC) values computed for each demographic model.

Population	Area	Life-style	Constant model	Exponential model	Logistic model	Expansion model
Akele	Africa	Sedentary farmers	1,820	1,853	1,821	1,811
Bangando	Africa	Sedentary farmers	1,748	1,735	1,730	1,723
C. Fang	Africa	Sedentary farmers	1,697	1,671	1,691	1,655
Ewondo	Africa	Sedentary farmers	1,688	1,695	1,691	1,687
G. Fang	Africa	Sedentary farmers	1,903	1,893	1,891	1,880
Kota	Africa	Sedentary farmers	1,684	1,697	1,693	1,669
Ngumba	Africa	Sedentary farmers	1,861	1,862	1,861	1,843
Nzebi	Africa	Sedentary farmers	1,534	1,548	1,555	1,524
Nzime	Africa	Sedentary farmers	1,828	1,846	1,830	1,819
Teke	Africa	Sedentary farmers	1,991	1,985	1,982	1,972
Tikar	Africa	Sedentary farmers	1,809	1,817	1,819	1,790
Tsogho	Africa	Sedentary farmers	1,616	1,628	1,644	1,609
Bezan	Africa	Nomadic HG	1,615	1,637	1,628	1,632
C. Baka	Africa	Nomadic HG	1,629	1,636	1,645	1,639
C.Bongo	Africa	Nomadic HG	1,526	1,555	1,547	1,539
E. Bongo	Africa	Nomadic HG	1,816	1,832	1,831	1,838
E. Baka	Africa	Nomadic HG	994	1,002	1,001	1,008
G.Baka	Africa	Nomadic HG	1,795	1,814	1,805	1,814
Kola	Africa	Nomadic HG	1,920	1,932	1,933	1,945
Koya	Africa	Nomadic HG	1,615	1,635	1,642	1,633
S.Bongo	Africa	Nomadic HG	1,620	1,651	1,652	1,643
S.Baka	Africa	Nomadic HG	1,738	1,762	1,758	1,762
H.Chinese	Eurasia	Sedentary farmers	18 861	18 563	19 853	18 354
Palestinians	Eurasia	Sedentary farmers	30 661	29 949	30 658	29 025
Pathans	Eurasia	Sedentary farmers	13 718	13 398	13 586	12 114
Russians	Eurasia	Sedentary farmers	13 792	12 798	13 598	11 178
Balochi	Eurasia	Nomadic herders	9 348	11 529	10 256	9 565
Bedouins	Eurasia	Nomadic herders	30 125	31 219	30 825	30 503
Mongols	Eurasia	Nomadic herders	2 203	2 754	2 587	2 325

Yakut	Eurasia	Nomadic herders	11 019	11 721	11 853	11 258
LUZ	C.Asia	Sedentary farmers	-26 711	-26 125	-25 916	-27 925
TAB	C.Asia	Sedentary farmers	-31 231	-31 298	-31 423	-31 508
TAM	C.Asia	Sedentary farmers	-24 946	-25 035	-25 174	-25 205
TDS	C.Asia	Sedentary farmers	-8 318	-9 217	-8 642	-9 724
TDU	C.Asia	Sedentary farmers	-6 891	-7 896	-7 658	-8 364
TJA	C.Asia	Sedentary farmers	-10 680	-11 257	-10 885	-11 554
TJE	C.Asia	Sedentary farmers	-7 189	-4 027	-6 528	-7 782
TJK	C.Asia	Sedentary farmers	-8 139	-8 632	-7 656	-8 852
TJN	C.Asia	Sedentary farmers	-16 245	-16 318	-16 395	-16 488
TJR	C.Asia	Sedentary farmers	-9 205	-6 877	-7 654	-10 082
TJT	C.Asia	Sedentary farmers	-15 591	-14 733	-14 853	-15 783
TJU	C.Asia	Sedentary farmers	-18 652	-17 274	-18 357	-19 285
TJY	C.Asia	Sedentary farmers	-14 452	-14 495	-14 796	-15 029
KAR	C.Asia	Nomadic herders	-26 903	-26 897	-26 751	-27 631
KAZ	C.Asia	Nomadic herders	-30 326	-30 644	-30 457	-31 033
KIB	C.Asia	Nomadic herders	-29 245	-29 347	-29 334	-29 781
KIM	C.Asia	Nomadic herders	-16 648	-16 279	-16 785	-17 052
KRA	C.Asia	Nomadic herders	-29 848	-29 596	-29 857	-29 991
KRB	C.Asia	Nomadic herders	-18 409	-18 166	-18 443	-18 996
KRG	C.Asia	Nomadic herders	-17 544	-17 632	-17 321	-18 199
KRL	C.Asia	Nomadic herders	-19 183	-18 732	-18 362	-19 631
KRM	C.Asia	Nomadic herders	-18 622	-18 392	-18 743	-19 653
KRT	C.Asia	Nomadic herders	-17 728	-18 959	-18 239	-19 235
LKZ	C.Asia	Nomadic herders	-8 218	-12 863	-12 641	-12 978
OTU	C.Asia	Nomadic herders	-18566	-17635	-18 864	-19 065
TKY	C.Asia	Nomadic herders	-20 561	-23 506	-22 586	-23 787
TUR	C.Asia	Nomadic herders	-17 407	-17 673	-18 356	-18 542
UZA	C.Asia	Nomadic herders	-6 662	-6 471	-7 002	-7 165
UZB	C.Asia	Nomadic herders	-11 412	-10 694	-11 358	-11 829
UZT	C.Asia	Nomadic herders	-7 991	-7 815	-8 139	-8 265

The model with the lowest DIC value which is indicated in bold for each population is considered as the best-fitting model for the observed data (Spiegelhalter et al., 2002). A difference of five points in DIC was considered as significant. The DIC formula is described with details in "Material and methods" section.

Table I.C-S3. Modes and 95% HPD of current effective population size (N_0), ancestral effective population size, per year growth rate (g) and expansion onset (in years BP) estimated under the best-fitting (BF) model for each population.

Population	Area	Life-style	BF model	N_1 95% lower	N_1 mode	N_1 95% upper	N_0 95% lower	N_0 mode	N_0 95% upper	g 95% lower	g mode	g 95% upper	t 95% lower	t mode	t 95% upper
Akele	Africa	Sedentary farmers	Expansion	16	18 003	103 298	32	36 023	206 692	1.15×10^{-03}	1.89×10^{-03}	6.86×10^{-03}	1217	4652	13204
Bangando	Africa	Sedentary farmers	Expansion	21	17 366	86 825	40	34 576	197 381	1.06×10^{-03}	4.00×10^{-03}	7.90×10^{-03}	2541	4865	23054
C. Fang	Africa	Sedentary farmers	Expansion	18	29 814	92 440	36	58 917	182 673	1.01×10^{-03}	3.84×10^{-03}	7.18×10^{-03}	3079	5426	22228
Ewondo ¹	Africa	Sedentary farmers	Constant	–	–	–	17 604	21 090	25 391	–	–	–	–	–	–
Ewondo ¹	Africa	Sedentary farmers	Expansion	13	20 318	86 848	35	42 182	197 228	1.71×10^{-03}	3.75×10^{-03}	7.22×10^{-03}	4024	6939	17190
G. Fang	Africa	Sedentary farmers	Expansion	11	13 786	82 590	34	26 993	161 712	1.25×10^{-03}	4.15×10^{-03}	8.18×10^{-03}	1557	3036	13919
Kota	Africa	Sedentary farmers	Expansion	17	12 148	98 281	45	30 929	177 493	2.02×10^{-03}	3.80×10^{-03}	7.09×10^{-03}	3254	5934	12975
Ngumba	Africa	Sedentary farmers	Expansion	21	12 741	87 346	43	37 146	180 909	1.15×10^{-03}	3.89×10^{-03}	7.67×10^{-03}	2978	4752	12005
Nzebi	Africa	Sedentary farmers	Expansion	10	26 553	98 472	20	52 032	192 958	2.05×10^{-03}	3.11×10^{-03}	3.97×10^{-02}	3786	6297	13046
Nzime	Africa	Sedentary farmers	Expansion	15	17 095	123 719	35	34 065	211 593	1.38×10^{-03}	3.74×10^{-03}	7.34×10^{-03}	1834	4560	27928
Teke	Africa	Sedentary farmers	Expansion	25	20 266	78 967	53	48 532	157 416	9.01×10^{-04}	1.78×10^{-03}	6.54×10^{-02}	3129	5439	18947

Tikar	Africa	Sedentary farmers	Expansion	14	18 472	108 676	27	35 232	209 827	1.97x10 ⁻⁰³	5.15 x10 ⁻⁰³	1.01 x10 ⁻⁰²	3478	6012	18752
Tsogho	Africa	Sedentary farmers	Expansion	15	14 579	73 264	47	47 819	159 269	1.03 x10 ⁻⁰³	2.54 x10 ⁻⁰³	4.78 x10 ⁻⁰³	3987	6159	20364
Bezan	Africa	Nomadic HG	Constant	–	–	–	11 611	14 128	17 348	–	–	–	–	–	–
C. Baka	Africa	Nomadic HG	Constant	–	–	–	13 402	16 328	18 354	–	–	–	–	–	–
C.Bongo	Africa	Nomadic HG	Constant	–	–	–	13 865	17193	21 493	–	–	–	–	–	–
E. Bongo	Africa	Nomadic HG	Constant	–	–	–	12 574	16 473	19 702	–	–	–	–	–	–
E. Baka	Africa	Nomadic HG	Constant	–	–	–	12 467	15 383	19 315	–	–	–	–	–	–
G.Baka	Africa	Nomadic HG	Constant	–	–	–	1 051	14 316	26 626	–	–	–	–	–	–
Kola	Africa	Nomadic HG	Constant	–	–	–	4 389	9 460	13 531	–	–	–	–	–	–
Koya	Africa	Nomadic HG	Constant	–	–	–	4 993	13 215	51 436	–	–	–	–	–	–
S.Bongo	Africa	Nomadic HG	Constant	–	–	–	12 036	16 793	18 731	–	–	–	–	–	–
S.Baka	Africa	Nomadic HG	Constant	–	–	–	4 644	10 187	45 730	–	–	–	–	–	–
H.Chinese	Eurasia	Sedentary farmers	Expansion	8 035	10 577	16 045	14 806	19 491	29 567	4.63 x10 ⁻⁰⁴	2.88 x10 ⁻⁰³	4.97 x10 ⁻⁰³	1609	2864	12842
Palestinians	Eurasia	Sedentary farmers	Expansion	14 047	18 444	26 336	26 134	34 313	48 996	1.32 x10 ⁻⁰⁴	1.50 x10 ⁻⁰³	2.86 x10 ⁻⁰³	2669	5111	30851
Pathans	Eurasia	Sedentary farmers	Expansion	12 514	17 152	25 594	22 941	31 445	46 920	2.37 x10 ⁻⁰⁴	1.80 x10 ⁻⁰³	3.26 x10 ⁻⁰³	2153	4589	22692
Russians	Eurasia	Sedentary farmers	Expansion	14 129	17 238	21 641	12 514	17 152	25 594	1.55 x10 ⁻⁰⁴	1.66 x10 ⁻⁰³	2.88 x10 ⁻⁰³	2415	4489	21896
Balochi	Eurasia	Nomadic herders	Constant	–	–	–	26 814	32 076	38 528	–	–	–	–	–	–

Bedouins	Eurasia	Nomadic herders	Constant	–	–	–	21 213	26 287	32 779	–	–	–	–	–	–
Mongols	Eurasia	Nomadic herders	Constant	–	–	–	21 233	26 347	32 877	–	–	–	–	–	–
Yakut	Eurasia	Nomadic herders	Constant	–	–	–	18 615	21 894	26 024	–	–	–	–	–	–
LUZ	C. Asia	Sedentary farmers	Expansion	15 626	25 543	47 899	27 184	34 043	48 298	9.79×10^{-05}	1.31×10^{-03}	2.87×10^{-03}	2494	5494	33131
TAB	C. Asia	Sedentary farmers	Expansion	13 922	27 562	55 570	28 246	38 833	56 019	1.39×10^{-04}	1.70×10^{-03}	2.78×10^{-03}	2640	5084	30918
TAM	C. Asia	Sedentary farmers	Expansion	13 040	24 237	36 350	21 089	26 363	37 088	3.04×10^{-04}	1.96×10^{-03}	3.49×10^{-03}	2541	5352	30907
TDS	C. Asia	Sedentary farmers	Expansion	11 166	19 479	35 531	19 651	26 183	37 317	2.62×10^{-04}	2.13×10^{-03}	3.54×10^{-03}	2033	4038	18754
TDU	C. Asia	Sedentary farmers	Expansion	9 262	23 242	42 444	21 847	30 956	47 927	2.04×10^{-04}	1.97×10^{-03}	3.54×10^{-03}	2048	4078	21649
TJA	C. Asia	Sedentary farmers	Expansion	14 791	24 379	35 639	22 703	31 139	37 940	2.32×10^{-04}	2.09×10^{-03}	3.93×10^{-03}	2318	4588	21275
TJE	C. Asia	Sedentary farmers	Expansion	9 487	24 377	57 499	24 606	32 726	60 235	2.65×10^{-04}	1.83×10^{-03}	3.17×10^{-03}	2661	4754	19415
TJK	C. Asia	Sedentary farmers	Expansion	13 272	29 070	53 646	26 886	34 587	54 380	2.41×10^{-04}	1.59×10^{-03}	2.90×10^{-03}	2740	5061	23966
TJN	C. Asia	Sedentary farmers	Expansion	9 156	24 922	51 186	22 921	35 305	52 820	2.48×10^{-04}	1.79×10^{-03}	3.35×10^{-03}	2480	4734	19701
TJR	C. Asia	Sedentary farmers	Expansion	9 186	27 452	62 537	25 645	34 952	65 452	2.51×10^{-04}	1.81×10^{-03}	3.12×10^{-03}	2486	5090	21918
TJT	C. Asia	Sedentary farmers	Expansion	5 911	26 313	67 954	22 690	37 716	69 887	2.73×10^{-04}	2.23×10^{-03}	3.16×10^{-03}	2444	4924	21872

TJU	C. Asia	Sedentary farmers	Expansion	12 833	26 372	40 767	23 703	32 431	44 667	1.30×10^{-04}	1.71×10^{-03}	3.28×10^{-03}	2165	4511	23410
TJY	C. Asia	Sedentary farmers	Expansion	13 040	21 892	35 350	21 089	29 363	37 088	3.04×10^{-04}	1.96×10^{-03}	3.49×10^{-03}	2132	4167	17356
KAR	C. Asia	Nomadic herders	Expansion	14 656	21 522	39 860	24 075	29 322	40 155	1.89×10^{-04}	1.81×10^{-03}	3.19×10^{-03}	2075	4672	24493
KAZ	C. Asia	Nomadic herders	Expansion	13 757	20 418	42 668	23 956	26 018	42 955	1.95×10^{-04}	1.39×10^{-03}	3.10×10^{-03}	2430	4772	27316
KIB	C. Asia	Nomadic herders	Expansion	11 197	21 968	41 824	22 143	27 465	42 177	3.25×10^{-04}	1.98×10^{-03}	3.44×10^{-03}	2399	4055	18062
KIM	C. Asia	Nomadic herders	Expansion	13 421	23 264	37 192	22 199	30 164	37 477	2.45×10^{-04}	1.97×10^{-03}	3.46×10^{-03}	2219	4122	19266
KRA	C. Asia	Nomadic herders	Expansion	11 803	19 522	35 246	20 131	27 385	35 508	1.95×10^{-04}	2.11×10^{-03}	3.78×10^{-03}	1537	3662	19931
KRB	C. Asia	Nomadic herders	Expansion	7 856	18 852	42 570	19 041	26 306	43 075	2.24×10^{-04}	2.24×10^{-03}	4.01×10^{-03}	1952	4029	17975
KRG	C. Asia	Nomadic herders	Expansion	8 371	21 654	52 958	21 790	29 367	53 747	3.17×10^{-04}	1.89×10^{-03}	4.44×10^{-03}	2145	4211	17319
KRL	C. Asia	Nomadic herders	Expansion	9 565	20 987	48 619	22 112	28 335	49 188	3.07×10^{-04}	2.06×10^{-03}	3.50×10^{-03}	2340	4693	18147
KRM	C. Asia	Nomadic herders	Expansion	6 237	18 745	45 648	19 059	25 843	46 158	3.78×10^{-04}	2.39×10^{-03}	3.79×10^{-03}	2166	3900	15774
KRT	C. Asia	Nomadic herders	Expansion	10 155	17 658	37 076	19 503	24 383	37 369	1.90×10^{-04}	1.87×10^{-03}	3.85×10^{-03}	1880	3891	20056
LKZ	C. Asia	Nomadic herders	Expansion	12 043	24 633	40 702	21 756	32 592	41 028	2.53×10^{-04}	1.92×10^{-03}	3.39×10^{-03}	2230	4196	20055
OTU	C. Asia	Nomadic herders	Expansion	12 284	19 845	37 976	21 894	27 220	38 229	1.22×10^{-04}	1.43×10^{-03}	3.38×10^{-03}	1952	4011	27337

TKY	C. Asia	Nomadic herders	Expansion	10 072	26 374	40 634	19 937	35 392	42 127	2.66 x10 ⁻⁰⁴	1.99 x10 ⁻⁰³	2.56 x10 ⁻⁰³	2016	4528	23587
TUR	C. Asia	Nomadic herders	Expansion	14 220	22 788	38 084	22 943	28 499	38 319	2.12 x10 ⁻⁰⁴	1.63 x10 ⁻⁰³	3.25 x10 ⁻⁰³	2338	4370	22784
UZA	C. Asia	Nomadic herders	Expansion	10 433	26 458	44 008	28 367	37 678	49 838	1.36 x10 ⁻⁰⁴	1.58 x10 ⁻⁰³	2.66 x10 ⁻⁰³	2832	5698	31930
UZB	C. Asia	Nomadic herders	Expansion	12 713	22 058	35 361	20 656	25 973	35 626	2.49 x10 ⁻⁰⁴	1.82 x10 ⁻⁰³	3.59 x10 ⁻⁰³	2132	3748	18918
UZT	C. Asia	Nomadic herders	Expansion	8 932	22 254	43 579	20 212	27 724	44 070	2.99 x10 ⁻⁰⁴	2.25 x10 ⁻⁰³	3.75 x10 ⁻⁰³	2071	4209	16828

[†]Note that for one population (Ewondo), the Expansion model and a Constant model could not be distinguished from their DIC values (**Table I.C-S2**).

I.D. Y-chromosome microsatellite markers reveal that sex-specific migration patterns impact coalescent-based demographic inferences in Central Asia.

Article en projet.

I.D.i. Abstract.

Studying the current distribution of genetic diversity in human populations has important implications for our understanding of the evolution and history of our species. Here, using both parametric and non-parametric coalescent-based methods, we analyzed a set of linked STR loci from the paternally inherited Y chromosome to infer the past demography of 56 African and Eurasian populations. In agreement with previous studies, we inferred expansion events in most sedentary farmer populations, while we found no significant changes in effective population sizes for both nomadic hunter-gatherers in Africa and semi-nomadic herders in Eurasia. Comparing our results with previous inferences from other type of markers presenting contrasted modes of transmission (maternal, paternal or biparental inheritance) brought complementary insights about the demographic history of populations. In particular, Central Asia stood out as an exception in our previous studies by presenting expansion signals for both sedentary farmers and nomadic herders when analyzing maternally (HVS-I sequences) and biparentally (autosomal microsatellites) inherited markers. Conversely, we showed here expansion signals only for sedentary farmers but not nomadic herders in this area. We suggest that previous inferences for nomadic herders could have been biased by migration processes between these exogamous populations. As Central-Asian populations are patrilocal, this process does not affect the inferences from the Y chromosome, thus explaining our contrasted results between different markers. We did not detect similar sex-specific effects in other studied populations either in the rest of Eurasia or in Africa.

I.D.ii. Introduction.

The demographic history of human populations has been punctuated by several periods of drastic changes in effective population sizes, as well as other complex processes such as selection, migration, admixture, *etc.* (Cavalli-Sforza *et al.*, 1994). As demographic processes are known to leave noticeable footprints on the current distribution of genetic diversity (Menozzi *et al.*, 1978 ; Cavalli-Sforza *et al.*, 1994 ; Beaumont, 1999), reconstructing the demographic history of human populations from contemporary genetic data is a challenge for population geneticists. Together with recent developments in sequencing technologies, which now give access to large multi-locus polymorphism datasets, modern statistical and bioinformatics tools can substantially improve our knowledge of past demographical events (Beaumont, 2004). In particular, the coalescent theory (Kingman, 1982) provides a powerful framework to assess how gene genealogies (*i.e.* “coalescent trees”) are affected by demographic parameters such as the effective population size and the growth rate of populations (Slatkin and Hudson, 1991; Griffiths and Tavaré, 1994). Numerical coalescent-based methods have then been developed, allowing the inference of demographic parameters from molecular data (see Excoffier and Heckel, 2006; Kuhner *et al.*, 2008; Ho and Shapiro, 2011 for reviews).

During the last decades, population genetic studies have provided a substantial contribution to the understanding of the demographic history of *Homo sapiens* (*e.g.* Cavalli-Sforza and Feldman, 2003; Pakendorf and Stoneking, 2005). In particular, the within- and among-population genetic polymorphisms of the mtDNA and nuclear genomes have provided strong evidence for the African origin of modern humans and subsequent expansion throughout the world (Cann *et al.*, 1987; Quintana-Murci *et al.*, 1999; Laval *et al.*, 2010). Other genetic studies have focused on intrapopulation changes in effective population sizes through time, and revealed that demographic expansions started in multiple African and Eurasian populations during the Paleolithic period (*e.g.* Voight *et al.*, 2005; Chaix *et al.*, 2008; Atkinson *et al.*, 2009; Batini *et al.*, 2011). This result challenges the view of paleoanthropologists and archeologists, who largely consider that human expansions began during the Neolithic period along with the processes of sedentarization and the emergence of farming and herding (*e.g.* Bocquet-Appel, 2011). In a previous work (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, and F. Austerlitz, submitted) using HVS-I and autosomal sequences, we inferred much stronger Paleolithic

expansion events for sedentary farmer populations than for nomadic herder populations in Eurasia, and Paleolithic expansions for sedentary farmers but not for nomadic hunter-gatherers in Africa. These differences in demographic patterns between populations with contrasted life-styles seem therefore to predate the emergence of agriculture and the sedentarization processes, thus suggesting that strong expansion events in some Paleolithic populations may have ultimately favored the emergence of farming in these populations. Nevertheless, previous analyses of autosomal microsatellites (C. Aimé, P. Verdu, L. Séguérel, B. Martinez-Cruz, E. Heyer, F. Austerlitz; submitted) also revealed expansion events in sedentary but not in nomadic populations, but these analyses pointed rather toward Neolithic events, probably linked with the emergence of agriculture. These results highlighted the fact that different markers can be affected in different ways by demographic processes.

Here, we investigate whether sex-specific demographic processes could be involved in previous inferences of the demographic history of human populations with contrasted life-styles in Africa and Eurasia (see above). Indeed, Chaix *et al.* (2007) showed contrasted patterns of expansion signals in nomadic herder populations and sedentary farmer populations in Central Asia when considering data from the paternally inherited Y chromosome, but not using a maternally transmitted marker (the hypervariable mitochondrial region HVS-I). Their results from HVS-I data are in agreement with our previous works (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguérel, R. Chaix, L. Quintana-Murci, E. Heyer, and F. Austerlitz, submitted; and C. Aimé, P. Verdu, L. Séguérel, B. Martinez-Cruz, E. Heyer, F. Austerlitz; submitted), as Central Asia stood out as an exception in these studies by presenting no differences in demographic patterns between the two groups. Up to our knowledge, no previous studies have compared demographic inferences from mitochondrial and Y chromosome polymorphism in populations with different life-styles for Africa and the rest of Eurasia.

Differences between inferences from paternally and maternally inherited markers may occur due to sex-specific migration patterns resulting from cultural rules. In particular, some authors have suggested that migration rates and degrees of isolation likely impact demographic inferences, as expansion signals could be attenuated or suppressed in isolated populations (Ray *et al.*, 2003, Excoffier, 2004). Migration rates are generally higher for females than males in human populations due to patrilocality (Seielstad *et al.*, 1998, Chaix *et al.*, 2007). Moreover, the extent of sex-specific differences in migration patterns vary among

populations. For instance, these differences are expected to be higher in exogamous and patrilocal populations than in endogamous and patrilocal populations, as the combination of exogamy and patrilocality increase the migration rates for women (Chaix *et al.*, 2007). As differences in life style are likely to be associated with differences in cultural practices among populations, resulting differences in sex-specific migration patterns could therefore be involved in the inferences of contrasted demographic histories between populations with contrasted life styles.

We present below new coalescent-based inferences from Y chromosome STR data from 21 African populations, 27 Central-Asian populations and eight populations from several other regions in Eurasia (Europe, Middle East, Pamir and East-Asia). We combined a parametric approach to infer demographic parameters assuming several demographic models, and a non-parametric approach (Bayesian Skyline Plots, BSP, Drummond *et al.*, 2005) to visualize the evolution of the effective populations sizes (N_e) through time. These analyses aimed to address the following questions: Do the analyses of Y-chromosome microsatellites reveal contrasted demographic patterns between populations with contrasted life-styles? What kinds of differences are observed between these inferences from Y-chromosome microsatellites and previous inferences from other markers? What do these differences tell us about sex-specific demographic processes in studied areas and their impact on coalescent-based demographic inferences?

I.D.iii. Material and Methods.

Population sampling and marker sets.

For Africa, we analyzed 364 male individuals from 11 sedentary farmer populations and 174 individuals from 10 nomadic hunter-gatherer populations (**Table I.D-S1**). As described by Verdu *et al.* (2013), these individuals were genotyped for six tetranucleotide microsatellites (DYS389I, DYS389II, DYS391, DYS390, DYS393, DYS19) on the non-recombining Y chromosome (NRY), chosen from the Y-chromosome Reference Database (<http://www.yhrd.org/>).

For Eurasia, we combined published data from several studies to analyze five sedentary farmer populations and three nomadic herder populations from Europe, Middle-East, Pamir and East-Asia (**Table I.D-S1**). For Middle East, we compared one sedentary

farmer population, the Turks ($N=124$), and one nomadic herder population, the Kurds ($N=21$). These data come from the study of Quintana-Murci *et al.* (2004). For Pamir, we compared one sedentary farmer population (the Pathans, $N=93$ individuals) and one nomadic herder population (the Balochi, $N=59$) from the study of Qamar *et al.* (2002). Finally, for East Asia, we compare one sedentary farmer population (the Han Chinese, $N=36$) and one nomadic herder population (the Mongols, $N=40$) from the study of Kayser *et al.* (2001). In addition, for comparison with our previous studies, we analyzed two European sedentary farmer populations: Germans ($N=166$) from Henke *et al.* (2001) and Bulgarians ($N=141$) from Zaharova *et al.* (2001). All populations have been genotyped for seven microsatellite loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393).

For our detailed study of Central Asia, we used a data set of 576 individuals from 15 semi-nomadic herder populations and 362 individuals from 12 sedentary farmer populations (**Table S1**). These data include previously published data (Chaix *et al.*, 2004; Heyer *et al.*, 2009; Ségurel *et al.*, 2008). The data for one farmer population (LUZ) and seven herder populations (LKZ, KRB, UZA, UZT, KRL, KRT, TKY) were genotyped for this study under the same condition as in Chaix *et al.* (2004). We used twelve NRY microsatellite loci (DYS19, DYS385I, DYS385II, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439).

Data analyses.

Parametric approach.

We computed the Bayesian posterior distribution of demographic parameters using the MCMC algorithm implemented in BEAST v1.7. (Drummond and Rambaut, 2007). We tested the four demographic models implemented in this software: constant effective population size (“Constant model”), population expansion with a constant growth rate (“Exponential model”), population expansion with a decreasing rate of growth (“Logistic model”) and expansion starting at a given moment in the past from an ancestral population of constant size to reach its current size under an exponential model (“Expansion model”). BEAST estimates composite parameters for each model, namely $N_0\mu$ and g/μ , where N_0 is the current effective population size, g the growth rate and μ the mutation rate. In addition, for the “Expansion model”, the ratio between the current (N_0) and ancestral (N_1) effective population size is also estimated. To infer N_0 and g from these composite parameters, we then need to assume a

value for the mutation rate μ . However, there is no consensus for human Y-chromosome microsatellite mutation rates in the literature, as different methods lead to different estimations. We used here both the pedigree-based rate of 2.1×10^{-3} /generation/site (Heyer *et al.*, 1997) and the phylogeny-based rate of $\mu = 6.9 \times 10^{-4}$ /generation/site (Zhivotovsky *et al.*, 2004). For the populations for which the “Expansion model” best-fitted the data, we inferred the dates of expansion onsets (t) using the following formula: $t = (1/g) \times \ln(N_1/N_0)$, applied to each step of the MCMC algorithm. We used a single step mutational model (SSM), which takes homoplasmy into account. The implementation of microsatellite models in BEAST v1.7.4 is described in Wu and Drummond (2011).

We performed three runs of 6×10^7 steps per population and per demographic model for the African populations, and three runs of 1.2×10^8 steps for the Eurasian populations (which corresponded for all populations to three runs of 10^7 steps per locus). We recorded one tree every 1 000 steps, resulting in a total of 10^5 trees per locus and per run. We then removed the first 10% of each run (burn-in period) and combined the runs to obtain reasonably high effective sample sizes (ESS of 100 or above, see *e.g.* Gignoux *et al.*, 2011). The convergence of these runs was assessed by visual inspection of traces using Tracer v1.5 (Rambaut et Drummond, 2007) to check for concordance between runs, and also by the computation of the Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992) using R v2.14.1 (R Development Core Team, 2011) with the function “gelman.diag” available in the add-on package coda (Plummer *et al.*, 2006). In order to facilitate a large exploration of the parameter space, we chose uniform priors between 1 and 10^5 for N_e and between -0.1 and 0.1 per generation for g . We assumed a generation time of 25 years, permitting the comparison with previous human population genetics studies (*e.g.* Patin *et al.*, 2009; Laval *et al.*, 2010, and our previous studies).

For each population and demographic model, we obtained the mode and the 95% Highest Probability Density (HPD) interval of each parameter, inferred from the posterior distributions obtained using the add-on package Locfit (Loader, 1999) in R v2.14.1. Then, we computed the Deviance Information Criteria (DIC) of each model, as the model with lower DIC is considered as the best-fitting model for the observed data (Spiegelhalter *et al.*, 2002). DIC is computed as $DIC = p_D + \bar{D}$, where p_D represents the effective number of parameters of the model, defined as $\bar{D} - D(\bar{\theta})$. $D(\bar{\theta})$ is the deviance parameter calculated as $-2 \log(p(y|\theta))$, where $p(y|\theta)$ represents the likelihood function, substituted here with the probability of

the data (y) given the posterior mean of the parameter values (θ). \bar{D} is the expectation of the likelihood associated with the means of posterior parameters distributions ($\bar{D} = E^\theta [D(\bar{\theta})]$). To obtain \bar{D} , we repeated all analyses using the posterior mean of each parameter as punctual priors. A difference of five points in DIC was considered as significant (Spiegelhalter *et al.*, 2002; Bolker, 2008).

Non-parametric approach.

Bayesian Skyline Plots (BSPs; Drummond *et al.*, 2005), also implemented in BEAST, estimate demographic changes in a population occurring continuously through time, using the time intervals between successive coalescent events. This method allows visualizing the evolution of N_e through time. As above, we combined three runs of 6×10^7 steps for African populations and three runs of 1.2×10^8 steps for Eurasian populations (which corresponded for all populations to three runs of 10^7 steps per locus) to obtain acceptable ESS (Effective Sample Size) values. We assumed also the same mutation rates as above, and a generation time of 25 years. Outputs were analyzed with Tracer v1.5 to visually check for convergence and ESS, We then used R v2.14.1 to compute Gelman and Rubin's (1992) convergence diagnostic as above, as well as to compute skyline plots. Finally, we used the population growth curves generated from Tracer to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 – 20 generation intervals. We considered that the population increase (or decrease) when both the median and HPD values for N_e increased (or decreased) between more than 2 successive data points. Although this method does not allow providing a confidence interval for inferred expansion timings, this conservative approach ensure that we consider only relevant expansion signals.

I.D.iv. Results.

For Africa, the parametric method showed clear expansion events for all sedentary farmer populations except two of them. Indeed, the “Expansion model” best fitted the data (*i.e.* had the lowest DIC value) for all sedentary populations (**Table I.D-S2, I.D-S3**), and the difference in DIC between the “Expansion model” and the others was higher than 5 points for all these populations except the Ewondo and Nzime. For these two populations, we could indeed not distinguish between an “Expansion model” or a “Constant” model (**Table I.D-S3**), which may be due to lower sample sizes for these populations than for the others (10 and 5

individuals, respectively, **Table I.D-S1**). Conversely, the “Constant model” or the “Exponential model” with negative modal values for growth rates best fitted the data for all nomadic hunter-gatherer populations, thus showing either a constant population size or slight contraction events. As the HPD (High Probability Density) interval for N_e also include 0 for all hunter-gatherer populations, we consider that we inferred no significant demographic changes for these populations. The inferred growth rates for sedentary farmer populations varied between 2.62×10^{-4} [1.31×10^{-8} - 6.71×10^{-4}] and 1.19×10^{-3} [1.52×10^{-4} - 2.79×10^{-3}] per year assuming a mutation rate of $\mu = 6.9 \times 10^{-4}$ /generation/site, or between 7.97×10^{-4} [3.99×10^{-8} - 2.04×10^{-3}] and 3.63×10^{-3} [4.63×10^{-4} - 8.51×10^{-3}] per year assuming a mutation rate $\mu = 2.1 \times 10^{-3}$ /generation/site (**Table I.D-S2**).

As the “expansion model” best fitted the data, we could estimate the expansion onset time under this parametric model. It varied between 40,312 [13,669 – 164,502] years before present (YBP) and 143,270 [41,553 – 275,099] YBP assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$ /generation/site, or between 122,548 [41,007 – 500,086] YBP and 435,540 [124,716 – 836,300] YBP assuming a mutation rate of $\mu = 6.9 \times 10^{-4}$ /generation/site (**Table I.D-1, I.D-S2**). Using the non-parametric method, the BSP graph (**Figure I.D-S1**) showed large HPD intervals including a potential constant population size for all populations. However, when considering the median values of N_e through time, six sedentary farmer populations (G.Fang, Kota, Nzebi, Teke, Tikar and Tsogho) stood out from the others by showing an increase in N_e starting at least 65,000 YBP or 197,600 YBP (**Table I.D-S4**), followed by a stabilization of N_e or a slight contraction since about 30,000 or 90,000 YBP. With both the parametric and the non parametric methods, all datings clearly predated the emergence of farming and the Neolithic transition in Central Africa (**Table I.D-1**).

Table I.D-1. Modes and 95% HPD of expansion onset times (t) inferred from the parametric method, and t inferred from the non-parametric method (BSP), compared with datations of the emergence of farming.

Area	Analyzed populations	Life-style	Emergence of farming or herding	t 95% lower	t mode	t 95% upper	t inferred from BSP plots
Africa	Central-African farmers*	Sedentary farmers	5 000 YBP	30 563 / 92 913	92 749 / 281 958	1 105 239 / 3 359 928	105 683 / 321 277
East-Asia	Han Chinese	Sedentary farmers	9 000 YBP	1 446 / 4 395	18 093 / 55 002	42 822 / 130 178	8 600 / 26 144
Middle-East	Turkish	Sedentary farmers	11 000 YBP	7 404 / 22 508	11 598 / 35 257	17 796 / 54 099	7 975 / 24 244
Pamir	Pathans	Sedentary farmers	8 000 YBP	5 282 / 16 057	9 580 / 29 123	18 127 / 55 106	8 700 / 26 448
Europe	Bulgarians	Sedentary farmers	9 000 YBP	5 006 / 15 218	8 442 / 25 663	15 248 / 46 353	6 800 / 20 672
Europe	Germans	Sedentary farmers	9 000 YBP	5 571 / 16 935	8 321 / 25 295	12 403 / 37 705	6 400 / 19 456
Central-Asia	indo-iranian farmers*	Sedentary farmers	9 000 YBP	4 271 / 12 983	9 491 / 28 855	42 750 / 129 961	9 925 / 30 172

For these area, we report the mean values over populations, considering only populations for which we inferred a signal of expansion. Detailed results for each population are presented in **Table I.D-S2 for the parametric method and **Table I.D-S4** for the non-parametric method. Values in bold are inferred assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$, the others are inferred using $\mu = 6.9 \times 10^{-4}$.*

For Eurasia, the “Expansion model” best fitted the data (**Table I.D-S2**) and the difference in DIC between this model and the others was higher than 5 points (**Table I.D-S3**) for all sedentary farmer populations, indicating a clear expansion signal. Conversely, the “Constant model” best fitted the data for the three nomadic herder populations (**Table I.D-S2, I.D-S3**). The inferred growth rates for sedentary farmer populations varied between 2.06×10^{-4} [1.25×10^{-6} - 6.02×10^{-4}] and 5.43×10^{-4} [3.49×10^{-4} - 7.86×10^{-4}] or between 6.28×10^{-4} [3.79×10^{-6} - 1.83×10^{-3}] and 1.65×10^{-3} [1.06×10^{-3} - 2.39×10^{-3}] per year (**Table I.D-1**). The BSP graph showed stronger expansions (*i.e.* higher N_e and growth rates) for sedentary farmers than nomadic herders (**Figure I.D-2**). Dating for expansion onsets from both the parametric and the non-parametric methods for sedentary farmers are compatible with the Neolithic revolution when assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$ /generation/site, but predate this transition when assuming $\mu = 6.9 \times 10^{-4}$ /generation/site (**Table I.D-1**). Indeed, using the parametric method, we inferred expansions starting between 8,321 [5,571 – 12,403] YBP and 18,093 [1,446 – 42,822] YBP, or between 25,295 [16,935 – 37,705] YBP and 55002 [4,395 – 130,178] (**Table I.D-S2**). Using the non-parametric method, we inferred expansions starting between 6400 YBP and 8700 YBP, or between 19,456 and 26,448 YBP (**Table I.D-S4**).

For Central Asia, we also inferred expansion events for sedentary farmer populations but not for nomadic herder populations using the parametric method. Indeed, the “Expansion model” best fitted the data and the difference in DIC between this model and the others was higher than five points for all sedentary populations except TJY (**Table I.D-1, S2**). Conversely, the “Constant model” or the “Exponential model” with negative modal values for growth rates best fitted the data for all nomadic herder populations (**Table I.D-1, I.D-S2**). The “Constant model” also best fitted the data for the TJY sedentary farmer population. The inferred growth rates for expanding sedentary populations varied between 1.26×10^{-4} [1.32×10^{-6} - 6.15×10^{-3}] and 8.39×10^{-3} [2.03×10^{-4} - 1.26×10^{-2}] or between 3.85×10^{-4} [4.00×10^{-6} - 1.87×10^{-2}] and 2.55×10^{-2} [6.18×10^{-3} - 3.82×10^{-2}] per year (**Table I.D-1**). With regard to the BSP graph, although the HPD intervals were rather wide, at least four populations (TAB, TJE, TJR, and TJT) showed a tendency for expansion (**Figure I.D-3**). These populations have a sedentary life-style based on farming. Again, using the parametric method, dating for expansion onsets for sedentary farmers are compatible with the Neolithic transition when assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$ /generation/site, but predate this transition when assuming $\mu = 6.9 \times 10^{-4}$ /generation/site (**Table I.D-1**). Indeed, we inferred expansion onsets

times varying between 8,245 [3,871 – 25,806] YBP and 10,334 [4,704 – 36,957] YBP or between 25,064 [11,767 – 78,450] YBP and 31,415 [14,300 – 112,349] YBP (**Table I.D-S2**). Using the non-parametric method, we inferred expansions starting between 9,400 YBP and 10,975 YBP or between 28,576 YBP and 33,364 YBP (**Table I.D-S4**), thus before the Neolithic transition.

I.D.v. Discussion.

In this study, we inferred expansion events in most sedentary farmer populations for both Africa and Eurasia using the parametric method, while we inferred no significant changes in effective population sizes for nomadic hunter-gatherers in Africa and nomadic herders in Eurasia and Central Asia. These results are in agreement with our previous inferences from other type of markers (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguirel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz; submitted; and C. Aimé, P. Verdu, L. Séguirel, B. Martinez-Cruz, E. Heyer, F. Austerlitz; submitted), except for Central-Asia. As discussed below, our inferences for Central Asia, together with our previous results, suggest an impact of migration patterns and sex-specific processes on demographic inferences for this area.

From a methodological point of view, the non-parametric method (BSPs) gave less clear results than the parametric method, as the HPD intervals on BSP graphs are very large, thus often not allowing to distinguish between a constant population size or an expansion event. However, the use of Bayesian skyline plots allowed inferring the demographic history of populations with no prior assumption of a particular demographic model. This is especially interesting for analyzing populations for which previous knowledge about demographic history is scarce, which was often the case here (especially for Eurasian nomadic herders). Here, the fact that BSP graphs showed expansions in some sedentary populations but no nomadic populations, together with the agreement between expansion onset times inferred from the parametric and the non-parametric method for these populations, confirm that the assumption of particular demographic models *a priori* seem unlikely to have strongly impacted our conclusions with the parametric method.

Contrasted demographic histories in sedentary and nomadic populations.

Some previous studies have already shown contrasted demographic histories between sedentary and nomadic populations. In particular, Batini *et al.* (2011) showed expansion events for farmers but not hunter-gatherers in Africa using mitochondrial sequences. Our previous works using different type of markers (autosomal sequences, mitochondrial sequences, and autosomal microsatellites) also showed this pattern for Africa and Eurasia, except for Central Asia (C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz; submitted; and C. Aimé, P. Verdu, L. Ségurel, B. Martinez-Cruz, E. Heyer, F. Austerlitz; submitted). This pattern could result from differences in local expansion dynamics and/or the combination of spatial expansion processes at a larger scale and differences in isolation degrees. As shown by Ray *et al.* (2003), expansion signals will be observed in a spatial expansion process if the number of migrants (Nm) is high enough, but not otherwise. For Africa, Verdu *et al.* (2013) showed a higher degree of isolation (higher population-specific F_{ST} values) in hunter-gatherer populations than in farmer populations, which might therefore explain the fact that we did not detect expansion signals for hunter-gatherers. However, as shown in C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz (submitted), although isolation and admixture processes could have some impact on the demographic inferences, these processes seem unlikely to explain alone the contrasted demographic history inferred in populations with different life-styles in Africa and in most areas in Eurasia. The intrinsic demographic growth patterns seem therefore different between sedentary and nomadic populations, except maybe for Central Asia (see below).

As in Batini *et al.* (2011) and in C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz (submitted), our inferred expansions for African farmer populations clearly predated the Neolithic transition and the process of sedentarisation in this area. This is consistent with the “demographic theory” (Sauer, 1952), which suggest that human populations could have started to increase during Paleolithic times, and strong Paleolithic expansions in some populations may have ultimately favored their shift toward agriculture during the Neolithic. Paleolithic demographic expansions could be linked with a rapid environmental change towards a dryer climate (Partridge *et al.*, 1997) and/or to the emergence of new hunting technologies (Mellars, 2006). For instance, the emergence of projectile weapons between 100,000 and 50,000 YBP (Shea, 2009) was probably part of a

broader pattern of ecological diversification of early *Homo sapiens* populations, which may have impacted demographic dynamics.

For Eurasia and Central Asia, the uncertainty in mutation rates did not allow us concluding about the chronology of inferred expansions and the emergence of agriculture. Assuming the phylogeny-based mutation rate, we found expansions predating the Neolithic transition, which is again consistent with some previous studies (Chaix *et al.*, 2008; C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguirel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz, submitted). Conversely, assuming the pedigree-based mutation rate, the inferred expansions were compatible for most populations with the emergence of farming and the process of sedentarization. Accordingly, in C. Aimé, P. Verdu, L. Séguirel, B. Martinez-Cruz, E. Heyer, F. Austerlitz (submitted), we showed using autosomal microsatellite markers that a second strong expansion event may have occurred during the Neolithic period in both African and Eurasian early farmer populations, while we inferred constant population sizes for African nomadic hunter-gatherers and Eurasian nomadic herders. Neolithic expansions in early farmer populations have also been demonstrated by paleoanthropologists and are widely considered to result from sedentarization processes (*e.g.* Bocquet-Appel, 2011). Conversely, nomadic herders and hunter-gatherers share several of the constraints of a nomadic way of life. For instance, birth intervals are generally longer (at least four years) in nomadic populations than in sedentary populations (*e.g.* Short, 1982). According to Bocquet-Appel (2011), these longer birth intervals may be mainly determined by diet differences. Indeed, Vallengia and Ellison (2009) showed that birth interval is mainly determined by the rapidity of post-partum energy recovery, which may be increased by high-carbohydrate food (like cereals) consumption. Moreover a nomadic way of life may offer less food security than sedentary farming, to the latter being more prone to efficient long term food storage. In addition, competition between farmers and nomadic populations for land may have led to a reduction in, and fragmentation of, available habitat and resources for nomads, thereby limiting their possibilities of expansion (Diamond, 2002).

Impact of sex-specific processes on demographic inferences.

Interestingly, our results for Central Asia are not consistent with our previous inferences from maternally and bi-parentally transmitted markers. Indeed, in both C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguirel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz

(submitted) and C. Aimé, P. Verdu, L. Ségurel, B. Martinez-Cruz, E. Heyer, F. Austerlitz (submitted), we found signals of demographic expansions for both sedentary farmers and semi-nomadic herders in this area. Conversely, we found here expansion events for Central Asian sedentary farmers but not for Central Asian nomadic herders, like in the rest of Eurasia. This pattern may result from sex-specific migration patterns of this area. Conversely, our results for Africa and the rest of Eurasia are consistent with our previous studies by showing expansions in sedentary but not in nomadic populations. Sex-specific demographic processes do not seem, therefore, to have affected our inferences for these areas.

As Central-Asian herder populations are mostly exogamous (*i.e.* marriages occurred between individuals from different clans or lineages) while farmer populations are mostly endogamous (*i.e.* marriages occurred between individuals from different clans or lineages), one can expect more gene flow among herder populations than among farmer populations (Chaix *et al.*, 2007). Accordingly, previous studies showed significantly lower F_{ST} values for the herders than for the farmers in Central Asia (Heyer *et al.*, 2009; C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz, submitted). Chaix *et al.* (2007) also found higher levels of mitochondrial genetic diversity in herder populations than in farmer populations. Moreover, as Central Asian populations are patrilocal, migration rates are expected to be higher for women than men, thus potentially resulting to contrasted impact of migration processes on inferences from mitochondrial DNA or the Y chromosome. Indeed, sex-specific migration processes may have impacted demographic inferences in two different ways. First, farmers and herders may both have experimented expansion events, but the remaining signal of these expansions for herder populations is stronger in the current polymorphism of maternally and bi-parentally markers as compared to the Y chromosome, due to lower migration rate for women than men. Indeed, expansion signals can be attenuated or suppressed due to isolation and a low number of immigrants (less than 20; Ray *et al.*, 2003; Excoffier, 2004). However, as farmer populations are more isolated than herder populations (Heyer *et al.*, 2009; C. Aimé, G. Laval, E. Patin, P. Verdu, L. Ségurel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz, submitted), this hypothesis is not supported by our inferences of expansion event for farmers using both the Y chromosome, the HVS-I region and autosomal microsatellite data. Alternatively, we suggest here that higher migration rates in herder than farmer populations, by increasing the proportion of rare alleles, may have led to spurious expansion signals in herder populations in our previous studies. As

migration rates are much lower for men than women, this hypothesis also explains the fact that we didn't detect any expansion event for herders from Y chromosome data.

Possible confounding factors.

Our approach makes the assumption that populations are isolated and panmictic, which is questionable for human populations. However, we analyzed a large set of populations sampled in very distant geographical regions (*i.e.* Central Africa, East Africa, Europe, Middle-East, Central Asia, Pamir, Siberia and East-Asia). The main conclusions of this study rely on consistent patterns between most of these areas, and it seems unlikely that processes such as admixture could have biased the estimates similarly everywhere. Moreover, in Central Africa, several studies have shown that hunter-gatherer populations show signals of admixture, while it is not the case for farmer populations (Patin *et al.*, 2009; Verdu *et al.*, 2009, 2013). If this introgression had been strong enough, this may have yielded a spurious expansion signal in the hunter-gatherer populations, which is not what we observed here.

Similarly, potential selection occurring on the whole NRY region seems unlikely to have impacted in the same way all the studied populations within each group (*e.g.* stronger positive selection on sedentary than on nomadic populations), as we analyzed different nomadic and sedentary populations living near each other, in several geographically distant areas. Moreover, as explained above, our result of different expansion patterns between sedentary and nomadic population have been also demonstrated using neutral markers in previous studies (*e.g.* C. Aimé, G. Laval, E. Patin, P. Verdu, L. Séguérel, R. Chaix, L. Quintana-Murci, E. Heyer, F. Austerlitz; submitted).

I.D.vi. Supplementary Figures.

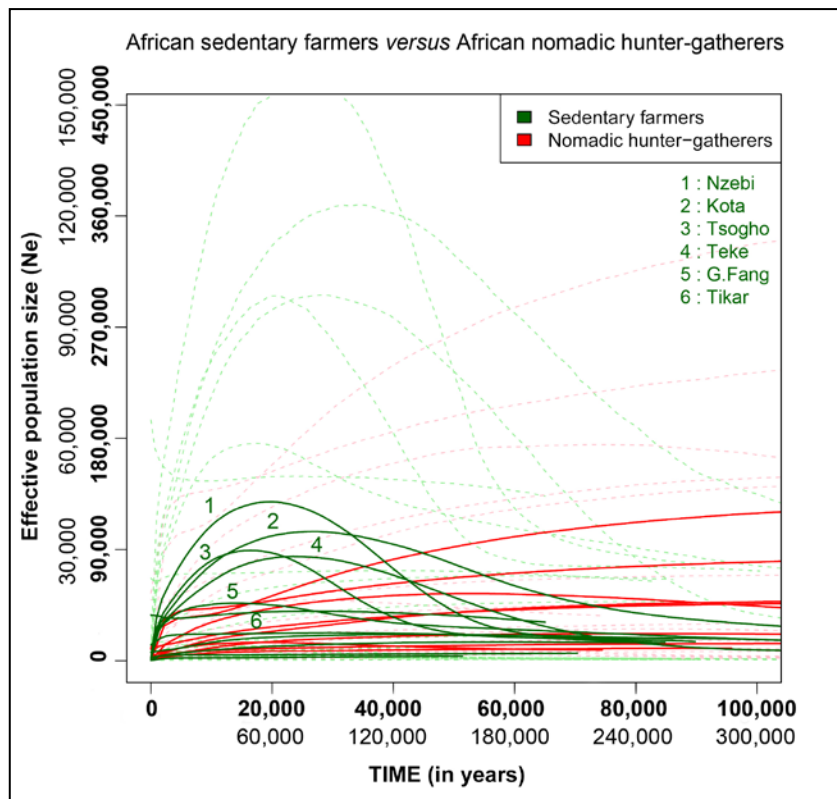


Figure I.D-S1. Bayesian Skyline Plot inferred from Y-chromosome microsatellites markers in African sedentary farmers versus African nomadic hunter-gatherers. The values indicated in bold on the axes are obtained assuming a mutation rate of $\mu=2.1 \times 10^{-3}$ / generation /site (Heyer et al., 1997), and the other values correspond to $\mu=6.3 \times 10^{-4}$ / generation /site (Zhivotovsky et al., 2004). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time is represented backward on the X axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines.

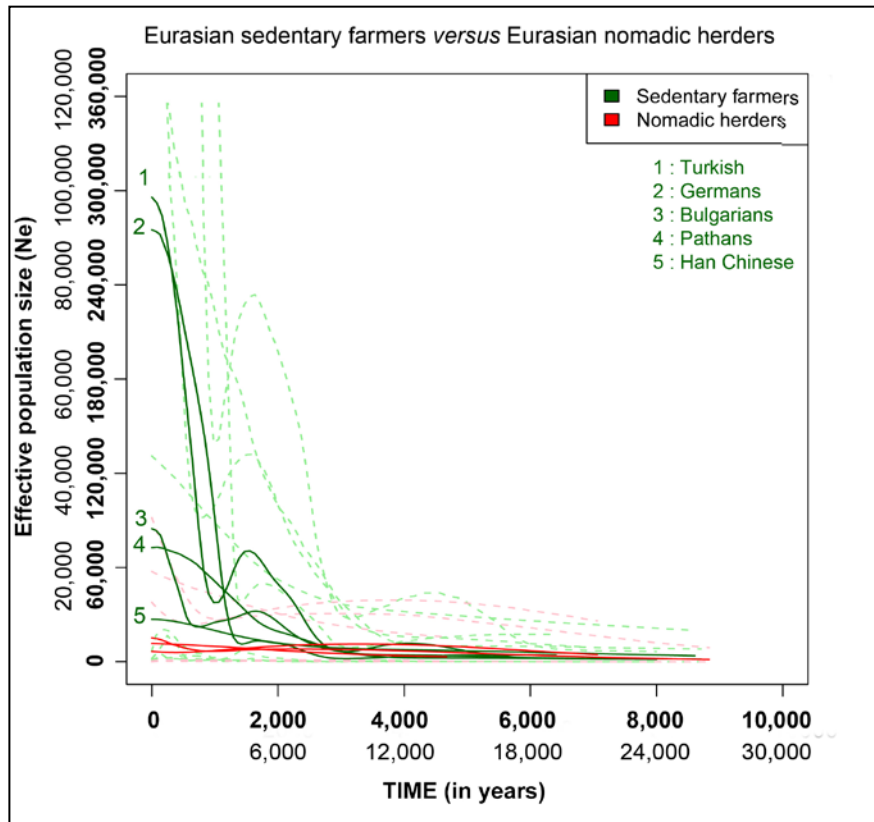


Figure I.D-S2. Bayesian Skyline Plot inferred from Y-chromosome microsatellites markers in Eurasian sedentary farmers versus Eurasian nomadic hunter-gatherers. *The values indicated in bold on the axes are obtained assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$ / generation /site (Heyer et al., 1997), and the other values correspond to $\mu = 6.3 \times 10^{-4}$ / generation /site (Zivotovsky et al., 2004). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time is represented backward on the X axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines.*

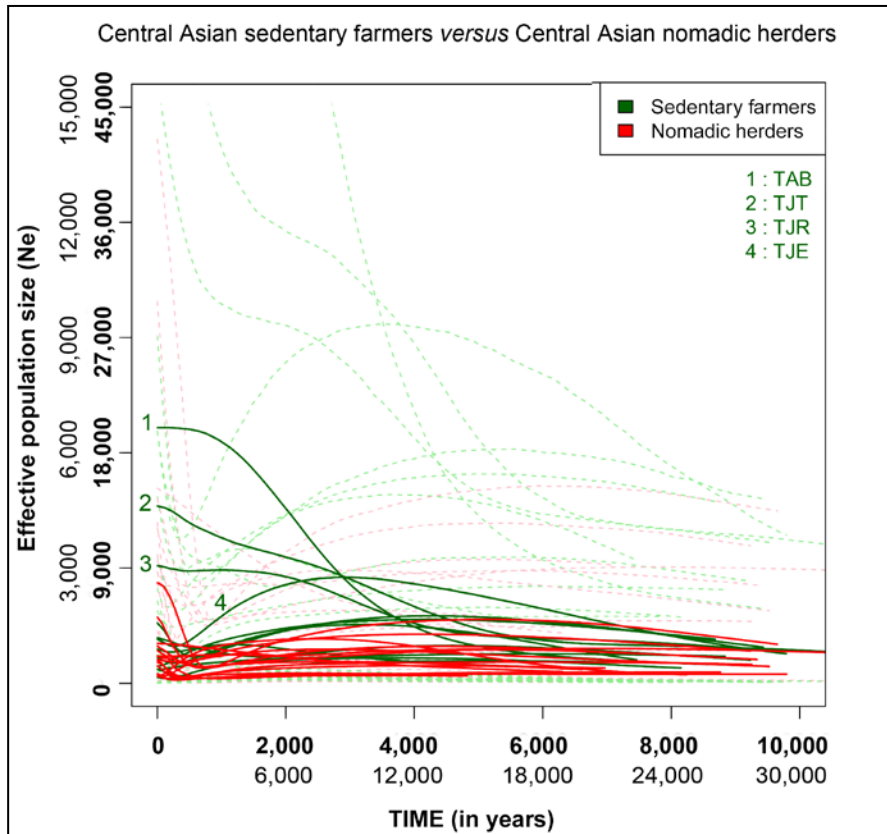


Figure I.D-S3. Bayesian Skyline Plot inferred from Y-chromosome microsatellites markers in Central Asian sedentary farmers versus Central Asian nomadic hunter-gatherers. The values indicated in bold on the axes are obtained assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$ / generation /site (Heyer et al., 1997), and the other values correspond to $\mu = 6.3 \times 10^{-4}$ / generation /site (Zhitovovskiy et al., 2004). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time is represented backward on the X axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines.

I.D.vii. Supplementary Tables.

Table I.D-S1. Description and location of studied populations.

Population	code	Area	Country	Life-style	number of markers	number of samples	Reference
Akele	–	Central Africa	Gabon	Sedentary farmers	6	12	Verdu <i>et al.</i> , 2013
Bangando	–	Central Africa	Cameroon	Sedentary farmers	6	24	Verdu <i>et al.</i> , 2013
C. Fang	–	Central Africa	Cameroon	Sedentary farmers	6	17	Verdu <i>et al.</i> , 2013
Ewondo	–	Central Africa	Cameroon	Sedentary farmers	6	10	Verdu <i>et al.</i> , 2013
G. Fang	–	Central Africa	Gabon	Sedentary farmers	6	60	Verdu <i>et al.</i> , 2013
Kota	–	Central Africa	Gabon	Sedentary farmers	6	53	Verdu <i>et al.</i> , 2013
Nzebi	–	Central Africa	Gabon	Sedentary farmers	6	57	Verdu <i>et al.</i> , 2013
Nzime	–	Central Africa	Cameroon	Sedentary farmers	6	5	Verdu <i>et al.</i> , 2013
Teke	–	Central Africa	Gabon	Sedentary farmers	6	48	Verdu <i>et al.</i> , 2013
Tikar	–	Central Africa	Cameroon	Sedentary farmers	6	19	Verdu <i>et al.</i> , 2013
Tsogho	–	Central Africa	Gabon	Sedentary farmers	6	60	Verdu <i>et al.</i> , 2013
Bezan	–	Central Africa	Cameroon	Nomadic HG	6	24	Verdu <i>et al.</i> , 2013
C. Baka	–	Central Africa	Cameroon	Nomadic HG	6	10	Verdu <i>et al.</i> , 2013
C.Bongo	–	Central Africa	Gabon	Nomadic HG	6	10	Verdu <i>et al.</i> , 2013
E. Bongo	–	Central Africa	Gabon	Nomadic HG	6	19	Verdu <i>et al.</i> , 2013
E. Baka	–	Central Africa	Cameroon	Nomadic HG	6	29	Verdu <i>et al.</i> , 2013
G.Baka	–	Central Africa	Gabon	Nomadic HG	6	25	Verdu <i>et al.</i> , 2013
Kola	–	Central Africa	Cameroon	Nomadic HG	6	17	Verdu <i>et al.</i> , 2013
Koya	–	Central Africa	Gabon	Nomadic HG	6	16	Verdu <i>et al.</i> , 2013

S.Bongo	–	Central Africa	Gabon	Nomadic HG	6	14	Verdu <i>et al.</i> , 2013
S.Baka	–	Central Africa	Cameroon	Nomadic HG	6	14	Verdu <i>et al.</i> , 2013
Bulgarians	–	Eurasia	Bulgaria	Sedentary farmers	7	140	Zaharova <i>et al.</i> , 2001
Germans	–	Eurasia	Germany	Sedentary farmers	7	165	Henke <i>et al.</i> , 2001
H.Chinese	–	Eurasia	China	Sedentary farmers	7	36	Kayser <i>et al.</i> , 2001
Pathans	–	Eurasia	Pakistan	Sedentary farmers	7	93	Qamar <i>et al.</i> , 2002
Turkish	–	Eurasia	Turkey	Sedentary farmers	7	124	Quintana-Murci <i>et al.</i> , 2004
Balochi	–	Eurasia	Pakistan	Nomadic herders	7	59	Qamar <i>et al.</i> , 2002
Kurds	–	Eurasia	Iran	Nomadic herders	7	21	Quintana-Murci <i>et al.</i> , 2004
Mongols	–	Eurasia	Mongolia	Nomadic herders	7	40	Kayser <i>et al.</i> , 2001
Tajiko-uzbeks (Bukhara)	LUZ	Central-Asia	uzbekistan	Sedentary farmers	12	31	This study
Tajiks (Boukhara)	TAB	Central-Asia	uzbekistan	Sedentary farmers	12	21	Ségurel <i>et al.</i> , 2008
Tajiks (Penjikent)	TDS	Central-Asia	Tajikistan/ Uzbekistan border	Sedentary farmers	12	31	Ségurel <i>et al.</i> , 2008
Tajiks (Penjikent)	TDU	Central-Asia	Tajikistan/ Uzbekistan border	Sedentary farmers	12	31	Ségurel <i>et al.</i> , 2008
Tajiks (Samarkand)	TJA	Central-Asia	Uzbekistan/ Tajikistan border	Sedentary farmers	12	32	Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJE	Central-Asia	Tajikistan	Sedentary farmers	12	27	Ségurel <i>et al.</i> , 2008
Tajiks (Ferghana)	TJK	Central-Asia	Uzbekistan/ Kyrgyzstan border	Sedentary farmers	12	35	Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJN	Central-Asia	Tajikistan	Sedentary farmers	12	30	Ségurel <i>et al.</i> , 2008
Tajiks (Ferghana)	TJR	Central-Asia	Uzbekistan/ Kyrgyzstan border	Sedentary farmers	12	29	Heyer <i>et al.</i> , 2009
Tajiks (Gharm)	TJT	Central-Asia	Tajikistan	Sedentary farmers	12	30	Ségurel <i>et al.</i> , 2008

Tajiks (Samarkand)	TJU	Central-Asia	Uzbekistan/ Tajikistan border	Sedentary farmers	12	29	Heyer <i>et al.</i> , 2009
Tajiks (Yagnobs from Dushanbe)	TJY	Central-Asia	Tajikistan	Sedentary farmers	12	36	Ségurel <i>et al.</i> , 2008
Karakalpaks (Karakalpokia)	KAR	Central-Asia	Uzbekistan	Nomadic herders	12	54	Chaix <i>et al.</i> , 2004
Kazaks (Karakalpokia)	KAZ	Central-Asia	Uzbekistan	Nomadic herders	12	50	Chaix <i>et al.</i> , 2004
Kyrgyz (Andijan)	KRA	Central-Asia	Uzbekistan/ Kyrgyzstan border	Nomadic herders	12	46	Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRB	Central-Asia	Kyrgyzstan	Nomadic herders	12	31	This study
Kyrgyz (Narin)	KRG	Central-Asia	Kyrgyzstan	Nomadic herders	12	20	Heyer <i>et al.</i> , 2009
Kyrgyz (Narin)	KRL	Central-Asia	Kyrgyzstan	Nomadic herders	12	40	This study
Kyrgyz (Narin)	KRM	Central-Asia	Kyrgyzstan	Nomadic herders	12	22	Heyer <i>et al.</i> , 2009
Kyrgyz (Issyk Kul)	KRT	Central-Asia	Kyrgyzstan	Nomadic herders	12	37	This study
Kazaks (Bukhara)	LKZ	Central-Asia	Uzbekistan	Nomadic herders	12	20	This study
Karakalpaks (Karakalpokia)	OTU	Central-Asia	Uzbekistan	Nomadic herders	12	54	Chaix <i>et al.</i> , 2004
Turkmen (Urgench)	TKY	Central-Asia	Kyrgyzstan	Nomadic herders	12	31	This study
Turkmen (Karakalpokia)	TUR	Central-Asia	Kyrgyzstan	Nomadic herders	12	51	Chaix <i>et al.</i> , 2004
Uzbeks (Ferghana)	UZA	Central-Asia	Uzbekistan/ Kyrgyzstan border	Nomadic herders	12	36	This study
Uzbeks (Karakalpokia)	UZB	Central-Asia	Uzbekistan	Nomadic herders	12	49	Chaix <i>et al.</i> , 2004
Uzbeks (Penjikent)	UZT	Central-Asia	Tajikistan/ Uzbekistan border	Nomadic herders	12	35	This study

Table LD-S2. Modes and 95% HPD of current effective population size (N_0), ancestral effective population size, per year growth rate (g) and expansion onset (in years BP) estimated under the best-fitting (BF) model for each population.

Population	Area	Life style	BF model	N_1 95% lower	N_1 mode	N_1 95% upper	N_0 95% lower	N_0 mode	N_0 95% upper	g 95% lower	g mode	g 95% upper	t 95% lower	t mode	t 95% upper
Akele	Africa	Sedentary farmers	Expansion	200/ 609	1866/ 5671	11502/ 34966	1254/ 3811	4663/ 14175	12231/ 37183	1.40x10 ⁻⁰⁵ / 4.25x10⁻⁰⁵	4.50x10 ⁻⁰⁴ / 1.37x10⁻⁰³	1.23x10 ⁻⁰³ / 3.72x10⁻⁰³	10409 / 31643	30235 / 91915	226306 / 687970
Bangando	Africa	Sedentary farmers	Expansion	263/ 799	648/ 1969	1380/ 4195	358/ 1089	882/ 2682	1880/ 5715	1.10x10 ⁻⁰⁴ / 3.34x10⁻⁰⁴	8.97x10 ⁻⁰⁴ / 2.73x10⁻⁰³	1.99x10 ⁻⁰³ / 6.05x10⁻⁰³	6932 / 21073	15891 / 48309	75360 / 229095
C. Fang	Africa	Sedentary farmers	Expansion	295/ 898	3259/ 9908	12401/ 37699	1665/ 5063	5056/ 15371	13060/ 39702	6.30x10 ⁻⁰⁷ / 1.92x10⁻⁰⁶	4.24x10 ⁻⁰⁴ / 1.29x10⁻⁰³	1.15x10 ⁻⁰³ / 3.49x10⁻⁰³	8294 / 25215	30799 / 93630	337389 / 1025663
Ewondo*	Africa	Sedentary farmers	Expansion	15 /45	554/ 1685	2945/ 8953	32/ 98	1237/ 3762	6574/ 19985	1.52x10 ⁻⁰⁴ / 4.63x10⁻⁰⁴	1.19x10 ⁻⁰³ / 3.63x10⁻⁰³	2.80x10 ⁻⁰³ / 8.51x10⁻⁰³	4496 / 13669	13260 / 40312	54113 / 164502
Ewondo*	Africa	Sedentary farmers	Constant	–	–	–	129/ 393	506/ 1538	1215/ 3695	–	–	–	–	–	–
G. Fang	Africa	Sedentary farmers	Expansion	1367/ 4156	1989/ 6047	3034/ 9223	1610/ 4895	3330/ 10123	5218/ 15863	1.00x10 ⁻⁰⁵ / 3.05x10⁻⁰⁵	4.24x10 ⁻⁰⁴ / 1.29x10⁻⁰³	1.13x10 ⁻⁰³ / 3.43x10⁻⁰³	11012 / 33477	28210 / 85759	265860 / 808213
Kota	Africa	Sedentary farmers	Expansion	2433/ 7396	3592/ 10921	5465/ 16612	2869/ 8722	4236/ 12879	6444/ 19590	2.80x10 ⁻⁰⁶ / 8.50x10⁻⁰⁶	3.03x10 ⁻⁰⁴ / 9.22x10⁻⁰⁴	7.97x10 ⁻⁰⁴ / 2.42x10⁻⁰³	16911 / 51408	40593 / 123404	530732 / 1613425
Nzebi	Africa	Sedentary farmers	Expansion	0 / 0	3650/ 11095	6117/ 18596	3071/ 9336	4777/ 14522	8007/ 24341	3.42x10 ⁻⁰⁸ / 1.04x10⁻⁰⁷	3.41x10 ⁻⁰⁴ / 1.04x10⁻⁰³	9.50x10 ⁻⁰⁴ / 2.89x10⁻⁰³	15979 / 48575	44383 / 134925	605740 / 1841450
Nzime*	Africa	Sedentary farmers	Expansion	33/ 100	1014/ 3083	10235/ 31115	33/ 101	3231/ 9822	11182/ 33992	2.17x10 ⁻⁰⁵ / 6.59x10⁻⁰⁵	6.01x10 ⁻⁰⁴ / 1.83x10⁻⁰³	1.55x10 ⁻⁰³ / 4.70x10⁻⁰³	7055 / 21446	23937 / 72768	149897 / 455686
Nzime*	Africa	Sedentary farmers	Constant	–	–	–	33/ 99	1815/ 5517	5466/ 16616	–	–	–	–	–	–
Teke	Africa	Sedentary farmers	Expansion	2213/ 6729	3473/ 10558	5588/ 16987	2770/ 8422	4676/ 14215	6993/ 21260	1.31x10 ⁻⁰⁸ / 3.99x10⁻⁰⁸	2.62x10 ⁻⁰⁴ / 7.97x10⁻⁰⁴	6.71x10 ⁻⁰⁴ / 2.04x10⁻⁰³	13495 / 41025	47128 / 143270	904933 / 2750995

Tikar	Africa	Sedentary farmers	Expansion	266/ 808	1934/ 5880	12438/ 37813	2276/ 6921	5693/ 17305	14127/ 42945	3.36x10 ⁻⁰⁸ / 1.02x10⁻⁰⁷	3.31x10 ⁻⁰⁴ / 1.01x10⁻⁰³	1.02x10 ⁻⁰³ / 3.11x10⁻⁰³	8630/ 26235	33057/ 100494	542750/ 1649961
Tsogho	Africa	Sedentary farmers	Expansion	1486/ 4516	2245/ 6826	3347/ 10175	1735/ 5276	2623/ 7975	3910/ 11887	4.75x10 ⁻⁰⁶ / 1.44x10⁻⁰⁵	5.04x10 ⁻⁰⁴ / 1.53x10⁻⁰³	1.12x10 ⁻⁰³ / 3.41x10⁻⁰³	7379/ 22433	28111/ 85458	306143/ 930674
Bezan*	Africa	Nomadic HG	Constant	—	—	—	1299/ 3949	2106/ 6401	3420/ 10398	—	—	—	—	—	—
Bezan*	Africa	Nomadic HG	Exponential	—	—	—	107/ 327	401/ 1218	2625/ 7980	-3.80x10- 06/ - 1.16x10-05	-1.48x10-06/ -4.51x10-06	3.86x10 ⁻⁰⁸ / 1.17x10⁻⁰⁷	—	—	—
C. Baka	Africa	Nomadic HG	Exponential	—	—	—	183/ 557	1128/ 3429	3498/ 10634	-5.39x10- 06/ - 1.64x10-05	-6.49x10-07/ -1.97x10-06	3.74x10 ⁻⁰⁸ / 1.14x10⁻⁰⁷	—	—	—
C.Bongo	Africa	Nomadic HG	Constant	—	—	—	396/ 1204	2490/ 7569	8231/ 25023	—	—	—	—	—	—
E. Bongo	Africa	Nomadic HG	Constant	—	—	—	1147/ 3488	2263/ 6879	4495/ 13665	—	—	—	—	—	—
E. Baka	Africa	Nomadic HG	Exponential	—	—	—	1631/ 4959	2674/ 8128	4527/ 13762	-2.11x10- 06/ - 6.42x10-06	-4.00x10-07/ -1.22x10-06	1.04x10 ⁻⁰⁷ / 3.16x10⁻⁰⁷	—	—	—
G.Baka	Africa	Nomadic HG	Exponential	—	—	—	2256/ 6857	3232/ 9825	8914/ 27099	-1.57x10- 06/ - 4.78x10-06	-1.64x10-07/ -5.00x10-07	1.33x10 ⁻⁰⁸ / 4.06x10⁻⁰⁸	—	—	—
Kola	Africa	Nomadic HG	Constant	—	—	—	811/ 2467	1591/ 4836	2876/ 8743	—	—	—	—	—	—
Koya	Africa	Nomadic HG	Constant	—	—	—	1014/ 3082	1839/ 5591	3380/ 10274	—	—	—	—	—	—
S.Bongo	Africa	Nomadic HG	Constant	—	—	—	1478/ 4493	3168/ 9630	7907/ 24037	—	—	—	—	—	—
S.Baka	Africa	Nomadic HG	Constant	—	—	—	524/ 2372/	2372/ 4817/	4817/ —	—	—	—	—	—	—

H.Chinese	Eurasia	Sedentary farmers	Expansion	0/ 0	101/ 307	1274/ 3872	821/ 2495	4601/ 13987	13048/ 39665	1.25×10^{-06} / 3.79×10^{-06}	2.06×10^{-04} / 6.28×10^{-04}	6.01×10^{-04} / 1.83×10^{-03}	476/ 1446	5952/ 18093	14086/ 42822
Turkish	Eurasia	Sedentary farmers	Expansion	44/ 134	228/ 694	715/ 2175	5842/ 17760	11359/ 34532	16080/ 48884	2.23×10^{-04} / 6.79×10^{-04}	3.72×10^{-04} / 1.13×10^{-03}	5.59×10^{-04} / 1.70×10^{-03}	2436/ 7404	3815/ 11598	5854/ 17796
Pathans	Eurasia	Sedentary farmers	Expansion	33/ 100	180/ 548	811/ 2465	3127/ 9505	10002/ 30406	15479/ 47056	1.87×10^{-04} / 5.68×10^{-04}	4.04×10^{-04} / 1.23×10^{-03}	6.89×10^{-04} / 2.09×10^{-03}	1737/ 5282	3151/ 9580	5963/ 18127
Germans	Eurasia	Sedentary farmers	Expansion	24/ 72	159/ 484	432/ 1312	6998/ 21274	13172/ 40042	16378/ 49790	3.47×10^{-04} / 1.06×10^{-03}	5.44×10^{-04} / 1.65×10^{-03}	7.87×10^{-04} / 2.39×10^{-03}	1832/ 5571	2737/ 8321	4080/ 12403
Bulgarians	Eurasia	Sedentary farmers	Expansion	14/ 42	163/ 495	595/ 1808	4011/ 12193	10711/ 32562	15765/ 47926	2.42×10^{-04} / 7.36×10^{-04}	5.05×10^{-04} / 1.53×10^{-03}	7.89×10^{-04} / 2.40×10^{-03}	1647/ 5006	2777/ 8442	5016/ 15248
Balochi	Eurasia	Nomadic herders	Constant	–	–	–	394/ 1197	539/ 1639	744/ 2261	–	–	–	–	–	–
Kurds	Eurasia	Nomadic herders	Constant	–	–	–	207/ 630	351/ 1067	595/ 1809	–	–	–	–	–	–
Mongols	Eurasia	Nomadic herders	Constant	–	–	–	50/ 153	203/ 618	1591/ 4837	–	–	–	–	–	–
LUZ	Central-Asia	Sedentary farmers	Expansion	0/ 0	342/ 1039	6854/ 20835	50/ 153	1553/ 4723	8494/ 25821	2.53×10^{-04} / 7.69×10^{-04}	4.25×10^{-03} / 1.29×10^{-02}	1.05×10^{-02} / 3.18×10^{-02}	1467/ 4461	3312/ 10068	20624/ 62696
TAB	Central-Asia	Sedentary farmers	Expansion	0/ 0	187/ 570	2810/ 8541	680/ 2068	4766/ 14487	14273/ 43389	1.32×10^{-06} / 4.00×10^{-06}	4.78×10^{-04} / 1.45×10^{-03}	3.27×10^{-03} / 9.93×10^{-03}	1322/ 4020	3076/ 9351	16285/ 49507
TDS	Central-Asia	Sedentary farmers	Expansion	0 / 0	104/ 316	4746/ 14429	44/ 135	657/ 1998	5948/ 18083	2.03×10^{-03} / 6.18×10^{-03}	8.40×10^{-03} / 2.55×10^{-02}	1.25×10^{-02} / 3.81×10^{-02}	1274/ 3871	2712/ 8245	8489/ 25806
TDU	Central-Asia	Sedentary farmers	Expansion	0 / 0	243/ 739	6649/ 20214	38/ 115	1327/ 4034	8364/ 25425	6.77×10^{-04} / 2.06×10^{-03}	4.73×10^{-03} / 1.44×10^{-02}	1.07×10^{-02} / 3.25×10^{-02}	1373/ 4173	3231/ 9822	15783/ 47979
TJA	Central-Asia	Sedentary farmers	Expansion	0 / 0	147/ 448	6605/ 20080	36/ 108	1235/ 3754	9660/ 29365	4.26×10^{-04} / 1.30×10^{-03}	4.77×10^{-03} / 1.45×10^{-02}	1.23×10^{-02} / 3.73×10^{-02}	1412/ 4292	2882/ 8762	12973/ 39438

TJE	Central-Asia	Sedentary farmers	Expansion	0 / 0	235/ 714	6445/ 19594	37/ 111	1414/ 4299	8361/ 25416	9.06x10 ⁻⁰⁴ / 2.76x10⁻⁰³	5.37x10 ⁻⁰³ / 1.63x10⁻⁰²	1.17x10 ⁻⁰² / 3.56x10⁻⁰²	1346/ 4091	3101/ 9428	14036/ 42670
TJK	Central-Asia	Sedentary farmers	Expansion	0 / 0	172/ 524	6808/ 20696	64/ 196	1230/ 3738	9383/ 28524	7.08x10 ⁻⁰⁴ / 2.15x10⁻⁰³	5.36x10 ⁻⁰³ / 1.63x10⁻⁰²	1.11x10 ⁻⁰² / 3.36x10⁻⁰²	1440/ 4377	3285/ 9987	15648/ 47570
TJN	Central-Asia	Sedentary farmers	Expansion	0 / 0	216/ 656	6874/ 20896	58/ 175	1510/ 4592	9164/ 27860	1.20x10 ⁻⁰³ / 3.66x10⁻⁰³	5.20x10 ⁻⁰³ / 1.58x10⁻⁰²	1.09x10 ⁻⁰² / 3.32x10⁻⁰²	1547/ 4704	3399/ 10334	12157/ 36957
TJR	Central-Asia	Sedentary farmers	Expansion	0 / 0	118/ 359	3337/ 10145	52/ 159	2160/ 6567	10123/ 30773	1.32x10 ⁻⁰⁶ / 4.00x10⁻⁰⁶	1.27x10 ⁻⁰⁴ / 3.85x10⁻⁰⁴	6.14x10 ⁻⁰³ / 1.87x10⁻⁰²	1468/ 4462	3201/ 9732	12879/ 39153
TJT	Central-Asia	Sedentary farmers	Expansion	0 / 0	236/ 717	6827/ 20753	34/ 103	1464/ 4451	8850 / 26904	9.45x10 ⁻⁰⁴ / 2.87x10⁻⁰³	5.55x10 ⁻⁰³ / 1.69x10⁻⁰²	1.18x10 ⁻⁰² / 3.60x10⁻⁰²	1393/ 4235	3192/ 9703	13976/ 42487
TJU	Central-Asia	Sedentary farmers	Expansion	0 / 0	252/ 765	6566/ 19959	63/ 190	1376/ 4185	8360 / 25413	1.31x10 ⁻⁰³ / 3.99x10⁻⁰³	8.33x10 ⁻⁰³ / 2.53x10⁻⁰²	1.25x10 ⁻⁰² / 3.80x10⁻⁰²	1413/ 4296	2953/ 8978	11839/ 35991
TJY	Central-Asia	Sedentary farmers	Constant	-	-	-	83/ 253	130/ 394	207/ 629	-	-	-	-	-	-
KAR	Central-Asia	Nomadic herders	Constant	-	-	-	259/ 788	353/ 1073	487/ 1482	-	-	-	-	-	-
KAZ	Central-Asia	Nomadic herders	Constant	-	-	-	134/ 407	190/ 579	273/ 831	-	-	-	-	-	-
KRA*	Central-Asia	Nomadic herders	Constant	-	-	-	133/ 405	191/ 582	275/ 836	-	-	-	-	-	-
KRA*	Central-Asia	Nomadic herders	Exponentiel	-	-	-	119/ 361	182/ 554	293/ 891	-3.11x10 ⁻⁰⁵ / -9.46x10⁻⁰⁵	-2.61x10 ⁻⁰⁷ / -7.95x10⁻⁰⁷	3.32x10 ⁻⁰⁵ / 1.01x10⁻⁰⁴	-	-	-
KRB	Central-Asia	Nomadic herders	Constant	-	-	-	49/ 150	84/ 254	135/ 410	-	-	-	-	-	-
KRG	Central-Asia	Nomadic herders	Constant	-	-	-	37/ 113	65/ 199	120/ 365	-	-	-	-	-	-
KRL	Central-Asia	Nomadic herders	Constant	-	-	-	75/ 229	120/ 364	185/ 564	-	-	-	-	-	-

KRM	Central-Asia	Nomadic herders	Exponentiel	–	–	–	41/ 125	88/ 266	181/ 552	-5.38x10 ⁻⁰⁵ / -1.64x10 ⁻⁰⁴	-1.31x10 ⁻⁰⁵ / -3.97x10 ⁻⁰⁵	4.11x10 ⁻⁰⁵ / 1.25x10 ⁻⁰⁴	–	–	–
KRT	Central-Asia	Nomadic herders	Exponentiel	–	–	–	56/ 171	98/ 299	171/ 519	-5.19x10 ⁻⁰⁵ / -1.58x10 ⁻⁰⁴	-9.34x10 ⁻⁰⁶ / -2.84x10 ⁻⁰⁵	2.87x10 ⁻⁰⁵ / 8.73x10 ⁻⁰⁵	–	–	–
LKZ	Central-Asia	Nomadic herders	Constant	–	–	–	107/ 326	234/ 710	484/ 1473	–	–	–	–	–	–
OTU	Central-Asia	Nomadic herders	Constant	–	–	–	187/568	258/ 783	361/ 1098	–	–	–	–	–	–
TKY	Central-Asia	Nomadic herders	Exponentiel	–	–	–	165/ 502	262/ 798	428/ 1302	–	–	–	–	–	–
TUR	Central-Asia	Nomadic herders	Constant	–	–	–	131/ 397	187/ 568	268/ 815	–	–	–	–	–	–
UZA	Central-Asia	Nomadic herders	Constant	–	–	–	489/ 1487	703/ 2136	1028/ 3126	–	–	–	–	–	–
UZB	Central-Asia	Nomadic herders	Constant	–	–	–	400/ 1217	549/ 1670	771/ 2343	–	–	–	–	–	–
UZT	Central-Asia	Nomadic herders	Exponentiel	–	–	–	41/ 123	73/ 221	130/ 396	-5.66x10 ⁻⁰⁵ / -1.72x10 ⁻⁰⁴	-1.13x10 ⁻⁰⁵ / -3.42x10 ⁻⁰⁵	2.99x10 ⁻⁰⁵ / 9.08x10 ⁻⁰⁵	–	–	–

* Note that for these populations, the Expansion model and a Constant model could not be distinguished from their DIC values (**Table I.D-S3**).

Table I.D-S3. Deviance Information Criterion (DIC) values computed for each demographic model.

Population	Area	Life-style	Constant model	Exponential model	Logistic model	Expansion model
Akele	Africa	Sedentary farmers	213	214	320	201
Bangando	Africa	Sedentary farmers	-969	-966	-939	-975
C. Fang	Africa	Sedentary farmers	278	275	283	268
Ewondo	Africa	Sedentary farmers	-68	-45	-21	-73
G. Fang	Africa	Sedentary farmers	-3760	-3751	-3487	-3769
Kota	Africa	Sedentary farmers	-1691	-1716	956	-1747
Nzebi	Africa	Sedentary farmers	-815	-879	-824	-889
Nzime	Africa	Sedentary farmers	49	76	87	54
Teke	Africa	Sedentary farmers	-1482	-1489	-1438	-1494
Tikar	Africa	Sedentary farmers	210	218	321	182
Tsogho	Africa	Sedentary farmers	-2424	-2546	-2478	-2574
Bezan	Africa	Nomadic HG	-1984	-1974	-1968	-1915
C. Baka	Africa	Nomadic HG	-40	-54	17	24
C.Bongo	Africa	Nomadic HG	385	516	524	487
E. Bongo	Africa	Nomadic HG	-241	-157	-156	-158
E. Baka	Africa	Nomadic HG	-1585	-2418	-1965	-1864
G.Baka	Africa	Nomadic HG	-570	-633	-477	-436
Kola	Africa	Nomadic HG	-580	-315	-268	-212
Koya	Africa	Nomadic HG	-46	-34	-16	3
S.Bongo	Africa	Nomadic HG	144	257	359	361
S.Baka	Africa	Nomadic HG	47	123	127	141
Bulgarians	Africa	Sedentary farmers	566	565	578	539
Germans	Eurasia	Sedentary farmers	449	439	397	371
H.Chinese	Eurasia	Sedentary farmers	500	484	498	469
Pathans	Eurasia	Sedentary farmers	730	721	727	691
Turkish	Eurasia	Sedentary farmers	559	538	558	530
Balochi	Eurasia	Nomadic herders	935	956	969	994
Kurd	Eurasia	Nomadic herders	303	316	332	340
Mongols	Eurasia	Nomadic herders	425	436	454	478
LUZ	Eurasia	Sedentary farmers	-78	-12	-69	-92

TAB	C.Asia	Sedentary farmers	262	114	13	-164
TDS	C.Asia	Sedentary farmers	-788	-785	-587	-796
TDU	C.Asia	Sedentary farmers	58	55	66	48
TJA	C.Asia	Sedentary farmers	-482	-492	-479	-498
TJE	C.Asia	Sedentary farmers	-381	-496	-348	-508
TJK	C.Asia	Sedentary farmers	-839	-860	-826	-879
TJN	C.Asia	Sedentary farmers	-335	-518	-479	-526
TJR	C.Asia	Sedentary farmers	-38	-204	-164	-217
TJT	C.Asia	Sedentary farmers	-684	-774	-469	-822
TJU	C.Asia	Sedentary farmers	-499	-567	-473	-639
TJY	C.Asia	Sedentary farmers	-790	-768	-762	-770
KAR	C.Asia	Nomadic herders	-2010	-1939	-1856	-1874
KAZ	C.Asia	Nomadic herders	-1459	-1190	-1105	-1079
KRA	C.Asia	Nomadic herders	-1766	-1761	-1035	-1036
KRB	C.Asia	Nomadic herders	-542	-534	-528	-521
KRG	C.Asia	Nomadic herders	-613	-505	-532	-578
KRL	C.Asia	Nomadic herders	-1104	-1014	-963	-1015
KRM	C.Asia	Nomadic herders	-358	-393	-241	-301
KRT	C.Asia	Nomadic herders	-824	-941	-841	-906
LKZ	C.Asia	Nomadic herders	187	205	216	203
OTU	C.Asia	Nomadic herders	-2054	-1664	-1756	-1748
TKY	C.Asia	Nomadic herders	-139	-289	-187	-138
TUR	C.Asia	Nomadic herders	-1599	-1264	-1233	-1378
UZA	C.Asia	Nomadic herders	-26	72	64	78
UZB	C.Asia	Nomadic herders	-759	-654	-603	-475
UZT	C.Asia	Nomadic herders	-1251	-1341	-1237	-1102

The model with the lowest DIC value, which is indicated in bold for each population, is considered as the best-fitting model for the observed data (Spiegelhalter et al., 2002). A difference of five points in DIC was considered as significant. The DIC formula is described with details in "Material and methods" section.

Table I.D-S4. Inferred dating for the expansion onsets using BSP.

Population	Area	Life-style	t
G. Fang	Africa	Sedentary farmers	105,525 / 320,796
Kota	Africa	Sedentary farmers	129,875 / 394,820
Nzebi	Africa	Sedentary farmers	89,800 / 272,992
Teke	Africa	Sedentary farmers	143,450 / 436,088
Tikar	Africa	Sedentary farmers	65,000 / 197,600
Tsogho	Africa	Sedentary farmers	100,450 / 305,368
Bulgarians	Africa	Sedentary farmers	6,800 / 20,672
Germans	Eurasia	Sedentary farmers	6,400 / 19,456
H.Chinese	Eurasia	Sedentary farmers	8,600 / 26,144
Pathans	Eurasia	Sedentary farmers	8,700 / 26,448
Turkish	Eurasia	Sedentary farmers	7,975 / 24,244
TAB	Central-Asia	Sedentary farmers	9,400 / 28,576
TJE	Central-Asia	Sedentary farmers	9,775 / 29,716
TJR	Central-Asia	Sedentary farmers	9,550 / 29,032
TJT	Central-Asia	Sedentary farmers	10,975 / 33,364

We used the population growth curves generated from BEAST to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 – 20 generation intervals. We considered that the population increased when both the median and HPD values for N_e increased gradually between more than three successive data points. Although this estimation method does not allow providing a confidence interval for inferred expansion timings, this conservative approach ensure that we consider only strong and significant expansion signals. Values in bold are inferred assuming a mutation rate of $\mu = 2.1 \times 10^{-3}$, the others are inferred using $\mu = 6.9 \times 10^{-4}$.

I.E. Conclusions de l'analyse des données réelles.

Il est temps à présent de reprendre les questions posées au début de ce chapitre afin de synthétiser les différentes informations apportées sur chaque point par nos trois études.

(1) Y a-t-il une association entre modes de vie et histoire démographique dans les populations humaines ?

Les différents types de marqueurs étudiés s'accordent généralement pour répondre à cette question par l'affirmative. En effet, nous avons détecté de forts signaux d'expansions pour la majorité des populations d'agriculteurs sédentaires en Afrique et en Eurasie. A l'inverse, aucune des populations de chasseurs-cueilleurs nomades étudiées n'a montré de signal d'expansion. Concernant les éleveurs nomades en Eurasie, ceux-ci ont montré des signaux d'expansion plus faibles que ceux des agriculteurs d'après l'étude des séquences mitochondriales HVS-I avec la méthode non-paramétrique. L'étude des données microsatellites ainsi que des séquences par la méthode paramétrique n'a montré aucun signal d'expansion pour ces populations (excepté en Asie centrale pour les microsatellites autosomaux), ce qui plaide pour une dynamique d'expansion plus faible ainsi qu'un plus faible pouvoir de détection par la méthode paramétrique comparé à la méthode non-paramétrique. Si la différence d'histoire démographique entre les populations agricoles sédentaires et les chasseurs-cueilleurs nomades avait déjà été démontrée en Afrique (*e.g.* Batini *et al.*, 2011), nous sommes à notre connaissance les premiers à comparer le passé démographique des populations agricoles et pastorales en Eurasie, qui représentent les deux principaux groupes post-Néolithiques dans cette partie du monde.

L'absence ou la faiblesse des signaux d'expansion inférés à la fois pour les populations de chasseur-cueilleurs en Afrique et d'éleveurs en Eurasie peut s'expliquer par le fait que ces populations partagent les contraintes liées au nomadisme. Notamment, l'intervalle entre les naissances est généralement plus court pour les populations nomades (Short, 1982), ce qui pourrait être dû à un régime alimentaire faible en carbohydrates ne favorisant pas la récupération rapide des femmes après l'accouchement (Valeggia and Ellison, 2009). De plus, après la transition Néolithique, la compétition avec les populations agricoles pour la terre et la fragmentation résultante de l'habitat et des ressources disponibles a pu limiter les possibilités d'expansion pour les populations restées nomades (Diamond, 2002).

La comparaison de populations géographiquement distantes en Eurasie nous a cependant permis de détecter une certaine variabilité locale des processus démographiques. Notamment en Asie centrale, les analyses des séquences ainsi que des microsatellites autosomaux ont montré des dynamiques d'expansion d'intensité similaire entre les agriculteurs sédentaires et les éleveurs nomades. Comme montré par l'analyse des données de séquences, cela semble dû tout d'abord à une plus faible expansion démographique des agriculteurs dans cette région par rapport au reste de l'Eurasie, phénomène qui pourrait être liée à l'aridité du climat (Dirksen and Van Geel, 2004). Par ailleurs, l'analyse des microsatellites autosomaux montre des évènements d'expansions chez les éleveurs d'Asie centrale mais pas du reste de l'Eurasie au Néolithique. Cela pourrait s'expliquer par une plus faible compétition pour la terre avec les agriculteurs, ou une origine agropastorale des éleveurs actuels de cette région, qui auraient donc connu une période de sédentarisation avant de retrouver un mode de vie nomade. Cependant, nous n'avons pas observé de traces d'expansions chez les éleveurs d'Asie centrale en étudiant le chromosome Y, contrairement aux autres types de marqueurs. On ne peut donc exclure que le signal d'expansion observé avec les autres marqueurs soit en partie dû à l'effet confondant des migrations entre populations, qui concernent davantage la lignée maternelle que la lignée paternelle dans ces populations exogames et patrilocales (Chaix *et al.*, 2007).

(2) Les grandes expansions ont-elles commencé avant ou pendant la transition Néolithique et les premières sédentarisations ?

L'étude des séquences autosomales et de la région HVS-I, en accord avec de précédentes études (*e.g.* Harpending *et al.*, 1993 ; Watson *et al.*, 1997 ; Pritchard *et al.*, 1999 ; Voight *et al.*, 2005; Chaix *et al.*, 2008 ; Atkinson *et al.*, 2009 ; Laval *et al.*, 2010 ; Batini *et al.*, 2011), suggère que les expansions démographiques des populations humaines auraient commencé dès le Paléolithique, donc avant l'apparition de l'agriculture et de l'élevage. De plus, les différences entre les groupes en termes de dynamique d'expansion seraient également antérieures au Néolithique. Par exemple, en Afrique, comme déjà montré par Batini *et al.* (2011), seuls les ancêtres des populations d'agriculteurs auraient connu une expansion au Paléolithique, contrairement aux populations restées nomades par la suite. Nous allons plus loin ici en démontrant que les différences inférées en terme d'intensité d'expansions entre populations sédentaires et nomades en Eurasie sont également antérieures à l'apparition de l'agriculture et de l'élevage dans toutes les régions étudiées. Ce résultat est

compatible avec la « théorie démographique » proposée par Sauer (1952), selon laquelle de fortes expansions paléolithiques auraient favorisé l'apparition de l'agriculture au Néolithique dans les populations concernées. Ces expansions paléolithiques ont pu être favorisées par l'émergence de nouveaux outils et de nouvelles techniques de chasse (Mellars, 2006). Notamment, l'apparition de l'arc entre 100 000 et 50 000 ans avant notre ère (Shea, 2009) a probablement fait partie d'un processus plus large de diversification écologique qui a pu impacter les dynamiques démographiques. Cependant, l'étude des microsatellites autosomaux nous montre également une deuxième phase d'expansion, plus récente, chez les populations d'agriculteurs en Afrique et en Eurasie. En accord avec les données des archéologues et des paléoanthropologues (*e.g.* Bar-Yosef et Belfer-Cohen, 1991 ; Mazoyet et Roudart, 2006 ; Bocquet-Appel, 2011) cette expansion pourrait coïncider avec l'apparition de l'agriculture et la sédentarisation de ces populations.

(3) Différents marqueurs apportent ils des informations similaires ou complémentaires sur le passé démographique des populations étudiées ?

Comme soulevé ci-avant, les microsatellites autosomaux nous ont permis de détecter des évènements plus récents que les autres types de marqueurs analysés. Ce résultat rappelle celui de Fontaine *et al.* (2012), qui infère des évènements plus récents en utilisant des microsatellites autosomaux qu'avec l'ADN mitochondrial chez le marsouin commun. Nous avons suggéré qu'une plus grande sensibilité des microsatellites autosomaux pour détecter les évènements récents peut être due à la combinaison d'un fort taux de mutation et de l'analyse conjointe de plusieurs loci indépendants.

De plus, au cours de notre première étude portant sur les données de séquences, nous avons montré des résultats différents entre les analyses des séquences autosomales et de la région HVS-I pour certaines populations. Notamment, les autosomes ont permis de détecter des évènements de contraction pour la population japonaise et la population danoise. Ces évènements pourraient coïncider avec des évènements historiques connus (*i.e.* la colonisation de l'archipel japonais et le dernier maximum glaciaire en Europe) qui ont pu entraîner des phénomènes de goulets d'étranglement (*i.e.* une brusque réduction d'effectif efficace suivi d'une nouvelle expansion). L'analyse de la région HVS-I montre cependant uniquement des évènements d'expansion pour ces populations. Les analyses sur données simulées présentées

en deuxième partie de cette thèse nous fourniront davantage d'informations sur les réponses spécifiques des différents marqueurs.

Finalement, pour l'Asie centrale, la comparaison des résultats obtenus à partir des deux marqueurs uni-parentaux (HVS-I et chromosome Y) nous a permis de détecter une influence des règles de mariage et des flux géniques entre populations sur les inférences démographiques. En effet, l'étude de la région HVS-I de l'ADN mitochondrial nous révèle des signaux d'expansions similaires chez les agriculteurs sédentaires et les éleveurs nomades dans cette région. A l'inverse, l'étude de marqueurs microsatellites du chromosome Y nous indique des signaux d'expansion chez les agriculteurs sédentaires mais pas chez les éleveurs nomades. Cependant, les signaux d'expansion détectés chez les éleveurs à partir de la région HVS-I pourraient être un artéfact dû aux migrations importantes entre ces populations, généralement exogames. Les populations d'Asie centrale étant patrilocales, ces migrations concernent principalement les femmes. Cette différence de taux de migration entre les deux sexes est à l'origine d'une plus grande diversité génétique pour l'ADN mitochondrial que pour le chromosome Y (Chaix *et al.*, 2007). En augmentant la proportion d'allèles rares dans les lignées maternelles, elle pourrait également être à l'origine des différences observées entre nos inférences avec les deux marqueurs.

(4) Dans quelle mesure les différences démographiques observées entre des populations aux modes de vie différents peuvent-elles être liées à des facteurs confondants tels que l'isolation et les migrations récentes ?

Au cours de notre étude portant sur la région mitochondriale HVS-I, nous avons calculé les F_{st} mesurant la différenciation de chaque population vis-à-vis de l'ensemble des autres populations étudiées dans la même région. Nous avons également estimé le nombre de d'immigrants dans chaque population. L'étude des corrélations entre ces paramètres et les taux de croissance inférés nous permet de détecter une influence potentielle de l'isolation et des migrations sur nos inférences. Notamment, comme montré par Ray *et al.* (2003) et Excoffier (2004), une forte isolation des populations peut atténuer dans une certaine mesure les traces d'expansions spatiales anciennes. Cependant, excepté peut-être en Asie centrale (Cf paragraphe précédent), les différences en termes d'isolation et de migrations récentes ne suffisent pas à expliquer les différences d'histoire démographique observées ici entre les groupes.

**Chapitre II : Vers une meilleure interprétation des inférences
démographiques à partir des données génétiques - Analyses
préliminaires sur données simulées.**

II.A. Introduction

Au cours de la première partie de cette thèse, nous avons montré que l'analyse de différents marqueurs génétiques permet d'appréhender de manière complémentaire différents aspects de l'histoire démographique des populations, notamment en nous renseignant sur des échelles de temps différentes. Dans cette deuxième partie, nous avons décidé de produire et analyser des données simulées afin de (i) tester les méthodes utilisées, et (ii) mieux comprendre les réponses spécifiques des différents marqueurs et améliorer l'interprétation des résultats. Il s'agit ici de travaux préliminaires qui seront amenés à être approfondis après la fin de cette thèse, notamment par l'ajout de scénarios incluant des processus de flux géniques, ainsi que par l'augmentation du nombre de réplifications de chaque scénario déjà étudié. Nous présenterons ici ces travaux en leur état actuel et discuterons plus en détail des futures perspectives envisagées.

Dans un premier temps, nous avons entrepris de tester, à l'aide de données simulées à partir de scénarios prédéfinis, les principales méthodes utilisées dans le chapitre 1 de ce travail, à savoir la méthode paramétrique ainsi que l'une des méthodes non-paramétriques (EBSP, Heled and Drummond, 2010) implémentées dans BEAST v1.6 (Drummond et Rambaut, 2007). Pour ce faire, nous avons appliqué ces méthodes à des données simulant des séquences autosomales et mitochondriales produites suivant des scénarios simples d'expansion, de déclin ou d'effectif constant (0 ou 1 évènement démographique). Cela nous a permis de déterminer la fréquence à laquelle les méthodes paramétrique et non-paramétrique nous permettent d'inférer correctement le scénario approprié. Des analyses similaires à partir de données microsatellites (autosomes et chromosome Y) sont également en cours.

Dans un deuxième temps, nous avons fait varier l'ancienneté et l'intensité des expansions simulées, afin de tester l'hypothèse selon laquelle les différents marqueurs seraient plus ou moins informatifs à des échelles de temps différentes. Pour cette partie, nous présenterons à la fois l'analyse de données de séquences simulées (autosomales et mitochondriales) et de données microsatellites simulées (autosomes et chromosome Y), afin de comparer les résultats obtenus pour chaque scénario avec les différents marqueurs. Cela nous a notamment permis de tester l'hypothèse soulevée en discussion du **chapitre I.C**, selon laquelle les microsatellites autosomaux permettraient de détecter des expansions plus récentes que les autres types de marqueurs, dû à la combinaison d'un fort taux de mutation et de la

possibilité d'analyser conjointement plusieurs locus indépendants. Cette hypothèse est supportée par nos analyses sur données réelles, qui nous ont permis d'inférer des expansions clairement paléolithiques à partir de données de séquences (**Chapitre I.B.**), alors que nous avons détecté des expansions plus récentes, compatibles avec la transition néolithique, lors de l'analyse de microsatellites autosomaux (**Chapitre I.C.**). Des études récentes portant sur une espèce animale (le marsouin commun, Fontaine *et al.*, 2012) ou sur des données simulées analysées grâce aux méthodes dites « ABC » (Approximate Bayesian Computation, Beaumont *et al.*, 2002) vont également en ce sens (Cornuet *et al.*, 2008). Dans cette même partie, nous avons également simulé et analysé des données issues de scénarios incluant deux expansions successives plus ou moins espacées dans le temps, afin de tester la possibilité de distinguer, avec les différents marqueurs, des expansions successives, comme notamment les expansions paléolithiques et néolithiques dans l'espèce humaine.

II.B. Test des méthodes à partir de séquences simulées.

II.B.i. Matériel et méthodes.

Production des données.

Nous avons utilisé le logiciel SIMCOAL v2 (Laval et Excoffier, 2004) afin de simuler des données correspondant à des scénarii d'histoires démographiques contrastées : (i) un effectif constant au cours du temps, (ii) une croissance exponentielle avec un taux de croissance $g = 10^{-5}/\text{génération}$, (iii) un déclin exponentiel ($g = -10^{-5}/\text{génération}$), (iv) une expansion ($g = 10^{-5}/\text{génération}$) depuis 20 000 ans faisant suite à une période d'effectif constant. Nous avons simulé dix jeux de données indépendants pour chaque scénario, afin d'obtenir plusieurs réplicats. Pour chaque scénario et chaque réplicat, nous avons produit pour l'équivalent de 100 individus ($N=100$) (i) 20 régions diploïdes et indépendantes de 1300 pb, simulant des séquences autosomales, et (ii) une région haploïde de 400 pb, simulant la région mitochondriale HVS-I. Pour les séquences mitochondriales, nous avons fixé le taux de mutation à $\mu = 10^{-5}$ par génération et par site (Howell *et al.*, 1996; Heyer *et al.*, 2001) et l'effectif efficace actuel à $N_e = 5\ 000$. En ce qui concerne les séquences autosomales, nous avons fixé $\mu = 2,5 \times 10^{-8}$ par génération et par site (Pluzhnikov *et al.*, 2002) et $N_e = 10\ 000$.

Analyse des données.

Les données ont été analysées à l'aide du logiciel BEAST v1.6 (Drummond and Rambaut, 2007) suivant les mêmes protocoles (méthode paramétrique et EBSP) que pour l'analyse des données de séquences autosomales et mitochondriales réelles. Les modèles démographiques testés pour la méthode paramétrique (modèle constant, modèle exponentiel, modèle logistique et modèle d'expansion), le modèle de mutation (GTR), les distributions *a priori* utilisés (uniforme de 0 à 10^6 pour N_0 et uniforme entre -1 et 1), le protocole de traitement des sorties du logiciel à l'aide des logiciels Tracer v1.5 (Rambaut and Drummond, 2007) et R v2.14.1 (R Development Core Team, 2011), ainsi que la nature des paramètres estimés (effectif efficace actuel N_0 , effectif efficace initial N_1 dans le cas du modèle d'expansion, et taux de croissance g) restent identiques ici (Cf. **Chapitre I.B.v.**). En complément, nous avons également utilisé le modèle d'expansion à partir d'une population constante pour estimer le moment du début des expansions démographiques inférées suivant la formule $t = (1/g) \times \ln(N_1/N_0)$ appliquée à chaque étape de l'algorithme MCMC, comme nous

l'avions fait précédemment pour l'analyse des microsatellites (**Chapitre I.C.iii. et I.D.iii.**). Nous avons choisi ici d'utiliser un seul taux de mutation pour chaque type de marqueur, le même que pour la production des données (soit $\mu = 10^{-5}$ par génération et par site pour les séquences mitochondriales et $\mu = 2.5 \times 10^{-8}$ par génération et par site pour les séquences autosomales). Comme pour les données réelles, nous avons choisi une longueur de chaîne de Markov de 10^7 multiplié par le nombre de locus. En ce qui concerne la sélection du meilleur modèle, nous avons choisi de tester ici la méthode basée sur le DIC, utilisée précédemment pour les microsatellites (**Chapitres I.C.iii et D.iii.**), plutôt que les méthodes de « Path sampling » et « Stepping-stone sampling » (**Chapitre I.B.v.**). En effet, ces dernières ont déjà été testées avec BEAST par Baele *et al.* (2012).

II.B.ii. Résultats.

Séquences « mitochondriales ».

Avec la méthode paramétrique, l'analyse des données simulées suivant un scénario correspondant à un effectif constant nous permet de distinguer sans ambiguïté le modèle approprié dans neuf cas sur dix (**Tableau II.B-1 et Tableau II.B-S1**). En effet, le modèle constant est alors sélectionné, avec une valeur de DIC inférieure de plus de 5 points à celles des autres modèles. Dans le dernier cas, le modèle constant présente également la plus faible valeur de DIC, mais l'écart entre les valeurs ne permet pas de distinguer ce modèle d'un modèle logistique. Cependant, le taux de croissance inféré avec le modèle logistique est alors extrêmement faible (8.62×10^{-8} [$-5.52 \times 10^{-12} - 4.57 \times 10^{-7}$]) et l'intervalle HPD inclut la valeur de 0, ce qui n'est donc pas interprétable comme un signal d'expansion significatif. Les effectifs efficaces inférés varient entre 3 030 [1 -6 248] et 6 193 [2 057 - 10 273], et l'intervalle HPD inclut la valeur de 5 000 (utilisée pour la simulation des données) dans dix cas sur dix. Conformément à l'attendu, la méthode non-paramétrique (EBSP, **Figure II.B-1A**) ne nous montre aucune variation significative d'effectif efficace au cours du temps. En effet, l'intervalle HPD du nombre de changements inférés inclut 0 pour chacun des répliquats (**Tableau II.B-S2**).

Tableau II.B-1. Paramètres inférés à partir du meilleur modèle démographique pour les séquences mitochondriales simulées sous un scénario d'effectif constant.

Réplicat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1	Constant	1320	4060	10015			
2	Constant	69	3908	9613			
3	Constant	2368	5975	9391			
4	Constant	3	3960	6101			
5	Constant	0	4025	6237			
6	Constant	2057	6193	10273			
7	Constant	51	3945	9758			
8	Constant	25	4059	10015			
9*	Constant	1	3030	6248			
9*	Exponentiel	759	5391	32394	-6.52×10^{-12}	8.62×10^{-08}	4.57×10^{-07}
10	Constant	83	4104	10165			

**Pour ces répliquats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : $N_e = 5\ 000$.*

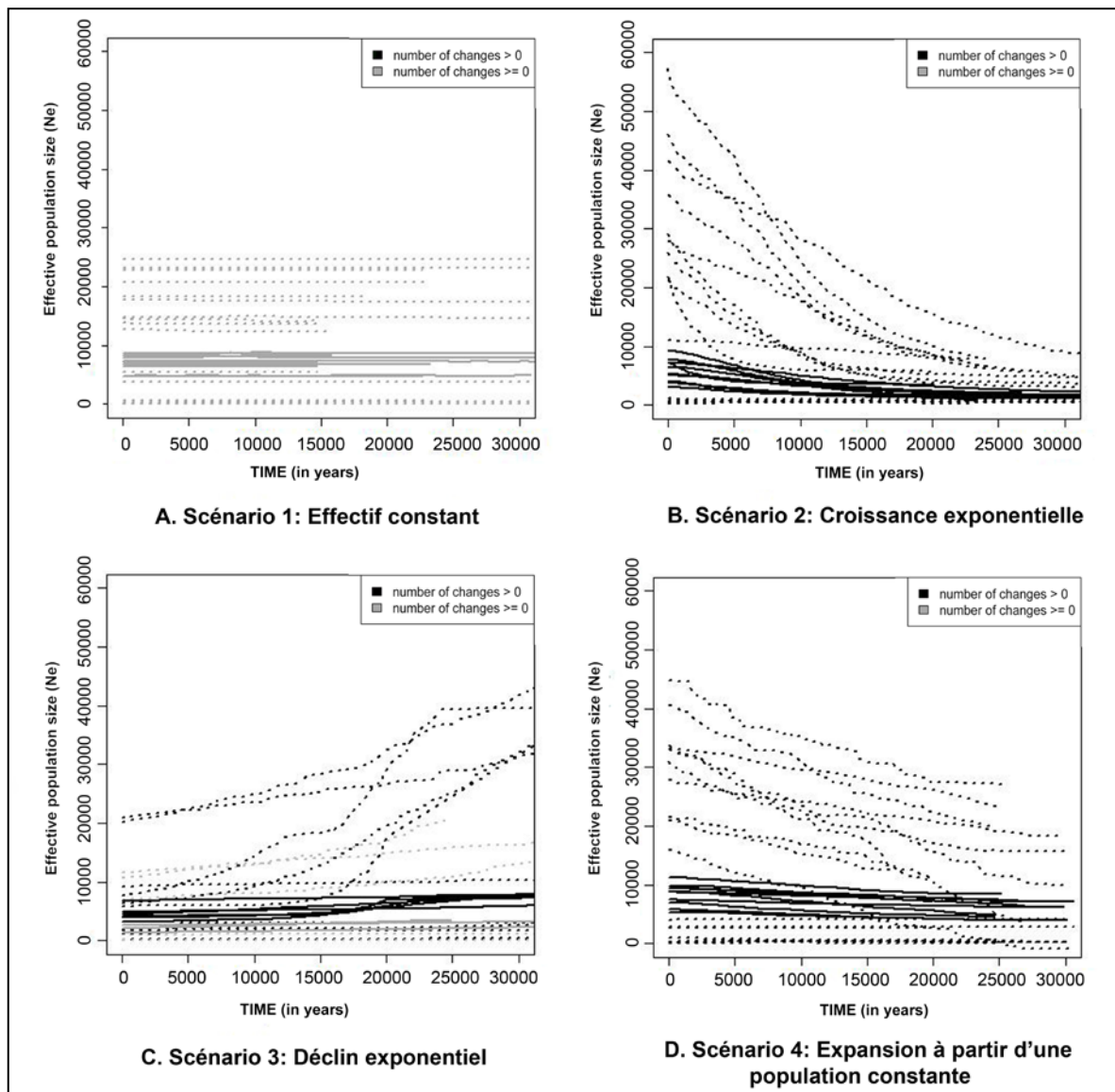


Figure II.B-1. EBSI inférés à partir de séquences HVS-I simulées sous différents scénarios. Paramètres utilisés pour la simulation des données : $N_0=5\ 000$, $g=10^{-5}$ (scénarios 2 et 4) ou -10^{-5} (scénario 3), $t=20\ 000$ (scénario 4). Le mode et l'intervalle HPD du nombre de changements démographiques inférés sont présentés dans le **Tableau II.B-S2**.

Dans le cas de données simulées sous un scénario de croissance exponentielle, le modèle exponentiel est également sélectionné comme meilleur modèle sans ambiguïté dans neuf cas sur dix (**Tableau II.B-2 et Tableau II.B-S1**). Dans un cas sur dix, on ne peut distinguer le modèle exponentiel d'un modèle logistique (différence de $DIC < 5$), ce qui constitue également un signal d'expansion. Les effectifs efficaces inférés varient entre 4 500 [1 364 -17 250] et 8 208 [1 044 - 20 382], et l'intervalle HPD inclut la valeur de 5 000 (utilisée pour la simulation des données) dans dix cas sur dix. Les taux de croissance inférés varient entre 8.78×10^{-6} [$1.37 \times 10^{-6} - 6.78 \times 10^{-5}$] et 7.55×10^{-5} [$1.56 \times 10^{-6} - 8.40 \times 10^{-4}$], et

l'intervalle de HPD inclut la valeur de 10^{-5} dans dix cas sur dix. De plus, l'intervalle HPD inclut uniquement des valeurs positives dans dix cas sur dix, ce qui représente un net signal de croissance démographique. En ce qui concerne la méthode non-paramétrique, le graphique EBSP montre une croissance significative pour chacun des dix réplicats, conformément à l'attendu (**Figure II.B-1B**). En effet, l'intervalle HPD du nombre de changements démographiques inférés au cours du temps n'inclus jamais la valeur de 0, ce qui permet de rejeter l'hypothèse d'une population constante (**Tableau II.B-S2**).

Tableau II.B-2. Paramètres inférés à partir du meilleur modèle démographique pour les séquences mitochondriales simulées sous un scénario de croissance exponentielle.

Réplicat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1	Exponentiel	1364	4500	17250	1.57×10^{-06}	3.15×10^{-05}	1.53×10^{-04}
2	Exponentiel	975	7104	14172	2.85×10^{-06}	5.20×10^{-05}	9.78×10^{-05}
3	Exponentiel	612	6444	15555	4.87×10^{-06}	4.00×10^{-05}	8.79×10^{-04}
4	Exponentiel	687	5859	22356	1.87×10^{-06}	9.30×10^{-06}	6.39×10^{-05}
5	Exponentiel	1428	7095	20955	1.56×10^{-06}	7.55×10^{-05}	8.40×10^{-04}
6	Exponentiel	1074	7947	22176	4.47×10^{-06}	1.80×10^{-05}	1.12×10^{-04}
7	Exponentiel	2022	6624	12369	6.38×10^{-06}	1.02×10^{-05}	5.59×10^{-04}
8	Exponentiel	1644	7548	22953	2.26×10^{-06}	2.35×10^{-05}	1.54×10^{-04}
9*	Exponentiel	1374	7062	21408	1.52×10^{-06}	2.90×10^{-05}	8.95×10^{-04}
9*	Logistique	1419	7254	16278	5.00×10^{-06}	1.12×10^{-05}	9.69×10^{-03}
10	Exponentiel	1044	8208	20382	1.37×10^{-06}	8.78×10^{-06}	6.78×10^{-05}

* Pour ces réplicats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de $DIC < 5$ points). Paramètres utilisés pour la simulation des données : $N_e = 5\ 000$, $g = 10^{-5}$.

Dans le cas de données simulées sous un scénario de déclin exponentiel, le modèle exponentiel avec un taux de croissance négatif est sélectionné comme meilleur modèle sans ambiguïté dans huit cas sur dix (**Tableau II.B-3 et Tableau II.B-S1**). Dans deux cas sur dix, on ne peut distinguer le modèle exponentiel d'un modèle logistique, avec également un taux de croissance négatif. Les effectifs efficaces inférés varient entre 4 052 [1 734 -15 181] et 8 259 [1 701 - 17 442], et l'intervalle HPD inclut la valeur de 5 000 (utilisée pour la simulation des données) dans dix cas sur dix. Les taux de croissance inférés varient entre -5.80×10^{-5} [-8.62×10^{-4} - -3.85×10^{-9}] et -8.26×10^{-6} [-5.25×10^{-5} - -6.54×10^{-9}], et l'intervalle HPD

inclut la valeur de -10^{-5} dans dix cas sur dix. De plus, l'intervalle HPD inclut exclusivement des valeurs négatives dans dix cas sur dix, ce qui, conformément à l'attendu, représente un net signal de déclin démographique. En ce qui concerne la méthode non-paramétrique, le graphique EBSF montre un déclin démographique significatif pour sept des dix répliquats (**Figure II.B-1C**). Les trois autres répliquats montrent également un signal de léger déclin, mais celui-ci n'est pas significatif. En effet, l'intervalle HPD du nombre de changements démographiques inférés au cours du temps inclut la valeur de 0 pour ces trois répliquats, contrairement aux sept autres (**Tableau II.B-S2**).

Tableau II.B-3. Paramètres inférés à partir du meilleur modèle démographique pour les séquences mitochondriales simulées sous un scénario de déclin exponentiel.

Répliquat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1	Exponentiel	1734	4052	15181	-5.25×10^{-05}	-8.26×10^{-06}	-6.54×10^{-09}
2	Exponentiel	1224	5685	20703	-4.90×10^{-03}	-2.30×10^{-05}	-8.68×10^{-07}
3	Exponentiel	1221	6573	18981	-1.59×10^{-03}	-2.78×10^{-05}	-6.65×10^{-09}
4	Exponentiel	2004	6876	16896	-6.25×10^{-04}	-1.28×10^{-05}	-5.85×10^{-09}
5*	Exponentiel	2394	6216	22002	-8.53×10^{-04}	-5.10×10^{-05}	-6.60×10^{-09}
5*	Logistique	2130	6534	20988	-6.59×10^{-04}	-9.25×10^{-06}	-1.72×10^{-09}
6	Exponentiel	1701	8259	17442	-8.62×10^{-04}	-5.80×10^{-05}	-3.85×10^{-09}
7	Exponentiel	1590	6540	20661	-7.82×10^{-05}	-9.44×10^{-06}	-2.14×10^{-09}
8	Exponentiel	2316	6924	18114	-8.68×10^{-04}	-2.54×10^{-05}	-2.62×10^{-09}
9*	Exponentiel	1155	6891	15465	-1.26×10^{-04}	-9.20×10^{-06}	-2.26×10^{-09}
9*	Logistique	1158	6993	15444	-6.04×10^{-03}	-9.24×10^{-06}	-2.87×10^{-07}
10	Exponentiel	1488	7143	17379	-8.80×10^{-04}	-2.66×10^{-05}	-5.55×10^{-09}

*Pour ces répliquats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de $DIC < 5$ points). Paramètres utilisés pour la simulation des données : $N_e = 5\ 000$, $g = -10^{-5}$.

Dans le cas de données simulées sous un scénario de croissance exponentielle à partir d'une population de taille constante, ce modèle d'expansion est sélectionné comme meilleur modèle sans ambiguïté dans sept cas sur dix (**Tableau II.B-4 et Tableau II.B-S1**). Dans les trois cas restants, on ne peut distinguer le modèle d'expansion à partir d'une population de taille constante d'un modèle de croissance logistique ou exponentielle, qui constitue également un signal d'expansion. Les effectifs efficaces inférés varient entre 2 045 [537 -5

068] et 2 905 [434 – 6 838], et l'intervalle HPD inclut la valeur de 5 000 (utilisée pour la simulation des données) dans dix cas sur dix. Les taux de croissance inférés varient entre 6.84×10^{-6} [2.85×10^{-8} – 9.88×10^{-5}] et 7.51×10^{-5} [4.25×10^{-8} – 1.55×10^{-3}], et l'intervalle HPD inclut la valeur de 10^{-5} dans dix cas sur dix. De plus, l'intervalle HPD inclut exclusivement des valeurs positives dans dix cas sur dix, ce qui représente un net signal de croissance démographique. Les dates de début d'expansion inférées varient entre 14 131 [8 238 - 40 341] et 27 623 [16 647 – 52 560], et l'intervalle HPD inclut la valeur de 20 000 dans dix cas sur dix. En ce qui concerne la méthode non-paramétrique, le graphique EBSP montre une croissance significative pour chacun des dix réplicats, conformément à l'attendu (**Figure II.B-1D, Tableau II.B-S2**).

Tableau II.B-4. Paramètres inférés à partir du meilleur modèle démographique pour les séquences mitochondriales simulées sous un scénario de croissance exponentielle à partir d'une population d'effectif initial constant.

Réplikat	Modèle	$N1$ (HPD 5 %)	$N1$ (mode)	$N1$ (HPD 95 %)	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)	t (HPD 5 %)	t (mode)	t (HPD 95 %)
1	Expansion	7	184	1267	93	2357	17624	7.51×10^{-08}	4.25×10^{-05}	7.28×10^{-03}	16647	27623	52560
2	Expansion	3	266	1650	124	2667	18995	4.25×10^{-08}	7.51×10^{-05}	1.55×10^{-03}	8238	14131	40341
3*	Exponentiel				572	2645	7458	9.78×10^{-09}	8.81×10^{-06}	4.25×10^{-04}			
3*	Expansion	0	375	2073	157	2569	18778	6.50×10^{-08}	6.25×10^{-05}	6.51×10^{-03}	6983	14174	49006
4	Expansion	0	225	2063	96	2727	19532	6.55×10^{-08}	5.88×10^{-05}	9.52×10^{-03}	10869	19533	26185
5*	Exponentiel				434	2905	6838	8.48×10^{-08}	7.90×10^{-06}	3.65×10^{-04}			
5*	Expansion	0	674	3236	101	2778	25194	3.29×10^{-08}	3.75×10^{-05}	7.34×10^{-03}	12871	17386	43561
6	Expansion	0	348	2247	122	2769	25843	6.85×10^{-08}	5.23×10^{-05}	8.75×10^{-03}	12846	18256	47301
7	Expansion	88	264	2045	226	2135	14531	1.68×10^{-08}	4.76×10^{-05}	1.27×10^{-03}	7107	20236	46875
8*	Exponentiel				537	2045	5068	2.85×10^{-08}	6.84×10^{-06}	9.88×10^{-05}			
8*	Expansion	0	268	3232	65	2170	15094	8.55×10^{-08}	1.80×10^{-05}	1.08×10^{-03}	6582	21797	71746
9	Expansion	0	145	3122	56	2075	20980	8.74×10^{-08}	6.32×10^{-05}	4.51×10^{-03}	5944	19383	38732
10	Expansion	0	104	4122	36	2204	20870	8.57×10^{-08}	4.56×10^{-05}	4.75×10^{-03}	5719	20801	39343

**Pour ces réplikat, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : $N_e = 5\ 000$, $g = 10^{-5}$, $t = 20\ 000$.*

Séquences « autosomales » simulées.

Dans le cas de données simulées sous un scénario correspondant à un effectif constant, le modèle constant est choisi comme meilleur modèle sans ambiguïté dans huit cas sur dix (**Tableau II.B-5**). Dans deux cas sur dix, le modèle constant ne peut être distingué du modèle exponentiel, la différence de DIC entre les deux modèles étant inférieure à 5 points (**Tableau II.B-S3**). Cependant, l'intervalle HPD des taux de croissance inférés avec le modèle exponentiel dans les deux derniers cas (respectivement -5.12×10^{-5} [$-5.3 \times 10^{-3} - 1.08 \times 10^{-2}$] et -3.52×10^{-5} [$-7.59 \times 10^{-4} - 6.62 \times 10^{-5}$]) inclut la valeur de zéro, ce qui n'est donc pas interprétable comme un signal d'expansion significatif. Les effectifs efficaces inférés varient entre 10 838 [7 823 -13 921] et 19 032 [9 453 – 30 319], et l'intervalle HPD inclut la valeur de 10 000 (utilisée pour la simulation des données) dans dix cas sur dix. Conformément à l'attendu, la méthode non-paramétrique n'indique de variations significatives d'effectif efficace au cours du temps pour aucun des dix répliqués (**Figure II.B-2A**). En effet, l'intervalle HPD du nombre de changements démographiques inférés inclut la valeur de zéro dans tous les cas (**Tableau II.B-S4**).

Tableau II.B-5. Paramètres inférés à partir du meilleur modèle démographique pour les séquences autosomales simulées sous un scénario d'effectif constant.

Répliquat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1*	Constant	7823	10838	13921			
1*	Exponentiel	9453	19032	30319	5.30×10^{-03}	-5.12×10^{-05}	1.08×10^{-02}
2	Constant	9514	12853	16339			
3	Constant	10082	13305	16775			
4	Constant	7472	13886	17203			
5	Constant	8492	13886	17203			
6*	Constant	10004	13445	16811			
6*	Exponentiel	8557	11847	15254	-7.59×10^{-04}	-3.52×10^{-05}	6.62×10^{-05}
7	Constant	7366	13658	17239			
8	Constant	9686	13094	16515			
9	Constant	8994	13468	18126			
10	Constant	10014	14488	18126			

**Pour ces répliqués, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : Ne=10 000.*

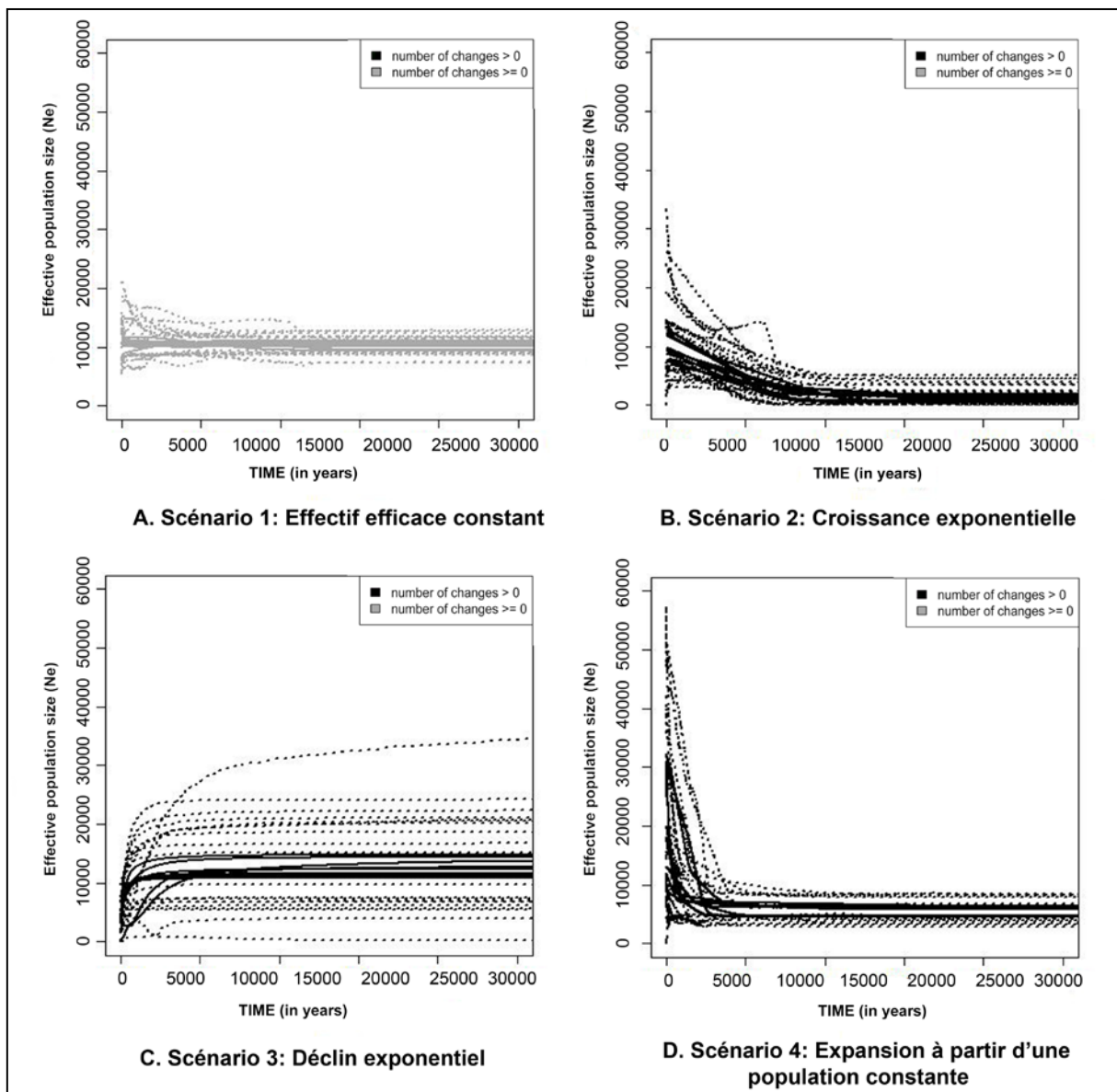


Figure II.B-2. EBSI inférés à partir de séquences autosomales simulées sous différents scénarios. Paramètres utilisés pour la simulation des données : $N_0=10\,000$, $g=10^{-5}$ (scénarios 2 et 4) ou -10^{-5} (scénario 3), $t=20\,000$ (scénario 4). Le mode et l'intervalle HPD du nombre de changements démographiques inférés sont présentés dans le **Tableau II.B-S4**.

Dans le cas de données simulées sous un scénario de croissance exponentielle, le modèle exponentiel est également sélectionné comme meilleur modèle sans ambiguïté dans huit cas sur dix (**Tableau II.B-6**). Dans deux cas sur dix, on ne peut distinguer le modèle exponentiel d'un modèle logistique (différence de $DIC < 5$), qui constitue également un signal d'expansion (**Tableau II.B-S3**). Les effectifs efficaces inférés varient entre 10 065 [7 922 -15 264] et 11 887 [7 899 – 16 285], et l'intervalle HPD inclut la valeur de 10 000 dans dix cas sur dix. Les taux de croissance inférés varient entre 8.07×10^{-6} [$3.81 \times 10^{-6} - 1.45 \times 10^{-4}$] et 6.25×10^{-5} [$1.98 \times 10^{-6} - 8.95 \times 10^{-5}$], et l'intervalle HPD inclut la valeur de 10^{-5} dans dix cas sur dix. De plus, l'intervalle HPD n'inclus que des valeurs positives dans dix cas sur dix, ce qui

représente un net signal de croissance démographique. Conformément à l'attendu, la méthode non-paramétrique indique une croissance démographique significative au cours du temps pour chacun des dix répliquats (EBSP, **Figure II.B-2B**). En effet, l'intervalle HPD du nombre de changements démographiques inférés n'inclus jamais la valeur de 0 (**Tableau II.B-S4**).

Tableau II.B-6. Paramètres inférés à partir du meilleur modèle démographique pour les séquences autosomales simulées sous un scénario de croissance exponentielle.

Répliquat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1	Exponentiel	7899	11887	16285	1.67×10^{-06}	1.43×10^{-05}	3.97×10^{-04}
2*	Exponentiel	8274	10938	14262	1.98×10^{-06}	6.25×10^{-05}	8.95×10^{-05}
2*	Logistique	7922	10065	15264	3.26×10^{-06}	2.05×10^{-05}	2.54×10^{-04}
3	Exponentiel	7938	11271	14938	2.58×10^{-06}	9.91×10^{-06}	9.87×10^{-05}
4	Exponentiel	8163	11873	15006	3.81×10^{-06}	8.07×10^{-06}	1.45×10^{-04}
5	Exponentiel	8437	11489	14483	4.65×10^{-06}	2.13×10^{-05}	4.53×10^{-04}
6*	Exponentiel	8289	10937	14936	1.65×10^{-06}	2.95×10^{-05}	4.74×10^{-04}
6*	Logistique	7512	10294	15127	4.65×10^{-06}	4.62×10^{-05}	8.82×10^{-04}
7	Exponentiel	7940	10494	15324	9.63×10^{-07}	1.47×10^{-05}	4.35×10^{-04}
8	Exponentiel	7932	10682	14321	8.51×10^{-07}	3.39×10^{-05}	1.07×10^{-04}
9	Exponentiel	8763	10948	14172	6.54×10^{-06}	1.75×10^{-05}	2.05×10^{-04}
10	Exponentiel	8396	11507	14683	1.45×10^{-06}	2.22×10^{-05}	3.04×10^{-04}

**Pour ces répliquats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : $N_e = 10\ 000$, $g = 10^{-5}$.*

Dans le cas de données simulées sous un scénario de déclin exponentiel, le modèle exponentiel avec un taux de croissance négatif est sélectionné comme meilleur modèle sans ambiguïté dans huit cas sur dix (**Tableau II.B-7, S3**). Dans deux cas sur dix, on ne peut distinguer le modèle exponentiel d'un modèle logistique, avec également un taux de croissance négatif. Les effectifs efficaces inférés varient entre 7 874 [4 327 -28 872] et 10 242 [3 204 – 26 984], et l'intervalle HPD inclut la valeur de 10 000 (utilisée pour la simulation des données) dans dix cas sur dix. Les taux de croissance inférés varient entre -5.57×10^{-5} [$-3.65 \times 10^{-4} - -7.85 \times 10^{-6}$] et -9.20×10^{-6} [$-6.55 \times 10^{-4} - -9.13 \times 10^{-7}$], et l'intervalle HPD inclut la valeur de -10^{-5} dans dix cas sur dix. De plus, l'intervalle HPD inclus uniquement des valeurs négatives dans dix cas sur dix, ce qui représente un net signal de déclin démographique.

Conformément à l'attendu, la méthode non-paramétrique indique un déclin démographique significatif au cours du temps pour chacun des dix répliquats (EBSP, **Figure II.B-2C**, **Tableau II.B-S4**).

Tableau II.B-7. Paramètres inférés à partir du meilleur modèle démographique pour les séquences autosomales simulées sous un scénario de déclin exponentiel.

Répliquat	Modèle	No (HPD 5 %)	No (mode)	No (HPD 95 %)	g (HPD 5 %)	g (mode)	g (HPD 95 %)
1	Exponentiel	3008	8187	25515	-2.87×10^{-04}	-2.56×10^{-05}	-1.53×10^{-06}
2	Exponentiel	4009	8215	19101	-3.65×10^{-04}	-5.57×10^{-05}	-7.85×10^{-06}
3	Exponentiel	4327	7875	28872	-3.25×10^{-04}	-2.34×10^{-05}	-1.24×10^{-06}
4*	Exponentiel	2917	7959	31113	-9.85×10^{-04}	-3.27×10^{-05}	-7.86×10^{-06}
4*	Logistique	2580	7719	28083	-4.86×10^{-05}	-4.87×10^{-06}	-6.45×10^{-12}
5	Exponentiel	1725	8377	29671	-4.52×10^{-04}	-8.90×10^{-06}	-9.58×10^{-07}
6	Exponentiel	1932	8153	24773	-6.55×10^{-04}	-9.20×10^{-06}	-9.13×10^{-07}
7	Exponentiel	2465	9315	27503	-6.45×10^{-04}	-3.59×10^{-05}	-3.28×10^{-06}
8*	Exponentiel	3204	10243	26985	-4.16×10^{-04}	-5.47×10^{-05}	-9.87×10^{-06}
8*	Logistique	3220	10265	26916	-5.41×10^{-04}	-5.70×10^{-06}	-1.90×10^{-12}
9	Exponentiel	2905	10165	28891	-5.87×10^{-04}	-1.21×10^{-05}	-8.85×10^{-06}
10	Exponentiel	2889	8321	26386	-6.58×10^{-04}	-4.89×10^{-05}	-9.86×10^{-06}

**Pour ces répliquats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : $N_e = 10\ 000$, $g = -10^{-5}$.*

Dans le cas de données simulées sous un modèle d'expansion, c'est-à-dire un scénario de croissance exponentielle à partir d'une population de taille constante, ce modèle d'expansion est sélectionné comme meilleur modèle sans ambiguïté dans sept cas sur dix (**Tableau II.B-8, S3**). Dans les trois cas restants, on ne peut distinguer le modèle d'expansion à partir d'une population de taille constante d'un modèle de croissance logistique ou exponentielle, qui constitue également un signal d'expansion. Les effectifs efficaces inférés varient entre 7 165 [6 449 - 10 732] et 15 857 [4 501 - 72 071], et l'intervalle HPD inclut la valeur de 10 000 dans dix cas sur dix. Les taux de croissance inférés varient entre 6.33×10^{-6} [$8.18 \times 10^{-7} - 9.87 \times 10^{-3}$] et 9.95×10^{-5} [$8.08 \times 10^{-6} - 4.58 \times 10^{-3}$], et l'intervalle HPD inclus la valeur de 10^{-5} dans dix cas sur dix. L'intervalle HPD n'inclut que des valeurs positives, ce qui

représente un net signal de croissance démographique, dans dix cas sur dix. Les dates de début d'expansion inférées varient entre 11 669 [1 278 - 38 640] et 32 461 [11 789 – 51 463], et l'intervalle HPD inclus la valeur de 20 000 dans dix cas sur dix. Conformément à l'attendu, la méthode non-paramétrique indique une croissance démographique significative au cours du temps pour chacun des dix réplicats (EBSP, **Figure II.B-2B, Tableau II.B-S4**).

Tableau II.B-8. Paramètres inférés à partir du meilleur modèle démographique pour les séquences autosomales simulées sous un scénario de croissance exponentielle à partir d'une population d'effectif efficace initial constant.

réplicat	modèle	<i>N</i> 1 (HPD 5 %)	<i>N</i> 1 (mode)	<i>N</i> 1 (HPD 95 %)	<i>N</i> 0 (HPD 5 %)	<i>N</i> 0 (mode)	<i>N</i> 0 (HPD 95 %)	<i>g</i> (HPD 5 %)	<i>g</i> (mode)	<i>g</i> (HPD 95 %)	<i>t</i> (HPD 5 %)	<i>t</i> (mode)	<i>t</i> (HPD 95 %)
1	Expansion	1	1205	3135	5608	10008	62425	1.94×10^{-05}	3.08×10^{-05}	8.48×10^{-03}	11789	32461	51463
2	Expansion	4	1754	4261	5751	9548	57791	1.45×10^{-06}	4.23×10^{-05}	5.21×10^{-03}	10042	18433	38804
3	Expansion	20	1177	2964	5156	11348	66019	2.87×10^{-06}	2.15×10^{-05}	9.60×10^{-03}	10418	26598	34788
4	Expansion	6	795	2309	5783	13748	37163	1.99×10^{-06}	9.98×10^{-06}	4.43×10^{-03}	12793	15291	30688
5*	Exponentiel				7503	12574	21051	8.08×10^{-06}	9.95×10^{-05}	4.58×10^{-03}			
5*	Expansion	9	1491	2351	6449	7165	10732	8.18×10^{-07}	6.33×10^{-06}	9.87×10^{-03}	10129	15810	23152
6*	Logistique				6701	12145	21715	4.25×10^{-11}	3.54×10^{-05}	7.83×10^{-04}			
6*	Expansion	15	1364	3901	5936	14643	41294	9.54×10^{-07}	3.21×10^{-05}	4.88×10^{-03}	12748	18233	43505
7	Expansion	12	793	3489	4908	8010	18384	1.53×10^{-06}	1.46×10^{-05}	3.99×10^{-03}	5078	18544	51890
8	Expansion	28	1024	1805	5296	7512	82064	1.26×10^{-06}	1.79×10^{-05}	2.29×10^{-03}	1278	11669	38640
9*	Exponentiel				6900	12298	20635	9.46×10^{-06}	9.56×10^{-05}	5.47×10^{-03}	80069		
9*	Expansion	31	583	2036	5549	15857	95035	2.86×10^{-06}	4.02×10^{-05}	9.86×10^{-03}	1014	24907	35007
10	Expansion	13	669	2063	4501	15137	72071	1.88×10^{-06}	8.64×10^{-06}	2.59×10^{-03}	8759	15326	33179

**Pour ces répliquats, deux modèles ne peuvent être distingués sur la base de leur DIC (différence de DIC < 5 points). Paramètres utilisés pour la simulation des données : $N_e = 10\ 000$, $g = 10^{-5}$, $t = 20\ 000$.*

II.B.iii. Supplementary Tables.

Tableau II.B-S1. Valeurs de DIC (Deviance Information Criteria) calculées pour chaque modèle démographique avec les séquences HVS-I simulées.

Scénario	Réplikat	DIC modèle constant	DIC modèle exponentiel	DIC modèle logistique	DIC modèle d'expansion
Constant	1	90189	90202	90215	90286
Constant	2	90177	90251	90246	90325
Constant	3	90180	90196	91005	91084
Constant	4	90241	90253	90257	90286
Constant	5	90230	90248	90241	90258
Constant	6	90312	90324	90325	90354
Constant	7	90285	90297	90299	91003
Constant	8	90297	90306	90314	90377
Constant	9	90358	90367	90359	90485
Constant	10	90285	90293	90296	90324
Croissance exponentielle	1	91877	91856	91862	91868
Croissance exponentielle	2	91825	91803	91816	91812
Croissance exponentielle	3	91858	91846	91856	91869
Croissance exponentielle	4	91885	91871	91879	91888
Croissance exponentielle	5	91865	91857	91866	91877
Croissance exponentielle	6	91865	91851	91861	91863
Croissance exponentielle	7	91837	91829	91836	91841
Croissance exponentielle	8	91832	91821	91829	91836
Croissance exponentielle	9	91923	91913	91911	91919
Croissance exponentielle	10	91815	91802	91815	91818
Déclin exponentiel	1	93375	94338	94347	93364
Déclin exponentiel	2	94372	94330	94338	94349
Déclin exponentiel	3	94369	94335	94342	94351
Déclin exponentiel	4	94364	94329	94337	94347
Déclin exponentiel	5	94374	94331	94329	94343
Déclin exponentiel	6	94368	94333	94341	94352
Déclin exponentiel	7	94372	94335	94344	94367
Déclin exponentiel	8	94374	94331	94346	94365

Déclin exponentiel	9	94365	94337	94340	94358
Déclin exponentiel	10	94361	94329	94337	94350
Expansion	1	100125	100084	100090	100078
Expansion	2	100147	100082	100090	100072
Expansion	3	100104	100081	100092	100083
Expansion	4	100175	100095	100099	100080
Expansion	5	100099	100084	100091	100084
Expansion	6	100132	100085	100099	100072
Expansion	7	100147	100082	100089	100069
Expansion	8	100995	100079	100088	100075
Expansion	9	100175	100081	100099	100072
Expansion	10	100095	100092	100099	100082

Le modèle présentant la plus faible valeur de DIC, indiqué en gras pour chaque réplicat, est considéré comme le mieux ajusté aux données (Spiegelhalter et al., 2002). Une différence de DIC d'au moins 5 points est considérée comme significative. Dans le cas contraire, il est impossible de distinguer le meilleur modèle.

Tableau II.B-S2. Modes et intervalle HPD du nombre de changements démographiques inférés par la méthode non-paramétrique pour les séquences HVS-I simulées.

Scénario	Réplikat	HPD 5%	Mode	HPD 95%
Constant	1	0	0	2
Constant	2	0	0	1
Constant	3	0	1	2
Constant	4	0	0	2
Constant	5	0	0	2
Constant	6	0	1	2
Constant	7	0	1	2
Constant	8	0	1	1
Constant	9	0	0	0
Constant	10	0	0	0
Croissance exponentielle	1	1	1	2
Croissance exponentielle	2	1	1	2
Croissance exponentielle	3	1	1	2
Croissance exponentielle	4	1	1	3
Croissance exponentielle	5	1	1	2
Croissance exponentielle	6	1	1	2
Croissance exponentielle	7	1	1	2
Croissance exponentielle	8	1	1	2
Croissance exponentielle	9	1	1	2
Croissance exponentielle	10	1	1	3
Déclin exponentiel	1	1	1	2
Déclin exponentiel	2	1	1	3
Déclin exponentiel	3	1	1	2
Déclin exponentiel	4	1	1	2
Déclin exponentiel	5	0	1	2
Déclin exponentiel	6	1	1	2
Déclin exponentiel	7	1	1	2
Déclin exponentiel	8	0	1	3
Déclin exponentiel	9	0	1	3
Déclin exponentiel	10	1	1	2

Expansion	1	1	1	2
Expansion	2	1	1	3
Expansion	3	1	1	3
Expansion	4	1	1	2
Expansion	5	1	1	1
Expansion	6	1	1	3
Expansion	7	1	1	2
Expansion	8	1	1	2
Expansion	9	1	1	2
Expansion	10	1	1	2

On considèrera ici que les évènements d'expansion ou de déclin observé sur les EBSP sont significatifs lorsque l'intervalle HPD du nombre de changements démographiques n'inclus pas la valeur de zéro (Heled et Drummond, 2010).

Tableau II.B-S3. Valeurs de DIC (Deviance Information Criteria) calculées pour chaque modèle démographique avec les séquences autosomales simulées.

Scénario	Réplikat	DIC (modèle constant)	DIC (modèle exponentiel)	DIC (modèle logistique)	DIC (modèle expansion)
Constant	1	56213	56211	56299	56305
Constant	2	56643	56692	56702	56675
Constant	3	56479	56513	56550	56541
Constant	4	55862	55889	55913	55963
Constant	5	56851	57155	57029	57197
Constant	6	56712	56710	56972	57129
Constant	7	57059	57185	57259	57322
Constant	8	56983	57097	57136	57612
Constant	9	57124	57197	57269	57263
Constant	10	56473	56504	56509	56616
Croissance exponentielle	1	75155	74982	75012	75062
Croissance exponentielle	2	75126	74951	74953	75019
Croissance exponentielle	3	75086	74919	75049	74980
Croissance exponentielle	4	75015	74921	74971	75014
Croissance exponentielle	5	75066	74904	75014	74983
Croissance exponentielle	6	75177	74913	75012	74922
Croissance exponentielle	7	75151	74981	75013	75068
Croissance exponentielle	8	75180	74908	75065	75032
Croissance exponentielle	9	75048	75002	75035	75028
Croissance exponentielle	10	75065	74989	75006	75023
Déclin exponentiel	1	63120	62548	64210	65985
Déclin exponentiel	2	66874	65875	67895	68521
Déclin exponentiel	3	62987	62545	67852	68026
Déclin exponentiel	4	63521	63581	65579	65897
Déclin exponentiel	5	65326	65282	66895	66744
Déclin exponentiel	6	69875	68514	70025	70036
Déclin exponentiel	7	63015	62829	65987	65521
Déclin exponentiel	8	61529	61525	63541	63878
Déclin exponentiel	9	68798	68587	70028	69987

Déclin exponentiel	10	66685	66572	68954	68597
Expansion	1	81328	81189	81126	81091
Expansion	2	81200	80702	80712	80685
Expansion	3	81954	80410	80391	80352
Expansion	4	81872	80853	80827	80745
Expansion	5	81604	80367	80445	80370
Expansion	6	81214	80312	80294	80292
Expansion	7	81607	80108	80120	80096
Expansion	8	81998	80121	80110	80101
Expansion	9	81808	80069	80052	80066
Expansion	10	81157	80734	80741	80716

Le modèle présentant la plus faible valeur de DIC, indiqué en gras pour chaque répliat, est considéré comme le mieux ajusté aux données (Spiegelhalter et al., 2002). Une différence de DIC d'au moins 5 points est considérée comme significative. Dans le cas contraire, il est impossible de distinguer le meilleur modèle.

Tableau II.B-S4. Modes et intervalle HPD du nombre de changements démographiques inférés par la méthode non-paramétrique pour les séquences autosomales simulées.

Scénario	Réplicat	HPD 5%	Mode	HPD 95%
Constant	1	0	0	2
Constant	2	0	0	1
Constant	3	0	1	2
Constant	4	0	0	2
Constant	5	0	0	2
Constant	6	0	1	2
Constant	7	0	1	2
Constant	8	0	1	1
Constant	9	0	0	0
Constant	10	0	0	0
Croissance exponentielle	1	1	1	2
Croissance exponentielle	2	1	1	2
Croissance exponentielle	3	1	1	2
Croissance exponentielle	4	1	1	3
Croissance exponentielle	5	1	1	2
Croissance exponentielle	6	1	1	2
Croissance exponentielle	7	1	1	2
Croissance exponentielle	8	1	1	2
Croissance exponentielle	9	1	1	2
Croissance exponentielle	10	1	1	3
Déclin exponentiel	1	1	1	2
Déclin exponentiel	2	1	1	3
Déclin exponentiel	3	1	1	2
Déclin exponentiel	4	1	1	2
Déclin exponentiel	5	0	1	2
Déclin exponentiel	6	1	1	2
Déclin exponentiel	7	1	1	2
Déclin exponentiel	8	0	1	3
Déclin exponentiel	9	0	1	3
Déclin exponentiel	10	1	1	2

Expansion	1	1	1	2
Expansion	2	1	1	3
Expansion	3	1	1	3
Expansion	4	1	1	2
Expansion	5	1	1	1
Expansion	6	1	1	3
Expansion	7	1	1	2
Expansion	8	1	1	2
Expansion	9	1	1	2
Expansion	10	1	1	2

On considèrera ici que les évènements d'expansion ou de déclin observé sur les EBSP sont significatifs lorsque l'intervalle HPD du nombre de changements démographiques n'inclus pas la valeur de 0 (Heled et Drummond, 2010).

II.C. Inférences démographiques à partir de différents types de marqueurs simulés sous différents scénarios d'expansions.

II.C.i. Matériel et méthodes.

Production des données.

Comme précédemment, nous avons utilisé le logiciel SIMCOAL v2 (Laval et Excoffier, 2004) pour la simulation des données.

Nous avons tout d'abord simulé des données suivant trois catégories de scénarios incluant un seul évènement d'expansion, variable par son ancienneté : (i) Une expansion « récente », débutant il y a 200 générations, soit 5000 ans (ii) Une expansion « intermédiaire », débutant il y a 800 générations, soit 20000 ans (iii) Une expansion « ancienne », débutant il y a 2000 générations, soit 50000 ans. Pour chaque cas, nous avons également fait varier l'intensité de l'expansion simulée, avec un taux de croissance $g=10^{-3}$ par génération (g_1) ou $g=10^{-2}$ par génération (g_2). Nous avons ensuite simulé des données suivant des scénarios à deux expansions successives plus ou moins espacées dans le temps : (i) deux expansions débutant il y a respectivement 200 et 800 générations, soit 5000 et 20000 ans, (ii) deux expansions débutant il y a respectivement 200 et 2000 générations, soit 5000 et 50000 ans. Pour chacun des deux cas, nous avons fait également varier les taux de croissance utilisés : (i) 10^{-4} par génération pour l'expansion la plus ancienne et 10^{-2} par génération pour la suivante (g_1'), (ii) 10^{-3} par génération pour l'expansion la plus ancienne et 10^{-2} par génération pour la suivante (g_2'). Les scénarii simulés et les paramètres utilisés sont détaillés dans le **Tableau II.C-1**.

Tableau II.C-1. Description des scénarii simulés et paramètres démographiques utilisés pour les différents marqueurs.

Marqueur simulé	Scénario simulé	N_e simulé	g simulé	t simulé
Séquences HVS-I	Expansion ancienne g1	50 000	10^{-3}	2000
Séquences HVS-I	Expansion ancienne g2	50 000	10^{-2}	2000
Séquences HVS-I	Expansion intermédiaire g1	50 000	10^{-3}	800
Séquences HVS-I	Expansion intermédiaire g2	50 000	10^{-2}	800
Séquences HVS-I	Expansion récente g1	50 000	10^{-3}	200
Séquences HVS-I	Expansion récente g2	50 000	10^{-2}	200
Séquences HVS-I	Intermédiaire + récente	50 000	$10^{-4} + 10^{-2}$	800+ 200
Séquences HVS-I	Intermédiaire + récente	50 000	$10^{-3} + 10^{-2}$	800+ 200
Séquences HVS-I	Ancienne + récente	50 000	$10^{-4} + 10^{-2}$	2000+ 200
Séquences HVS-I	Ancienne + récente	50 000	$10^{-3} + 10^{-2}$	2000+ 200
Séquences autosomales	Expansion ancienne g1	100 000	10^{-3}	2000
Séquences autosomales	Expansion ancienne g2	100 000	10^{-2}	2000
Séquences autosomales	Expansion intermédiaire g1	100 000	10^{-3}	800
Séquences autosomales	Expansion intermédiaire g2	100 000	10^{-2}	800
Séquences autosomales	Expansion récente g1	100 000	10^{-3}	200
Séquences autosomales	Expansion récente g2	100 000	10^{-2}	200
Séquences autosomales	Intermédiaire + récente	100 000	$10^{-4} + 10^{-2}$	800+ 200
Séquences autosomales	Intermédiaire + récente	100 000	$10^{-3} + 10^{-2}$	800+ 200
Séquences autosomales	Ancienne + récente	100 000	$10^{-4} + 10^{-2}$	2000+ 200
Séquences autosomales	Ancienne + récente	100 000	$10^{-3} + 10^{-2}$	2000+ 200
Microsatellites Y	Expansion ancienne g1	50 000	10^{-3}	2000
Microsatellites Y	Expansion ancienne g2	50 000	10^{-2}	2000
Microsatellites Y	Expansion intermédiaire g1	50 000	10^{-3}	800
Microsatellites Y	Expansion intermédiaire g2	50 000	10^{-2}	800
Microsatellites Y	Expansion récente g1	50 000	10^{-3}	200
Microsatellites Y	Expansion récente g2	50 000	10^{-2}	200
Microsatellites Y	Intermédiaire + récente g1'	50 000	$10^{-4} + 10^{-2}$	800+ 200
Microsatellites Y	Intermédiaire + récente g2'	50 000	$10^{-3} + 10^{-2}$	800+ 200
Microsatellites Y	Ancienne + récente g1'	50 000	$10^{-4} + 10^{-2}$	2000+ 200
Microsatellites Y	Ancienne + récente g2'	50 000	$10^{-3} + 10^{-2}$	2000+ 200
Microsatellites autosomaux	Expansion ancienne g1	100 000	10^{-3}	2000
Microsatellites autosomaux	Expansion ancienne g2	100 000	10^{-2}	2000
Microsatellites autosomaux	Expansion intermédiaire g1	100 000	10^{-3}	800
Microsatellites autosomaux	Expansion intermédiaire g2	100 000	10^{-2}	800
Microsatellites autosomaux	Expansion récente g1	100 000	10^{-3}	200
Microsatellites autosomaux	Expansion récente g2	100 000	10^{-2}	200
Microsatellites autosomaux	Intermédiaire + récente g1'	100 000	$10^{-4} + 10^{-2}$	800+ 200
Microsatellites autosomaux	Intermédiaire + récente g2'	100 000	$10^{-3} + 10^{-2}$	800+ 200
Microsatellites autosomaux	Ancienne + récente g1'	100 000	$10^{-4} + 10^{-2}$	2000+ 200
Microsatellites autosomaux	Ancienne + récente g2'	100 000	$10^{-3} + 10^{-2}$	2000+ 200

Nous avons simulé 100 individus ($N=100$) pour chaque scénario et chaque type de marqueur, soit, pour chacun des dix scénarios: (i) 100x 20 régions diploïdes et indépendantes de 1300 pb, simulant des séquences autosomales, (ii) 100 régions haploïdes de 400 pb, simulant la région mitochondriale HVS-I, (iii) 100x 20 locus microsatellites indépendants diploïdes, simulant des microsatellites autosomaux, ainsi que (iv) 100x 10 locus microsatellites liés et haploïdes, simulant les marqueurs microsatellites du chromosome Y. Nous présenterons ici les résultats d'un seul réplicat pour chaque marqueur pour chacun des six scénarios, ce qui constitue en l'état un travail préliminaire. Nous avons fixé les taux de mutation à respectivement $\mu=10^{-5}$ par génération et par site (Howell *et al.*, 1996; Heyer *et al.*, 2001) et $\mu = 2,5 \times 10^{-8}$ par génération et par site (Pluzhnikov *et al.*, 2002) pour les séquences mitochondriales et autosomales. Pour les microsatellites autosomaux, nous avons utilisé un taux de $\mu = 10^{-4}$ par génération et par site, ce qui correspond à la borne supérieure de la distribution *a priori* utilisée pour l'analyse des données réelles (**Chapitre I.B.v**). Pour le chromosome Y, nous avons considéré un taux de $\mu = 2.1 \times 10^{-3}$ par génération et par site (Heyer *et al.*, 1997). Enfin, nous avons spécifié un effectif efficace actuel de $N_e = 50\ 000$ individus pour les marqueurs uniparentaux (séquences mitochondriales et chromosome Y) et $2N_e = 100\ 000$ individus pour les séquences et les microsatellites autosomaux.

Analyse des données

Les données ont été analysées suivant les mêmes méthodes et protocoles que lors des analyses précédentes (**Chapitre II.B.i**). Cependant, le calcul des DIC pour chaque modèle démographique étant encore en cours, nous présenterons et discuterons ici pour la méthode paramétrique uniquement les paramètres inférés avec le modèle de croissance exponentielle débutant à un temps donné à partir d'une population d'effectif constant (modèle d'expansion). En effet, les analyses précédentes ont déjà montré l'efficacité du DIC pour la sélection du modèle démographique approprié. Dans le cadre de ce travail préliminaire, on peut donc raisonnablement supposer que le modèle d'expansion, suivant lequel les données ont été simulées, serait également sélectionné dans la grande majorité des cas ici.

II.C.ii. Resultats.

Scénari à un évènement d'expansion.

Les paramètres démographiques inférés par la méthode paramétrique à partir du modèle d'expansion sont présentés dans le **Tableau II.C-2**. Les paramètres utilisés pour simuler les données sont toujours inclus dans l'intervalle de confiance des paramètres estimés, quelque soient le scénario et le type de marqueur.

Tableau II.C-2. Modes et intervalles HPD des paramètres estimés avec la méthode paramétrique sous un modèle d'expansion à partir d'une population constante, comparés aux paramètres simulés.

Marqueur simulé	Expansion simulée	N_e simulé	g simulé	t simulé	N_e estimé	g estimé	t estimé
Séquences HVS-I	Exp. ancienne g1	50 000	10^{-3}	2000	45316 [27468 - 99521]	7.97×10^{-04} [2.56×10^{-04} - 7.52×10^{-03}]	1932 [1592 - 3004]
Séquences HVS-I	Exp. ancienne g2	50 000	10^{-2}	2000	43353 [20676 - 80156]	3.08×10^{-02} [6.15×10^{-03} - 7.26×10^{-02}]	2311 [1442 - 5122]
Séquences HVS-I	Exp. intermédiaire g1	50 000	10^{-3}	800	47391 [27778 - 94350]	8.27×10^{-04} [1.63×10^{-04} - 5.74×10^{-03}]	941 [352 - 1279]
Séquences HVS-I	Exp. intermédiaire g2	50 000	10^{-2}	800	63362 [26637 - 84787]	9.89×10^{-03} [4.27×10^{-03} - 5.05×10^{-02}]	828 [346 - 1168]
Séquences HVS-I	Exp. récente g1	50 000	10^{-3}	200	62597 [17508 - 81594]	1.30×10^{-03} [7.63×10^{-04} - 9.35×10^{-03}]	191 [41 - 273]
Séquences HVS-I	Exp. récente g2	50 000	10^{-2}	200	59367 [21326 - 93141]	9.58×10^{-03} [2.09×10^{-03} - 4.82×10^{-02}]	178 [52 - 250]
Séquences autosomales	Exp. ancienne g1	100 000	10^{-3}	2000	107968 [89606 - 163582]	8.65×10^{-04} [2.77×10^{-04} - 1.65×10^{-03}]	2272 [1313 - 4878]
Séquences autosomales	Exp. ancienne g2	100 000	10^{-2}	2000	112497 [79937 - 166528]	9.49×10^{-03} [1.83×10^{-03} - 6.14×10^{-02}]	2104 [1519 - 3271]
Séquences autosomales	Exp. intermédiaire g1	100 000	10^{-3}	800	109722 [94155 - 150652]	2.52×10^{-03} [8.26×10^{-04} - 9.65×10^{-03}]	863 [426 - 1516]
Séquences autosomales	Exp. intermédiaire g2	100 000	10^{-2}	800	101311 [80476 - 149625]	1.37×10^{-02} [6.72×10^{-03} - 7.19×10^{-02}]	609 [475 - 1389]
Séquences autosomales	Exp. récente g1	100 000	10^{-3}	200	92194 [93105 - 138547]	2.19×10^{-03} [6.68×10^{-04} - 8.53×10^{-03}]	173 [90 - 339]
Séquences autosomales	Exp. récente g2	100 000	10^{-2}	200	87854 [87706 - 163594]	2.06×10^{-02} [7.13×10^{-03} - 6.91×10^{-02}]	217 [66 - 362]
Microsatellites Y	Exp. ancienne g1	50 000	10^{-3}	2000	36939 [29718 - 105315]	8.70×10^{-04} [1.95×10^{-04} - 5.32×10^{-03}]	1809 [1524 - 7339]
Microsatellites Y	Exp. ancienne g2	50 000	10^{-2}	2000	36448 [29008 - 108398]	9.41×10^{-03} [2.79×10^{-03} - 5.27×10^{-02}]	1910 [1366 - 7144]
Microsatellites Y	Exp. intermédiaire g1	50 000	10^{-3}	800	62735 [22515 - 117842]	8.43×10^{-03} [8.62×10^{-04} - 6.55×10^{-02}]	693 [489 - 1359]
Microsatellites Y	Exp. intermédiaire g2	50 000	10^{-2}	800	61573 [20826 - 92092]	9.08×10^{-03} [2.68×10^{-03} - 4.89×10^{-02}]	723 [494 - 1540]
Microsatellites Y	Exp. récente g1	50 000	10^{-3}	200	37312 [17518 - 99270]	2.73×10^{-03} [7.33×10^{-04} - 8.93×10^{-03}]	175 [92 - 248]
Microsatellites Y	Exp. récente g2	50 000	10^{-2}	200	48720 [21302 - 109006]	8.46×10^{-03} [3.78×10^{-03} - 3.99×10^{-02}]	179 [44 - 291]
Microsatellites autosomaux	Exp. ancienne g1	100 000	10^{-3}	2000	88224 [87759 - 189614]	1.30×10^{-03} [7.25×10^{-04} - 9.63×10^{-03}]	1746 [1424 - 6379]
Microsatellites autosomaux	Exp. ancienne g2	100 000	10^{-2}	2000	121595 [80622 - 183870]	8.96×10^{-03} [1.60×10^{-03} - 5.69×10^{-02}]	1931 [1168 - 5293]
Microsatellites autosomaux	Exp. intermédiaire g1	100 000	10^{-3}	800	115597 [98334 - 141147]	2.14×10^{-03} [6.44×10^{-04} - 8.46×10^{-03}]	682 [413 - 1268]
Microsatellites autosomaux	Exp. intermédiaire g2	100 000	10^{-2}	800	108573 [78803 - 137871]	8.90×10^{-03} [2.97×10^{-03} - 3.94×10^{-02}]	992 [409 - 1105]
Microsatellites autosomaux	Exp. récente g1	100 000	10^{-3}	200	98244 [79462 - 123710]	2.72×10^{-03} [1.86×10^{-04} - 3.66×10^{-02}]	209 [59 - 343]
Microsatellites autosomaux	Exp. récente g2	100 000	10^{-2}	200	86228 [83792 - 131068]	8.59×10^{-03} [2.44×10^{-03} - 2.93×10^{-02}]	165 [84 - 375]

Ne : Effectif efficace actuel, *g* : Taux de croissance, *t* : Début des expansions démographiques. Les temps et les taux de croissance sont exprimés en générations.

En ce qui concerne les données de séquences HVS-I et autosomales, la méthode non-paramétrique nous permet également d'inférer de manière satisfaisante les différents scénarios d'expansion. En effet, l'intervalle HPD du nombre de changements démographiques inférés n'inclut dans aucun cas la valeur de 0, mais inclut toujours la valeur de 1, ce qui indique un signal d'expansion significatif (**Tableau II.C-S1**). De plus, l'inspection visuelle des graphiques permet de distinguer sans ambiguïté les différents scénarios, y compris en tenant compte des intervalles HPD (**Figure II.C-1A et B**).

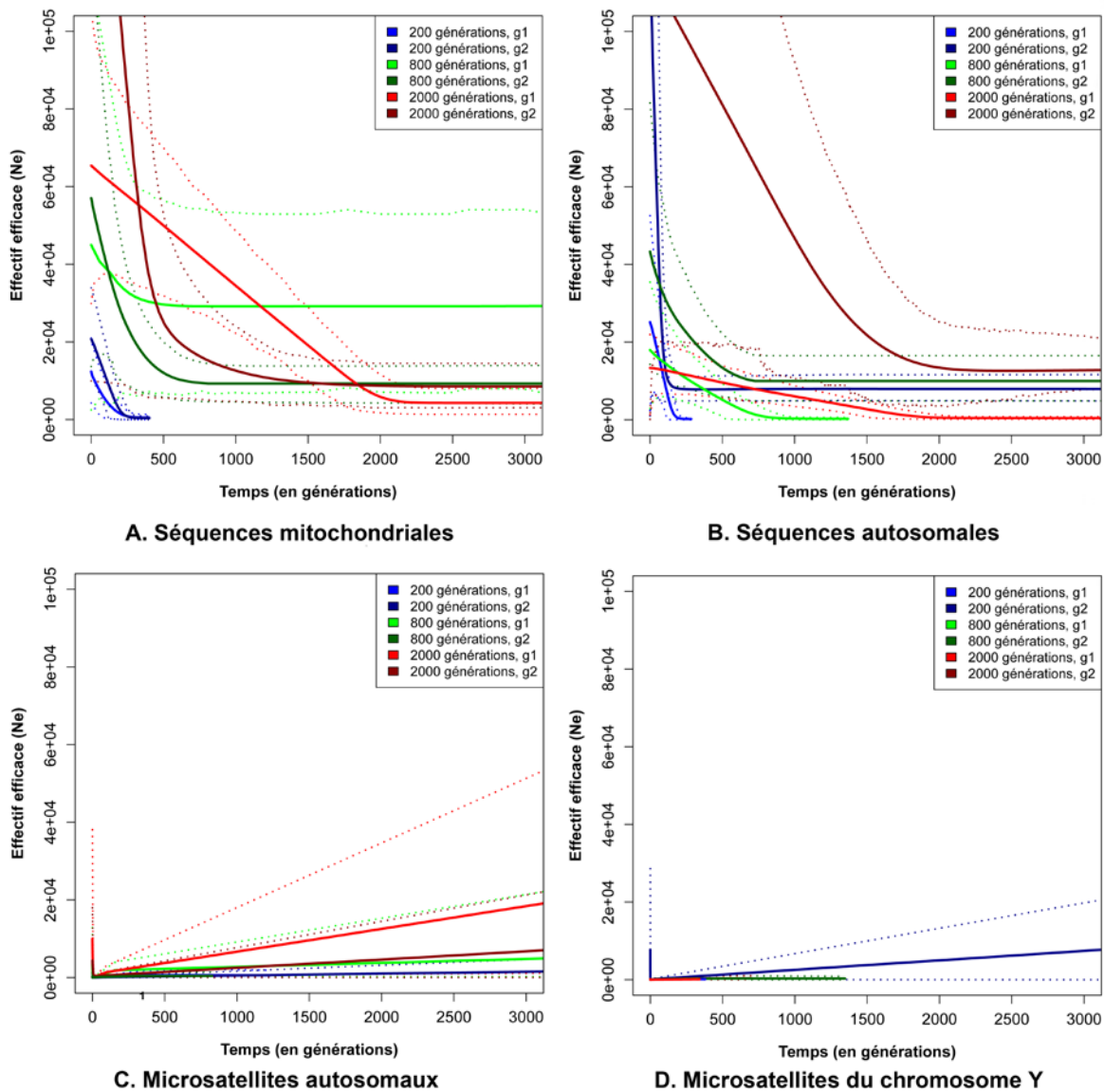


Figure 2.C-1. EBSP inférés à partir des différents marqueurs simulés sous des scénarios d'expansions variables par leur intensité et leur ancienneté. Taux de croissance : $g1 = 10^{-3}$ par génération, $g2 = 10^{-2}$ par génération.

A l'inverse, en ce qui concerne les données microsattellites, la méthode non-paramétrique EBSP ne nous permet pas de détecter les expansions simulées. En effet, l'intervalle HPD du nombre de changements démographiques inclut dans tous les cas la valeur de zéro (**Tableau II.C-S1**), et l'inspection des graphiques ne montre pas de signaux d'expansion (**Figure II.C-1C et D**).

Scénarii à deux évènements d'expansions successives.

Les paramètres démographiques inférés par la méthode paramétrique à partir du modèle d'expansion sont présentés dans le **Tableau II.C-3**.

Tableau II.C-3. Modes et intervalles HPD des paramètres estimés avec la méthode paramétrique sous un modèle d'expansion à partir d'une population constante, comparés aux paramètres simulés.

Marqueur simulé	Expansions simulées	N_e simulé	g simulé	t simulé	N_e estimé	g estimé	t estimé
Séquences HVS-I	Intermédiaire + récente	50 000	$10^{-4} + 10^{-2}$	800+ 200	126065 [47682 - 176234]	1.58×10^{-03} [7.17×10^{-04} - 7.61×10^{-03}]	821 [289 - 1558]
Séquences HVS-I	Intermédiaire + récente	50 000	$10^{-3} + 10^{-2}$	800+ 200	110562 [54249 - 163282]	8.45×10^{-03} [1.64×10^{-03} - 8.66×10^{-02}]	783 [313 - 1482]
Séquences HVS-I	Ancienne + récente	50 000	$10^{-4} + 10^{-2}$	2000+ 200	84253 [48797 - 153935]	4.27×10^{-03} [5.26×10^{-04} - 1.23×10^{-02}]	1632 [706 - 2506]
Séquences HVS-I	Ancienne + récente	50 000	$10^{-3} + 10^{-2}$	2000+ 200	89254 [30446 - 169225]	6.28×10^{-03} [2.49×10^{-03} - 7.13×10^{-02}]	1769 [819 - 3001]
Séquences autosomales	Intermédiaire + récente	100 000	$10^{-4} + 10^{-2}$	800+ 200	133822 [42927 - 188906]	1.53×10^{-03} [8.21×10^{-04} - 7.52×10^{-03}]	749 [251 - 1406]
Séquences autosomales	Intermédiaire + récente	100 000	$10^{-3} + 10^{-2}$	800+ 200	140662 [52401 - 199696]	8.64×10^{-03} [4.73×10^{-03} - 4.67×10^{-02}]	737 [283 - 1391]
Séquences autosomales	Ancienne + récente	100 000	$10^{-4} + 10^{-2}$	2000+ 200	125856 [41988 - 180898]	7.17×10^{-03} [1.86×10^{-03} - 1.85×10^{-02}]	1831 [675 - 2379]
Séquences autosomales	Ancienne + récente	100 000	$10^{-3} + 10^{-2}$	2000+ 200	110663 [57740 - 208073]	3.65×10^{-03} [5.92×10^{-03} - 4.26×10^{-02}]	1752 [806 - 2752]
Microsatellites Y	Intermédiaire + récente	50 000	$10^{-4} + 10^{-2}$	800+ 200	96512 [37131 - 146285]	5.48×10^{-03} [8.16×10^{-04} - 5.93×10^{-02}]	268 [86 - 817]
Microsatellites Y	Intermédiaire + récente	50 000	$10^{-3} + 10^{-2}$	800+ 200	72692 [36843 - 147244]	6.53×10^{-03} [1.68×10^{-03} - 6.83×10^{-02}]	241 [52 - 531]
Microsatellites Y	Ancienne + récente	50 000	$10^{-4} + 10^{-2}$	2000+ 200	83261 [35789 - 136256]	5.28×10^{-03} [6.45×10^{-04} - 4.66×10^{-02}]	287 [64 - 1067]
Microsatellites Y	Ancienne + récente	50 000	$10^{-3} + 10^{-2}$	2000+ 200	100226 [32612 - 140667]	7.29×10^{-03} [2.73×10^{-03} - 5.29×10^{-02}]	296 [76 - 2083]
Microsatellites autosomaux	Intermédiaire + récente	100 000	$10^{-4} + 10^{-2}$	800+ 200	152066 [39291 - 208263]	2.37×10^{-03} [3.47×10^{-04} - 9.47×10^{-03}]	196 [62 - 809]
Microsatellites autosomaux	Intermédiaire + récente	100 000	$10^{-3} + 10^{-2}$	800+ 200	140366 [32870 - 197328]	6.92×10^{-03} [1.68×10^{-03} - 5.49×10^{-02}]	226 [91 - 555]
Microsatellites autosomaux	Ancienne + récente	100 000	$10^{-4} + 10^{-2}$	2000+ 200	97332 [37949 - 206934]	7.50×10^{-03} [8.25×10^{-04} - 8.48×10^{-02}]	219 [59 - 1170]
Microsatellites autosomaux	Ancienne + récente	100 000	$10^{-3} + 10^{-2}$	2000+ 200	138928 [36711 - 210005]	5.52×10^{-03} [1.62×10^{-03} - 5.19×10^{-02}]	273 [43 - 937]

Ne : Effectif efficace actuel, *g* : Taux de croissance, *t* : Début des expansions démographiques. Les temps et les taux de croissance sont exprimés en générations.

Pour les données de séquences HVS-I et autosomales, dans le cas de deux expansions successives débutant à des périodes plus ou moins espacées dans le temps (il y a respectivement 200 et 800 générations ou 200 et 2000 générations), l'intervalle HPD du début d'expansion inclut la plus ancienne dans tous les cas, mais jamais la plus récente. Les taux de croissance estimés sont intermédiaires entre celui de la première expansion simulée et celui de la deuxième, avec un intervalle HPD incluant dans tous les cas seulement le taux de croissance le plus fort, qui est aussi ici le plus récent. Les effectifs efficaces estimés ont tendance à être surestimés. Notamment, l'effectif efficace utilisé pour simuler les données n'est pas inclus dans l'intervalle HDP estimé dans les trois cas correspondant à une expansion intermédiaire ($t=800$ générations) ou ancienne ($t=2000$ générations) forte ($g=10^{-3}$ par génération) suivie d'une expansion plus récente ($t=200$ générations, $g=10^{-2}$ par génération).

Avec la méthode non-paramétrique EBSP, l'expansion simulée la plus ancienne est correctement inférée quel que soit le scénario considéré (**Figure 2.C-1 A et B**). Dans le cas d'une expansion intermédiaire ($t=800$ générations) suivie d'une expansion récente ($t=200$ générations), ni les séquences autosomales ni les séquences HVS-I ne révèlent de traces de l'expansion la plus récente. Lorsque les deux événements sont plus espacés dans le temps (2000 et 200 générations), la présence et l'importance du signal correspondant à l'expansion simulée la plus récente dépend de l'intensité de l'expansion la plus ancienne. Avec un taux de croissance de $g=10^{-4}$ par génération pour l'expansion ancienne, on peut distinguer visuellement grâce aux séquences mitochondriales une sensible accélération récente de la croissance. Cependant, cette accélération est progressive et ne peut donc être datée de manière rigoureuse. A l'inverse, si l'expansion ancienne est plus marquée ($g=10^{-3}$ par génération), on n'observe alors aucune trace d'expansion plus récente. Avec les séquences autosomales, lorsque l'expansion ancienne est simulée avec un taux de croissance de $g=10^{-4}$ par génération, on observe, conformément à l'attendu, une accélération de la croissance il y a environ $t=200$ générations. Lorsque l'expansion ancienne est plus intense ($g=10^{-3}$), on observe une accélération progressive. Dans tous les cas, l'intervalle HPD du nombre de changement démographiques inférés n'inclut pas la valeur de 0, ce qui indique au moins un signal significatif, mais comprend à la fois les valeurs 1 et 2, ce qui ne permet pas de distinguer entre 1 ou 2 signaux d'expansion (**Tableau 2.C-S2**).

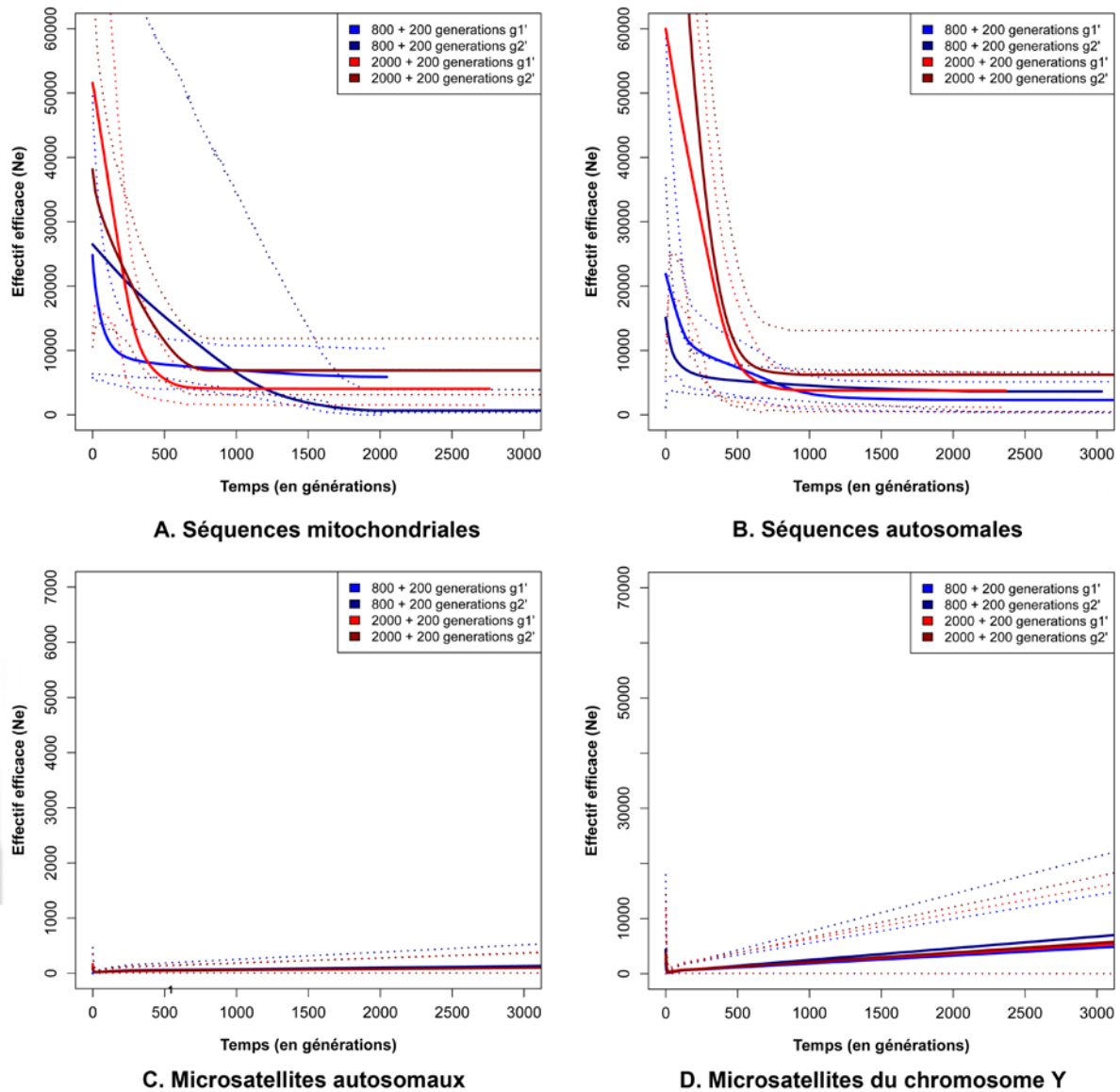


Figure 2.C-1. EBSI inférés à partir des différents marqueurs simulés sous des scénarios comprenant 2 expansions variables par leur intensité et leur espacement dans le temps. $g1' = 10^{-4}$ et 10^{-2} par génération, $g2' = 10^{-3}$ et 10^{-2} par génération.

A l'inverse, avec les données microsatellites Y et autosomales, l'intervalle HPD du début d'expansion inclus systématiquement l'expansion la plus récente. Les effectifs efficaces estimés sont également toujours supérieurs à ceux utilisés pour les simulations, bien que ceux-ci soit cette fois inclus dans l'intervalle HPD. Les taux de croissance inférés sont intermédiaires entre celui de la première expansion simulée et celui de la deuxième, avec un intervalle HPD incluant dans tous les cas seulement le taux de croissance le plus fort (qui est aussi le plus récent). Comme pour les scénarios à une seule expansion, la méthode non-

paramétrique (EBSF) ne nous permet de détecter aucun des évènements de croissance simulés ici pour les données microsatellites (Figure II.C-2C, D et Tableau II.C-S1).

II.C.iii. Tableau supplémentaire.

Tableau II.C-S1. Modes et intervalles HPD du nombre de changements démographiques inférés à partir de la méthode non-paramétrique pour les différents scénarios.

Marqueur simulé	Scénario	HPD 5%	Mode	HPD 95%
Séquences HVS-I	Expansion ancienne g1	1	1	1
Séquences HVS-I	Expansion ancienne g2	1	1	1
Séquences HVS-I	Expansion intermédiaire g1	1	1	1
Séquences HVS-I	Expansion intermédiaire g2	1	1	1
Séquences HVS-I	Expansion récente g1	1	1	1
Séquences HVS-I	Expansion récente g2	1	1	1
Séquences autosomales	Expansion ancienne g1	1	1	1
Séquences autosomales	Expansion ancienne g2	1	1	1
Séquences autosomales	Expansion intermédiaire g1	1	1	1
Séquences autosomales	Expansion intermédiaire g2	1	1	1
Séquences autosomales	Expansion récente g1	1	1	1
Séquences autosomales	Expansion récente g2	1	1	1
Microsatellites Y	Expansion ancienne g1	0	0	1
Microsatellites Y	Expansion ancienne g2	0	0	1
Microsatellites Y	Expansion intermédiaire g1	0	0	1
Microsatellites Y	Expansion intermédiaire g2	0	0	1
Microsatellites Y	Expansion récente g1	0	0	1
Microsatellites Y	Expansion récente g2	0	1	1
Microsatellites autosomales	Expansion ancienne g1	0	0	1
Microsatellites autosomales	Expansion ancienne g2	0	0	1
Microsatellites autosomales	Expansion intermédiaire g1	0	0	1
Microsatellites autosomales	Expansion intermédiaire g2	0	0	1
Microsatellites autosomales	Expansion récente g1	0	0	1
Microsatellites autosomales	Expansion récente g2	0	0	1
Séquences HVS-I	Expansion intermédiaire g1' + récente	1	1	2
Séquences HVS-I	Expansion intermédiaire g2' + récente	1	1	2
Séquences HVS-I	Expansion ancienne g1' + récente	1	1	2
Séquences HVS-I	Expansion ancienne g2' + récente	1	1	2
Séquences autosomales	Expansion intermédiaire g1' + récente	1	1	2
Séquences autosomales	Expansion intermédiaire g2' + récente	1	1	2
Séquences autosomales	Expansion ancienne g1' + récente	1	2	2
Séquences autosomales	Expansion ancienne g2' + récente	1	1	2
Microsatellites Y	Expansion intermédiaire g1' + récente	0	0	1
Microsatellites Y	Expansion intermédiaire g2' + récente	0	0	1
Microsatellites Y	Expansion ancienne g1' + récente	0	0	1
Microsatellites Y	Expansion ancienne g2' + récente	0	0	1
Microsatellites autosomales	Expansion intermédiaire g1' + récente	0	0	1
Microsatellites autosomales	Expansion intermédiaire g2' + récente	0	0	1
Microsatellites autosomales	Expansion ancienne g1' + récente	0	1	1
Microsatellites autosomales	Expansion ancienne g2' + récente	0	0	1

II.D. Discussion et conclusions des analyses sur données simulées.

Premièrement, l'analyse de séquences HVS-I et autosomales simulées suivant différents scénarios simples (effectif constant, croissance exponentielle, décroissance exponentielle, croissance débutée il y a 20 000 ans à partir d'une population d'effectif constant) nous confirme l'aptitude des méthodes testées à inférer correctement la nature (dans sept à neuf cas sur dix selon les scénarii) et les paramètres (dans dix cas sur dix pour tous les scénarii) des évènements simulés. Il est à noter que nous n'avons détecté aucun faux positif, c'est-à-dire aucun faux signal d'expansion ou de déclin. Dans le cas de données simulant un déclin démographique, celui-ci a été détecté significativement dans sept cas sur dix avec les séquences mitochondriales, et dix cas sur dix avec les séquences autosomales. En complément des analyses sur données réelles présentées précédemment (**Chapitre I.B**), ce dernier résultat suggère une plus grande sensibilité des analyses à partir des séquences autosomales par rapport à la région HVS-I pour la détection d'évènements de déclin démographique. En effet, lors de l'analyse des données réelles, nous avons inféré des signaux de contractions pour deux populations (Japonais et Danois) à partir des autosomes mais pas à partir de la région HVS-I.

Dans un deuxième temps, l'analyse de différents marqueurs (séquences HVS-I, séquences autosomales, microsatellites du chromosome Y et microsatellites autosomaux) simulés sous différents scénarios incluant un seul évènement d'expansion nous montre l'aptitude de la méthode paramétrique à estimer correctement différentes valeurs de paramètres démographiques (*i.e.* taux de croissance de 10^{-3} ou 10^{-2} par génération, ancienneté des expansions comprise entre 2000 et 200 générations, soit 50 000 et 5 000 ans), et ce quelque soit le type de marqueur utilisé. Les résultats de la méthode non-paramétrique sont également conformes aux scénarii simulés pour les séquences autosomales et mitochondriales. Cependant, nous avons montré l'incapacité de détecter les évènements d'expansions par la méthode EBSP à partir de données microsatellites pour l'ensemble des scénarios étudiés. Ce constat avait également été fait précédemment dans le cadre de l'analyse des données réelles (**Chapitre I.C et I.D**). Cela pourrait être dû au fait que les marqueurs microsatellites soient moins informatifs que les séquences pour reconstituer les arbres de coalescence. Dans le cas des microsatellites du chromosome Y, le fait que les différents marqueurs ne soit pas indépendants devrait permettre d'utiliser une autre méthode non-paramétrique également implémentée dans BEAST, les « Bayesian Skyline Plots » (BSP, Drummond *et al.*, 2005).

Cette méthode, déjà utilisée pour l'analyse des données microsatellites réelles dans le cas du chromosome Y (**Chapitre I.D**), sera envisagée par la suite ici afin de compléter nos travaux préliminaires sur données simulées.

Enfin, l'analyse de données simulées suivant des scénarii à deux expansions nous apporte des informations particulièrement intéressantes à la fois d'un point de vue méthodologique et pour l'interprétation de nos précédentes analyses portant sur les données réelles. Premièrement, dans le cas de deux expansions successives, les dates de début d'expansion inférées par la méthode paramétrique tendent vers l'expansion la plus ancienne lorsque l'on analyse des données de séquences (HVS-I ou autosomales), et la plus récente lorsque l'on analyse des données microsatellites (Y ou autosomales). Ce résultat est cohérent avec l'analyse des données réelles, qui nous a permis d'inférer pour les populations sédentaires à la fois des événements d'expansion paléolithiques à partir des données de séquence (**Chapitre I.B**), et néolithiques à partir des microsatellites autosomaux (**Chapitre I.C**) et des microsatellites du chromosome Y (si on considère le taux de mutation de 2.1×10^{-3} par génération, également utilisé ici pour produire les données simulées) (**Chapitre I.D**). L'ensemble de nos résultats sur données réelles et simulées suggère donc deux phases d'expansion successives dans ces populations. Il est important de noter ici que l'existence d'une deuxième phase d'expansion plus récente ou plus ancienne, même si celle-ci ne peut être identifiée en tant que telle, influence les paramètres inférés. Elle entraîne notamment une tendance à surestimer les effectifs efficaces actuels. Cette observation ouvre la voie à une nouvelle hypothèse pour l'interprétation de nos résultats obtenus avec les données de séquence HVS-I réelles en Eurasie. En effet, si les populations d'agriculteurs ont connu dans leur histoire deux phases d'expansions successives, on peut se demander si la différence d'intensité des expansions Paléolithiques entre agriculteurs et éleveurs (**Chapitre I.B**) ne serait pas due en partie à un effet confondant de l'expansion Néolithique ultérieure connue par les agriculteurs. Cette deuxième expansion plus récente, inférée à partir des données microsatellites (**Chapitre I.C**) pourrait avoir entraîné une augmentation de l'intensité du signal de croissance inféré chez les agriculteurs au Paléolithique. Une fois de plus, la méthode non-paramétrique EBSP n'a pas permis de détecter les expansions simulées pour les données microsatellites. A partir des données de séquences, si l'expansion la plus ancienne est correctement identifiée dans tous les cas, le pouvoir de détection de l'expansion la plus récente semble dépendre à la fois de l'écart temporel entre les deux événements et de

l'intensité (taux de croissance) du plus ancien. En effet, une expansion récente pourrait être d'autant plus difficile à détecter qu'elle est précédée d'une expansion forte et rapprochée dans le temps. Cependant, il faut rappeler ici que nos analyses sur données simulées constituent un travail préliminaire. Nous envisageons à l'avenir d'augmenter le nombre de réplicats analysés, ainsi que de tester de plus larges gammes de paramètres, afin de confirmer ou infirmer ces premières hypothèses.

Conclusion générale.

Tout au long de ce travail, l'analyse de données génétiques contemporaines nous a permis d'obtenir de nombreuses informations sur l'histoire démographique des populations humaines. D'un point de vue méthodologique, l'analyse de données simulées nous a permis de vérifier l'efficacité des méthodes utilisées, mais aussi de mieux comprendre les réponses spécifiques des différents marqueurs et d'améliorer notre interprétation des résultats de l'analyse des données réelles.

Premièrement, l'ensemble de nos résultats démontre deux grandes phases d'expansion successives au cours de l'histoire humaine en Afrique et en Eurasie. Contrairement au point de vue majoritaire chez les archéologues et paléanthropologues, qui considèrent l'apparition de l'agriculture et du mode de vie sédentaire comme préalable aux expansions démographiques, l'analyse des données de séquence autosomales et mitochondriales (HVS-I) révèle ici, pour certaines populations, une première phase de croissance intervenue durant la période Paléolithique, donc antérieure à l'apparition de l'agriculture. Grâce aux microsatellites autosomiaux, nous démontrons également une deuxième phase d'expansion plus récente, compatible avec la transition néolithique, chez les populations d'agriculteurs sédentaires mais pas chez les populations nomades. Ce résultat confirme l'association entre mode de vie et régime démographique, déjà suggérée et expliquée par différentes hypothèses selon les paléanthropologues. Notamment, l'intervalle entre les naissances est généralement plus court pour les populations nomades (Short, 1982), ce qui pourrait être dû à un régime alimentaire faible en carbohydrates ne favorisant pas la récupération rapide des femmes après l'accouchement (Valeggia and Ellison, 2009). De plus, après la transition Néolithique, la compétition avec les populations agricoles pour la terre et la fragmentation résultante de l'habitat et des ressources disponibles a pu limiter les possibilités d'expansion pour les populations restées nomades (Diamond, 2002). Néanmoins, à notre connaissance, une seule étude de génétique des populations, basée sur l'analyse de séquences mitochondriales complètes, était parvenue jusqu'à présent à démontrer un épisode d'expansion néolithique dans les populations humaines, et ce uniquement sur le continent Européen (Fu *et al.*, 2012). Nos analyses sur données simulées confirment le fait que, dans le cas de deux expansions successives au cours du temps, les données de séquences tendront à nous renseigner sur l'évènement le plus ancien tandis que les données microsatellites dévoileront l'évènement le plus récent. A l'inverse, l'ancienneté des expansions ne semble pas influencer le pouvoir de

détection de nos méthodes dans le cas d'un évènement unique, du moins dans la gamme de paramètres que nous avons testée.

Nos analyses ont porté sur de nombreuses populations distantes géographiquement et vivant dans des environnements parfois très différents (forêt équatoriale africaine, steppes d'Asie centrale, etc.). De manière prévisible, nous observons un certain degré de variabilité entre les régions étudiées en termes d'histoires démographiques. En Afrique, l'expansion démographique ancienne, bien que précédant la transition Néolithique et l'apparition de l'agriculture, ne concerne que les ancêtres des populations actuellement agricoles. De précédentes études (Verdu *et al.*, 2009 ; Patin *et al.*, 2010) avaient déjà montré une divergence entre les ancêtres des agriculteurs et des chasseurs-cueilleurs contemporains il y a environ 60 000 ans, donc bien avant l'apparition de l'agriculture. L'expansion Paléolithique connue par les ancêtres des populations d'agriculteurs pourrait avoir favorisé le développement de l'agriculture dans ces populations (Sauer, 1952), qui aurait favorisé à son tour une deuxième phase d'expansion au Néolithique. A l'inverse, les populations de chasseurs-cueilleurs contemporaines ne semblent pas avoir connu de grands changements démographiques au Paléolithique ni au Néolithique. Alternativement, si l'on suppose que l'expansion relève de processus spatiaux, les traces d'éventuelles expansions pourraient avoir été effacées par un plus fort degré d'isolation de ces populations (Ray *et al.*, 2003 ; Excoffier, 2004). Bien que les différences de F_{st} et de taux de migration nous apparaissent insuffisantes pour expliquer la différence de régime démographique observée ici entre les deux groupes (**Chapitre I.B.iv**), cette question devrait être traitée dans le futur par de nouvelles analyses sur données simulées.

En Eurasie, l'expansion Paléolithique concerne à la fois les populations agricoles sédentaires et les populations d'éleveurs semi-nomades. Cependant, nous observons ici aussi une différence de régime démographique entre les deux groupes intervenant antérieurement à leur différenciation en termes de mode de vie. En effet, bien que l'agriculture et l'élevage se soient développés à partir du Néolithique, les expansions paléolithiques inférées sont plus fortes chez les populations agricoles sédentaires que chez les éleveurs nomades. De plus fortes expansions chez les ancêtres des agriculteurs pourraient donc avoir favorisé ultérieurement l'apparition de l'agriculture (théorie démographique, Sauer 1952). Alternativement, l'existence d'une deuxième phase d'expansion au Néolithique chez les agriculteurs pourrait avoir agi comme un facteur confondant en augmentant l'intensité du signal d'expansion Paléolithique inféré pour ces populations. Cette deuxième hypothèse est

supportée par nos analyses sur données simulées, qui montrent une tendance à la surestimation des effectifs efficaces actuels après deux expansions successives. Enfin, en Asie centrale, l'analyse des données de séquences mitochondriales et des microsatellites autosomiaux suggère deux phases d'expansions successives (Paléolithique et Néolithique) à la fois chez les agriculteurs sédentaires et les éleveurs nomades. De plus, l'analyse des séquences mitochondriales montre une tendance à des taux de croissance plus forts chez les éleveurs que chez les agriculteurs, contrairement à ce que l'on observe pour le reste de l'Eurasie. Nous avons suggéré qu'un fort taux de migration entre les populations d'éleveurs voisines, dû à la règle culturelle d'exogamie dans ces populations, pourrait avoir agi comme un facteur confondant augmentant les signaux d'expansion inférés chez les populations d'éleveurs. Les populations d'Asie centrale étant patrilocales, il serait logique que ce phénomène n'influence pas les inférences utilisant le chromosome Y, car les migrations entre populations concernent majoritairement les femmes. C'est effectivement ce que nous montrons ici, l'analyse des microsatellites du chromosome Y ne dévoilant pas d'évènement d'expansion chez les éleveurs contrairement aux agriculteurs.

Perspectives

Différentes perspectives sont envisagées pour poursuivre ce travail. En ce qui concerne les données réelles, les nouvelles méthodes de génotypage entraînent un accès de plus en plus facilité à de nouveaux types de données couvrant le génome à une plus grande échelle (puces à ADN, séquences complètes). Notamment, j'ai pu participer durant mon doctorat à la production (étapes préliminaires de contrôle qualité) de données de puces à ADN (Illumina Omni 1) pour 352 individus appartenant à six populations d'Asie centrale et six populations d'Altaï. Bien que les inférences démographiques à partir de données de type SNP (« Single Nucleotide Polymorphism ») posent de nouvelles difficultés méthodologiques (taille des jeux de données produits, prise en compte nécessaire du biais d'ascertainment, *etc.*), plusieurs méthodes ont été récemment proposées (*e.g.* Gutenkunst *et al.*, 2009 ; Mc Evoy *et al.*, 2011 ; Lukic *et al.*, 2012) afin de tirer parti efficacement de ce nouveau type de données et de la quantité considérable d'information qu'elles sont susceptibles de contenir. L'utilisation conjointe de plusieurs types de marqueurs, par exemple de type séquences et microsatellites, pour reconstituer le processus de coalescence, pourrait également apporter plus de précision dans les inférences démographiques, et peut-être la possibilité d'inférer à la fois expansions Paléolithiques et Néolithiques au sein d'une même analyse grâce à la méthode non-

paramétrique. Enfin, il serait intéressant d'étendre nos investigations à d'autres régions du monde, par exemple les Amériques ou l'Océanie, où des populations aux modes de vies contrastés coexistent également.

Enfin, concernant les analyses de données simulées, nous avons déjà montré que la méthode non-paramétrique peut permettre dans certains cas de détecter des traces de deux expansions successives à partir de données de séquences. Cependant, la capacité à détecter l'évènement le plus récent semble dépendre à la fois de l'écart temporel entre les deux évènements et de l'intensité de l'expansion la plus ancienne. En effet, une forte expansion ancienne et un écart temporel limité entre les deux évènements pourrait diminuer la probabilité de détecter l'expansion la plus récente. Par ailleurs, les analyses de séquences autosomales semblent plus sensibles aux évènements de déclin démographique que les analyses de séquences mitochondriales. Ces observations constituent des résultats préliminaires qui devront être complétés par de nouvelles analyses, incluant notamment une plus large gamme de paramètres testés et davantage de réplicats pour chaque scénario. De plus, l'utilisation de la méthode non-paramétrique « Bayesian Skyline Plots » (BSP, Drummond *et al.*, 2005) est envisagée pour les données microsatellites du chromosome Y, pour lesquelles la méthode EBSP s'est avérée inefficace, alors qu'elle a fonctionné de manière satisfaisante sur les données de séquences. Enfin, il serait intéressant d'ajouter à nos analyses des scénarios incluant différents degrés de flux géniques entre populations, afin de vérifier l'impact des processus migratoires sur les inférences démographiques.

Bibliographie

- Ambrose, S. H. (2001) Paleolithic technology and human evolution. *Science*, 291(5509), 1748-1753.
- Arenas, M., Francois, O., Currat, M., Ray, N. & Excoffier, L. (2013) Influence of admixture and paleolithic range contractions on current european diversity gradients. *Mol Biol Evol*, 30(1), 57-61.
- Atkinson, Q. D., Gray, R. D. & Drummond, A. J. (2009) Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in africa. *Proc. R. Soc. B* 276,367-373.
- Austerlitz, F., Kalaydjieva, L. & Heyer, E. (2003) Detecting population growth, selection and inherited fertility from haplotypic data in humans. *Genetics*, 165(3), 1579-86.
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. & Alekseyenko, A. V. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*, 29(9), 2157-67.
- Balaresque, P. L., Ballereau, S. J. & Jobling, M. A. (2007) Challenges in human genetic diversity: Demographic history and adaptation. *Hum Mol Genet*, 16,134-139.
- Bar-Yosef, O. & Belfer-Cohen, A. (1991) From sedentary hunter-gatherers to territorial farmers in the levant. In *between bands and states*. . S. A. Gregg, ed. pp. 181-202.
- Batini, C., Lopes, J., Behar, D. M., Calafell, F., Jorde, L. B., Van Der Veen, L., Quintana-Murci, L., Spedini, G., Destro-Bisol, G. & Comas, D. (2011) Insights into the demographic history of african pygmies from complete mitochondrial genomes. *Mol Biol Evol*, 28(2), 1099-1110.
- Beaumont, M. A. (1999) Detecting population expansion and decline using microsatellites. *Genetics*, 153(4), 2013-29.
- Beaumont, M. A. (2004) Recent developments in genetic data analysis: What can they tell us about human demographic history? *Heredity*, 92(5), 365-379.
- Beaumont, M. A., Zhang, W. & Balding, D. J. (2002) Approximate bayesian computation in population genetics. *Genetics*, 162(4), 2025-35.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B-Meth*, 57(1), 289-300.

- Bermisheva, M., Salimova, A., Korshunova, T., Svyatova, G., Berezina, G., Villems, R. & Khusnutdinova, E. (2002) Mitochondrial DNA diversity in the populations of middle asia and northern caucasus. *Eur J Hum Genet* 10(179-180).
- Binford, L. R. & Chasko, W. J. (1976) Nunamiut demography history: A provocative case. In *demographic anthropology: Quantitative approaches e. B. W. Zubrow, ed. . Albuquerque: University of New Mexico Press., pp. 63-143.*
- Blurton-Jones, N. (1986) Bushman birth spacing: A test for optimal birth intervals. *Ethology and Sociobiology*, 7,91-105.
- Bocquet-Appel, J. P. (2011) When the world's population took off: The springboard of the neolithic demographic transition. *Science*, 333(6042), 560-561.
- Bocquet-Appel, J. P. & Bar-Yosef, O. (2008) *The neolithic demographic transition and its consequences.* Dordrecht: Springer. Bolker, B. (2008) *Ecological models and data in r,* (Princeton, Princeton University Press).
- Bregel, Y. (2003) *An historical atlas of central asia leiden.* Boston: Brill.
- Brunet, F. (1999) La néolithisation en asie centrale: Un état de la question. *Paléorient*, 24,27-48.
- Cann, H. M., De Toma, C., Cazes, L., Legrand, M. F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W. F., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., Chu, J., Carcassi, C., Contu, L., Du, R., Excoffier, L., Ferrara, G. B., Friedlaender, J. S., Groot, H., Gurwitz, D., Jenkins, T., Herrera, R. J., Huang, X., Kidd, J., Kidd, K. K., Langaney, A., Lin, A. A., Mehdi, S. Q., Parham, P., Piazza, A., Pistillo, M. P., Qian, Y., Shu, Q., Xu, J., Zhu, S., Weber, J. L., Greely, H. T., Feldman, M. W., Thomas, G., Dausset, J. & Cavalli-Sforza, L. L. (2002) A human genome diversity cell line panel. *Science*, 296(5566), 261-2.
- Cann, R. L., M., S. & A.C., W. (1987) Mitochondrial DNA and human evolution. *Nature*, 325,31-36.
- Cavalli-Sforza, L. (1986) *African pygmies.* Academic Press.
- Cavalli-Sforza, L. (Ed.) (1996) *The spread of agriculture and nomadic pastoralism: Insights from genetics, linguistics and archeology.* In *the origins and spread of agriculture and pastoralism in eurasia.* David r. Harris. pp.51-70.
- Cavalli-Sforza, L. L. & Feldman, M. W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet*, 33,266-75.

- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) The history and geography of human genes. (Princeton: Princeton University Press). pp. 47.
- Chaix, R., Austerlitz, F., Hegay, T., Quintana-Murci, L. & Heyer, E. (2008) Genetic traces of east-to-west human expansion waves in eurasia. *Am J Phys Anthropol*, 136(3), 309-17.
- Chaix, R., Austerlitz, F., Khegay, T., Jacquesson, S., Hammer, M. F., Heyer, E. & Quintana-Murci, L. (2004) The genetic or mythical ancestry of descent groups: Lessons from the y chromosome. *Am J Hum Genet*, 75(6), 1113-6.
- 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. *Curr Biol*, 17(1), 43-48.
- Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W. & McCabe, A. M. (2009) The last glacial maximum. *Science*, 325(5941), 710-714.
- Cohen, M. N. (1977) The food crisis in prehistory. New Haven, Connecticut: Yale University press. pp. 351.
- Conrad, D. F., Keebler, J. E. M., Depristo, M. A., Lindsay, S. J., Zhang, Y. J., Casals, F., Idaghdour, Y., Hartl, C. L., Torroja, C., Garimella, K. V., Zilversmit, M., Cartwright, R., Rouleau, G. A., Daly, M., Stone, E. A., Hurles, M. E., Awadalla, P. & Project, G. (2011) Variation in genome-wide mutation rates within and between human families. *Nature Genet*, 43(7), 712-U137.
- Cornuet, J. M., Ravigne, V. & Estoup, A. (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software diyabc (v1.0). *Bmc Bioinformatics*, 11 :401.
- Derenko, M. V., Malyarchuk, B. A., Dambueva, I. K., Shaikhaev, G. O., Dorzhu, C. M., Nimaev, D. D. & Zakharov, I. A. (2000) Mitochondrial DNA variation in two south siberian aboriginal populations: Implications for the genetic history of north asia. *Hum Biol*, 72(6), 945-973.
- Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, 418(6898), 700-707.
- Diamond, J. & Bellwood, P. (2003) Farmers and their languages: The first expansions. *Science*, 300(5619), 597-603.

- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. & Weissenbach, J. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, 380(6570), 152-4.
- Dirksen, V. G. & Van Geel, B. (2004) Mid to late holocene climate change and its influence on cultural development in south central siberia. ; in: Scott EM, Alekseev AY, Zaitseva GI (eds): *Impact of the Environment on Human Migration in Eurasia*. NATO Science Series: IV: Earth and Environmental Sciences. Volume 42. Dordrecht: Kluwer Academic Publishers, 2004, pp. 291-307.
- Drummond, A. J. & Rambaut, A. (2007) Beast: Bayesian evolutionary analysis by sampling trees. *Bmc Evol Biol*, 7:214.
- Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*, 22(5), 1185-1192.
- Emerson, B. C., Paradis, E. & Thebaud, C. (2001) Revealing the demographic histories of species using DNA sequences. *Trends Ecol Evol*, 16(12), 707-716.
- Endicott, P., Ho, S. Y., Metspalu, M. & Stringer, C. (2009) Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol*, 24(9), 515-21.
- Excoffier, L. (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: Lessons from the infinite-island model. *Mol Ecol*, 13(4), 853-64.
- Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: A survival guide. *Nature Rev Genet*, 7(10), 745-758.
- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinfo*, 1(47-50).
- Fontaine, M. C., Snirc, A., Frantzis, A., Koutrakis, E., Ozturk, B., Ozturk, A. A. & Austerlitz, F. (2012) History of expansion and anthropogenic collapse in a top marine predator of the black sea estimated from genetic data. *Proc Natl Acad Sci U S A*, 109(38), E2569-76.
- Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. (1996) Origin and evolution of native american mtDNA variation: A reappraisal. *Am J Hum Genet*, 59(4), 935-945.
- Fu, Q., Rudan, P., Paabo, S. & Krause, J. (2012) Complete mitochondrial genomes reveal neolithic expansion into europe. *Plos one*, 7(3), e32473.

- Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-25.
- Fu, Y. X. & Li, W. H. (1993) Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693-709.
- Fu, Y. X. & Li, W. H. (1999) Coalescing into the 21st century: An overview and prospects of coalescent theory. *Theor Pop Biol* 56,1-10.
- Gelfand, A. E. & Dey, D. K. (1994) Bayesian model choice: Asymptotics and exact calculations. *J Roy Stat Soc, B* 56,501-514.
- Gelman, A. & Meng, X. L. (1998) Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Stat. Sci.*, 13(2), 163-185.
- Gelman, A. & Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.*, 7(457-511).
- Gignoux, C. R., Henn, B. M. & Mountain, J. L. (2011) Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci U S A*, 108(15), 6044-9.
- Griffiths, R. C. & Tavaré, S. (1994) Sampling theory for neutral alleles in a varying environment. *Phil Tr Roy Soc London B*, 344(1310), 403-410.
- Gumilyov, L. N. (1967) *Ancient turks*. Moscow: Institute of Ethnology and Anthropology of the Academy of Sciences of USSR.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. (2009) Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *Plos Genet*, 5(10), e1000695.
- Harpending, H., Sherry, S., Rogers, A. & Stoneking, M. (1993) The genetic structure of ancient human populations. *Curr Anthropol*, 34,483-496.
- Heled, J. & Drummond, A. J. (2010) Bayesian inference of species trees from multilocus data. *Mol Biol and Evol*, 27(3), 570-580.
- Henke, J., Henke, L., Chatthopadhyay, P., Kayser, M., Dulmer, M., Cleef, S., Poche, H. & Felske-Zech, H. (2001) Application of y-chromosomal STR haplotypes to forensic genetics. *Croat Med J*, 42(3), 292-7.
- Hewlett, B. (1996) Cultural diversity among african pygmies. In: Kent s (ed) *culturaldiversity among twentieth-century foragers. An african perspective*. Vol 1. Cambridge university press, Cambridge, 215-244.

- Heyer, E., Balaesque, P., Jobling, M. A., Quintana-Murci, L., Chaix, R., Segurel, L., Aldashev, A. & Hegay, T. (2009) Genetic diversity and the emergence of ethnic groups in central asia. *Bmc Genet*, 10, doi:10.1186/1471-2156-10-49.
- Heyer, E., Chaix, R., Pavard, S. & Austerlitz, F. (2012) Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol*, 21(3), 597-612.
- Heyer, E., Puymirat, J., Dietjes, P., Bakker, E. & De Knijff, P. (1997) Estimating y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet*, 6,799-803.
- Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J. & Labuda, D. (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: Study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet*, 63,1113–1126.
- Ho, S. Y. W. & Shapiro, B. (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Res*, 11(3), 423-434.
- Howell, N., Kubacka, I. & Mackey, D. A. (1996) How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet*, 59(3), 501-509.
- Hu, Y., Guo, Y., Qi, D., Zhan, X., Wu, H., Bruford, M. W. & Wei, F. (2011) Genetic structuring and recent demographic history of red pandas (*ailurus fulgens*) inferred from microsatellite and mitochondrial DNA. *Mol Ecol*, 20(13), 2662-75.
- Hudson, R. R. & Kaplan, N. L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics*, 111(1), 147-164.
- Imaizumi, K., Parsons, T. J., Yoshino, M. & Holland, M. M. (2002) A new database of mitochondrial DNA hypervariable regions i and ii sequences from 162 japanese individuals. *Int J Leg Med*, 116(2), 68-73.
- Jorde, L. B., Bamshad, M. J., Watkins, W. S., Zenger, R., Fraley, A. E., Krakowiak, P. A., Carpenter, K. D., Soodyall, H., Jenkins, T. & Rogers, A. R. (1995) Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet*, 57(3), 523-38.
- Kayser, M., Brauer, S., Weiss, G., Schiefenovel, W., Underhill, P. A. & Stoneking, M. (2001) Independent histories of human y chromosomes from melanesia and australia. *Am J Hum Genet*, 68(1), 173-190.
- Kingman, J. F. C. (1982) The coalescent. *Stochast Proc App*, 13,235-248.

- Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L. & Zhang, Y. P. (2003) Phylogeny of east asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet*, 73(3), 671-676.
- Kuhner, M. K. (2006) Lamarc 2.0: Maximum likelihood and bayesian estimation of population parameters. *Bioinformatics*, 22(6), 768-70.
- Kuhner, M. K. (2008) Coalescent genealogy samplers: Windows into population history. *Tr Ecol Evol*, 24(2), 86-93.
- Kuhner, M. K., Yamato, J. & Felsenstein, J. (1995) Estimating effective population size and mutation rate from sequence data using metropolis-hastings sampling. *Genetics*, 140(4), 1421-1430.
- Kumar, S., Skjaeveland, A., Orr, R. J. S., Enger, P., Ruden, T., Mevik, B. H., Burki, F., Botnen, A. & Shalchian-Tabrizi, K. (2009) Air: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *Bmc Bioinformatics*, 10, doi:10.1186/1471-2105-10-357.
- Kumar, V., Langstieh, B. T., Madhavi, K. V., Naidu, V. M., Singh, H. P., Biswas, S., Thangaraj, K., Singh, L. & Reddy, B. M. (2006) Global patterns in human mitochondrial DNA and y-chromosome variation caused by spatial instability of the local cultural processes. *Plos Genet*, 2(4), e53.
- Laval, G. & Excoffier, L. (2004) Simcoal 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, 20(15), 2485-2487.
- Laval, G., Patin, E., Barreiro, L. B. & Quintana-Murci, L. (2010) Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *Plos one*, 5(4), e10284.
- Lavery, S., Moritz, C. & Fielder, D. R. (1996) Genetic patterns suggest exponential population growth in a declining species. *Mol Biol Evol*, 13(8), 1106-1113.
- Lee, R. B. & Daly, R. (1999) *The cambridge encyclopedia of hunters and gatherers*. Cambridge University press.
- Li, H. & Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493-U84.
- Li, W. & Sadler, W. L. (1991) Low nucleotide diversity in man. *Genetics*, 129,513-523.
- Librado, P. & Rozas, J. (2009) Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452.

- Liu, W., Jin, C. Z., Zhang, Y. Q., Cai, Y. J., Xing, S., Wu, X. J., Cheng, H., Edwards, R. L., Pan, W. S., Qin, D. G., An, Z. S., Trinkaus, E. & Wu, X. Z. (2010) Human remains from zhirendong, south china, and modern human emergence in east asia. *Proc Natl Acad Sci U S A*, 107(45), 19201-6.
- Loader, C. (1999) *Local regression and likelihood*, (New-York, Springer).
- Lukic, S. & Hey, J. (2012) Demographic inference using spectral methods on snp data, with an analysis of the human out-of-africa expansion. *Genetics*, 192(2), 619-39.
- Mardis, E. R. (2008) Next-generation DNA sequencing methods. *An Rev Genom Hum Genet*, 9,387-402.
- Martinez-Cruz, B., Vitalis, R., Ségurel, L., Austerlitz, F., Georges, M., Théry, S., Quintana-Murci, L., Hegay, T., Aldashev, A., Nasyrova, F. & Heyer, E. (2011) In the heartland of eurasia: The multilocus genetic landscape of central asian populations. *Eur J Hum Genet*, 19(2), 216-23.
- Mazoyer, M. & Roudart, L. (2006) *A history of world agriculture: From the neolithic age to the current crisis*, translated by J.-H. Membrez, London: Earthscan.
- Mcevoy, B. P., Powell, J. E., Goddard, M. E. & Visscher, P. M. (2011) Human population dispersal "Out of africa" Estimated from linkage disequilibrium and allele frequencies of snps. *Genome Res*, 21(6), 821-9.
- Mellars, P. (2006) A new radiocarbon revolution and the dispersal of modern humans in eurasia. *Nature*, 439(7079), 931-935.
- Mellars, P. (2006) Why did modern human populations disperse from africa ca. 60,000 years ago? A new model. *P Natl Acad Sci USA* 103(25), 9381-9386.
- Menozzi, P., Piazza, A. & Cavalli-Sforza, L. (1978) Synthetic maps of human gene frequencies in europeans. *Science*, 201(4358), 786-92.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) Equations of state calculations by fast computing machines. *J Chem Phys*, 21(6), 1087-1092.
- Nei, M. (1995) The origins of human populations: Genetic, linguistic, and archeological data, in: S. Brenner & K. Hanihara (Eds) *The origin and past of modern humans as viewed from DNA*. Singapore, World Scientific, 71-91.
- Nei, M. & Roychoudhury, A. K. (1993) Evolutionary relationships of human-populations on a global-scale. *Mol Biol Evol*, 10(5), 927-943.
- Nelson, S. M. (1993) *The archaeology of korea*, Cambridge, Cambridge University Press.

- Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity* (Edinb), 86, 641-7.
- Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. & Stoneking, M. (2001) Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet*, 29(1), 20-21.
- Oslisly, R. (2001) The history of human settlement in the middle Ohio valley. Weber, W., White, J.T., Veder, A., Naughton-Treves, L., eds. New Haven (Connecticut) and London: Yale University Press.
- Pakendorf, B. & Stoneking, M. (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet*, 6, 165-183.
- Partridge, T. C., Demenocal, P. B., Lorentz, S. A., Paiker, M. J. & Vogel, J. C. (1997) Orbital forcing of climate over South Africa: A 200,000-year rainfall record from the Pretoria saltpan. *Quaternary Science Reviews*, 16(10), 1125-1133.
- Patin, E., Laval, G., Barreiro, L. B., Salas, A., Semino, O., Santachiara-Benerecetti, S., Kidd, K. K., Kidd, J. R., Van Der Veen, L., Hombert, J. M., Gessain, A., Froment, A., Bahuchet, S., Heyer, E. & Quintana-Murci, L. (2009) Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *Plos Genet*, 5(4), e1000448.
- Pavlov, P., Svendsen, J. & Indrelid, S. I. (2001) Human presence in the European Arctic nearly 40,000 years ago. *Nature*, 413, 64-67.
- Phillipson, D. W. (1993) *African archaeology*, (Cambridge, Cambridge University Press).
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7-11.
- Pluzhnikov, A., Di Rienzo, A. & Hudson, R. R. (2002) Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics*, 161(3), 1209-1218.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. (1999) Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol Biol Evol*, 16(12), 1791-8.
- Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C. & Mehdi, S. Q. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet*, 70(5), 1107-24.
- Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A. S., Coppa, A., Ayub, Q.,

- Mohyuddin, A., Tyler-Smith, C., Mehdi, S. Q., Torroni, A. & Mcelreavey, K. (2004) Where west meets east: The complex mtDNA landscape of the southwest and central asian corridor. *Am J Hum Genet*, 74(5), 827-845.
- Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mougiamama-Daouda, P., Comas, D., Tzur, S., Balanovsky, O., Kidd, K. K. K., Kidd, J. R., Van Der Veen, L., Hombert, J. M., Gessain, A., Verdu, P., Froment, A., Bahuchet, S., Heyer, E., Dausset, J., Salas, A. & Behar, D. M. (2008) Maternal traces of deep common ancestry and asymmetric gene flow between pygmy hunter-gatherers and bantu-speaking farmers. *Proc. Natl. Acad. Sci. USA*, 105, 1596-1601.
- Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., Mcelreavey, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence of an early exit of homo sapiens sapiens from africa through eastern africa. *Nat Genet*, 23(4), 437-41.
- R.Development.Core.Team (2011) R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.
- Rambaut, A. & Drummond, A. J. (2007) Tracer, ver. 1.5.
- Ramos-Onsins, S. E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol*, 19(12), 2092-100.
- Ray, N., Currat, M. & Excoffier, L. (2003) Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol*, 20(1), 76-86.
- Rebernig, C. A., Weiss-Schneeweiss, H., Schneeweiss, G. M., Schonswetter, P., Obermayer, R., Villasenor, J. L. & Stuessy, T. F. (2010) Quaternary range dynamics and polyploid evolution in an arid brushland plant species (*melampodium cinereum*, asteraceae). *Mol Phylogenet Evol*, 54(2), 594-606.
- Renfrew, C. (1996) Language families and the spread of farming. In D.R. Harris (ed.) *The origins and spread of agriculture and pastoralism in Eurasia*. London: UCL Press. pp. 70-92.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. & Bandelt, H. J. (2000) Tracing

- European founder lineages in the near eastern mtDNA pool. *Am J Hum Genet*, 67(5), 1251-1276.
- Rodriguez, F., Oliver, J. L., Marin, A. & Medina, J. R. (1990) The general stochastic model of nucleotide substitution. *J Theor Biol*, 142(4), 485-501.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. & Feldman, M. W. (2002) Genetic structure of human populations. *Science*, 298(5602), 2381-5.
- Sauer, C. O. (1952) *Agricultural origins and dispersals*, (Cambridge, MA).
- Scally, A. & Durbin, R. (2012) Revising the human mutation rate: Implications for understanding human evolution. *Nat Rev Genet*, 13(10), 745-53.
- Ségurel, L., Austerlitz, F., Toupance, B., Gautier, M., Joanna, L., Kelley, J. L., Pasquet, P., Lonjou, C., Georges, M., Voisin, S., Cruaud, C., Coulloux, A., Hegay, T., Aldashev, A., Vitalis, R. & Heyer, E. (2013) Selection of protective variants for type 2 diabetes from the neolithic onward: A case study in central Asia. *Eur J Hum Genet*: online early. doi: 10.1038/ejhg.2012.295.
- Ségurel, L., Martinez-Cruz, B., Quintana-Murci, L., Balaresque, P., Georges, M., Hegay, T., Aldashev, A., Nasyrova, F., Jobling, M. A., Heyer, E. & Vitalis, R. (2008) Sex-specific genetic structure and social organization in central Asia: Insights from a multi-locus study. *PLoS Genet*, 4(9), e1000200.
- Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genet*, 20, 278 - 280.
- Shea, J. J. (2009) The impact of projectile weaponry on late Pleistocene hominin evolution. In: Hublin J-J, Richards MP, editors. *Evolution of Hominin Diets*. Dordrecht: Springer. pp. 189-199.
- Short, R. (1982) The biological basis for the contraceptive effects of breast feeding. *Int J Gynaecol Obstet.*, 25, 207-17.
- Sinsheimer, J. S., Lake, J. A. & Little, R. J. (1996) Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. *Biometrics*, 52(1), 193-210.
- Slatkin, M. & Hudson, R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129, 555-562.
- Sokal, R. R., Oden, N. L. & Wilson, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*, 351(6322), 143-145.

- Spiegelhalter, D. J., Best, N. G., Carlin, B. R. & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *J Roy Stat Soc B-Stat Methodo*, 64,583-616.
- Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001) Bayesian selection of continuous-time markov chain evolutionary models. *Mol Biol Evol*, 18(6), 1001-13.
- Tajima, F. (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595.
- The.1000.Genomes.Project.Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.
- The.International.Hapmap.Consortium. (2003) The international hapmap project. *Nature*, 426(6968), 789-96.
- Tremblay, M. & Vezina, H. (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am J Hum Genet*, 66,651-658.
- Valeggia, C. & Ellison, P. T. (2009) Interactions between metabolic and reproductive functions in the resumption of postpartum fecundity. *Am J Hum Biol* 21,559-566.
- Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., They, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J. M., Van Der Veen, L., Quintana-Murci, L., Bahuchet, S. & Heyer, E. (2009) Origins and genetic diversity of pygmy hunter-gatherers from western central africa. *Curr Biol*, 19(4), 312-8.
- Verdu, P., Becker, N. S., Froment, A., Georges, M., Grugni, V., Quintana-Murci, L., Hombert, J. M., Van Der Veen, L., Le Bomin, S., Bahuchet, S., Heyer, E. & Austerlitz, F. (2013) Sociocultural behavior, sex-biased admixture, and effective population sizes in central african pygmies and non-pygmies. *Mol Biol Evol*, Online early. doi: 10.1093/molbev/mss328.
- Voight, B. F., Adams, A. M., Frisse, L. A., Qian, Y. D., Hudson, R. R. & Di Rienzo, A. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Nat Acad SciUSA*, 102(51), 18508-18513.
- Watson, E., Forster, P., Richards, M. & Bandelt, H. J. (1997) Mitochondrial footprints of human expansions in africa. *Am J Hum Genet*, 61(3), 691-704.
- Weber, J. L. & Wong, C. (1993) Mutation of human short tandem repeats. *Hum Mol Genet*, 2(8), 1123-8.
- Weir, B. S. & Hill, W. G. (2002) Estimating f-statistics. *Annu Rev Genet*, 36(721-50).
- Wilson, A. C. & Cann, R. L. (1992) The recent african genesis of humans. *Sci Am*, 266(4), 68-73.

- Woerner, A. E., Cox, M. P. & Hammer, M. F. (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, 23(14), 1851-1853.
- Wood, F. (2002) *The silk road: Two thousand years in the heart of asia*. London : British Library.
- Wright, S. (1931) Evolution in mendelian populations. *Genetics*, 16,97-159.
- Wu, C. H. & Drummond, A. J. (2011) Joint inference of microsatellite mutation models, population history and genealogies using transdimensional markov chain monte carlo. *Genetics*, 188(1), 151-64.
- Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M. H. (2011) Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Syst Biol*, 60(2), 150-60.
- Yao, Y. G., Kong, Q. P., Bandelt, H. J., Kivisild, T. & Zhang, Y. P. (2002) Phylogeographic differentiation of mitochondrial DNA in han chinese. *Am J Hum Genet*, 70(3), 635-651.
- Zaharova, B., Andonova, S., Gilissen, A., Cassiman, J. J., Decorte, R. & Kremensky, I. (2001) Y-chromosomal str haplotypes in three major population groups in bulgaria. *Forensic Sci Int*, 124(2-3), 182-6.
- Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G. K., Herrera, R. J., Yong, K. K., Gresham, D., Tournev, I., Feldman, M. W. & Kalaydjieva, L. (2004) The effective mutation rate at y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet*, 74(1), 50-61.

Inférence des processus démographiques passés à partir de différents marqueurs génétiques pour des populations humaines aux modes de vie contrastés

Résumé

Reconstruire l'histoire démographique de notre espèce est un défi pour de nombreuses disciplines. Notamment, l'émergence de l'agriculture et de l'élevage au Néolithique est largement considérée par les archéologues et paléanthropologues comme le déclencheur des grandes expansions démographiques. A l'inverse, peu d'études de génétique des populations ont détecté des traces d'expansions néolithiques dans le polymorphisme génétique actuel, soulignant plutôt des expansions plus anciennes (Paléolithique moyen ou supérieur). Ici, nous avons inféré l'histoire démographique de populations d'Afrique et d'Eurasie aux modes de vie contrastés, à l'aide de plusieurs méthodes issues de la théorie de la coalescence appliquées à différents marqueurs génétiques. L'analyse de séquences autosomales et mitochondriales révèle une première expansion au Paléolithique, excepté chez les ancêtres des chasseurs-cueilleurs actuels en Afrique. Grâce aux microsatellites autosomaux, nous démontrons également une deuxième phase d'expansion plus récente, compatible avec la transition néolithique, chez les populations d'agriculteurs sédentaires mais pas chez les populations d'éleveurs nomades. Nous avons également montré que les processus de migration et d'isolation peuvent influencer dans une certaine mesure les inférences démographiques pour certaines populations. Enfin des analyses sur données simulées nous ont permis de tester les méthodes utilisées et ont confirmé le fait que, dans le cas de deux expansions successives au cours du temps, les données de séquences tendront à nous renseigner sur l'évènement le plus ancien tandis que les données microsatellites dévoileront l'évènement le plus récent.

Mots clés : génétique des populations, inférences démographiques, coalescence, modes de vie, transition Néolithique, expansions.

Abstract

Reconstructing the demographic History of human populations remains a strongly investigated issue in many disciplines. In particular, the transition from hunting and gathering to plant and animal domestication during the Neolithic period is widely assumed by paleoanthropologists and archeologists to have driven recent human population expansions. Conversely, although demographic changes leave footprints on genetic polymorphism, few population genetic studies have found traces of Neolithic expansions in the current repartition of human genetic diversity, pointing rather toward more ancient (*i.e.* Middle or Upper Paleolithic) expansions. Here, we inferred the demographic history of multiple African and Eurasian populations with contrasted life styles, using several coalescent-based methods applied to different types of genetic markers. The analyses on autosomal and mitochondrial sequences revealed a Paleolithic expansion event for most populations, except for the ancestors of contemporary African hunter-gatherers. Using autosomal microsatellites, we also inferred a more recent expansion event, likely concomitant with the Neolithic transition, in sedentary farmer but not in nomadic herder populations. We also found that, in some cases, isolation and migration patterns can have an impact on coalescent-based inferences. Finally, using simulated data, we confirmed the fact that, when two consecutive expansions occur, sequence data generally give information about the oldest one while microsatellite data can bring information about the most recent one.

Keyword: population genetics, demographic inferences, coalescent, life-style, Neolithic transition, expansions.