

L'ADN ancien suggère que les loups modernes trouvent leur origine dans une expansion de la Béringie au Pléistocène supérieur

Received: 5 October 2018 | Revised: 16 October 2019 | Accepted: 3 December 2019




DOI: 10.1111/mec.15329

Molecular Ecology, 2020, 29, 1596–1610

FROM THE COVER

MOLECULAR ECOLOGY WILEY

Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia

Liisa Loog^{1,2,3,4}  | Olaf Thalmann⁵ | Mikkel-Holger S. Sinding^{6,7,8}  | Verena J. Schuenemann^{9,10,11} | Angela Perri¹² | Mietje Germonpré¹³ | Herve Bocherens^{10,14} | Kelsey E. Witt¹⁵ | Jose A. Samaniego Castruita⁶ | Marcela S. Velasco⁶ | Inge K. C. Lundstrøm⁶ | Nathan Wales^{6,16} | Gontran Sonet¹⁷  | Laurent Frantz² | Hannes Schroeder⁶ | Jane Budd¹⁸ | Elodie-Laure Jimenez¹³ | Sergey Fedorov¹⁹ | Boris Gasparyan²⁰ | Andrew W. Kandel²¹ | Martina Lázníčková-Galetová^{22,23,24} | Hannes Napierala²⁵ | Hans-Peter Uerpmann⁹ | Pavel A. Nikolskiy^{26,27} | Elena Y. Pavlova^{27,28} | Vladimir V. Pitulko²⁷ | Karl-Heinz Herzig^{5,29} | Ripan S. Malhi³⁰ | Eske Willerslev^{2,31,32} | Anders J. Hansen^{8,31} | Keith Dobney^{33,34,35} | M. Thomas P. Gilbert^{6,36} | Johannes Krause^{9,37} | Greger Larson¹ | Anders Eriksson^{2,38,39}  | Andrea Manica²

Résumé

Le loup gris (*Canis lupus*) est l'un des rares grands carnivores terrestres à avoir conservé une large répartition géographique dans l'hémisphère nord tout au long du Pléistocène et de l'Holocène. Des études génétiques récentes ont suggéré que, malgré cette présence continue, des changements démographiques majeurs se sont produits dans les populations de loups entre le Pléistocène tardif et le début de l'Holocène, et que les loups actuels remontent à une seule population du Pléistocène tardif. L'origine géographique de cette population ancestrale et la façon dont elle s'est répandue restent inconnues. Nous avons utilisé un cadre de modélisation spatialement et temporellement explicite pour analyser un ensemble de données comprenant 90 génomes mitochondriaux de loups modernes et 45 anciens provenant de l'hémisphère nord et couvrant les 50 000 dernières années. Nos résultats suggèrent que les populations de loups contemporaines ont pour ancêtre une expansion à partir de la Béringie à la fin du dernier maximum glaciaire, et que ce processus a très probablement été déterminé par les fluctuations écologiques du Pléistocène tardif qui se sont produites dans l'hémisphère nord. Cette étude fournit des preuves génétiques anciennes directes que la migration à longue distance a joué un rôle important dans l'histoire de la population d'un grand carnivore, et permet de comprendre comment les loups ont survécu à la vague d'extinctions de la mégafaune à la fin de la dernière glaciation. De plus, étant donné que les loups gris du Pléistocène supérieur étaient la source probable de l'origine de tous les chiens modernes, l'histoire démographique décrite dans cette étude a des implications fondamentales pour comprendre l'origine géographique du chien.

1 | INTRODUCTION

L'époque du Pléistocène a abrité une grande diversité de prédateurs supérieurs, bien que la plupart se soient éteints pendant ou peu après le Dernier Maximum Glaciaire (LGM), il y a environ 21 000 ans (Barnosky, Koch, Feranec, Wing, & Shabel, 2004 ; Clark et al., 2012). Le loup gris (*Canis lupus*) était l'un des rares grands carnivores qui a survécu et maintenu une large gamme géographique tout au long de la période (Puzachenko & Markova, 2016), et les archives paléontologiques et archéologiques attestent de la présence continue de loups gris à travers l'hémisphère nord depuis au moins les 300 000 dernières années (Sotnikova & Rook, 2010) (examiné dans l'Annexe S1). Cette continuité géographique et temporelle à travers l'hémisphère nord contraste avec les analyses de génomes modernes complets, qui ont suggéré que tous les loups et chiens contemporains descendent d'une population ancestrale commune qui existait il y a seulement 20 000 ans (Fan et al., 2016 ; Freedman et al., 2014 ; Skoglund, Ersmark, Palkopoulou, & Dalén, 2015).

Ces analyses mettent en évidence un goulot d'étranglement suivi d'une radiation rapide à partir d'une population ancestrale autour ou juste après le LGM. L'origine géographique et la dynamique de cette radiation restent inconnues. La résolution de ces changements démographiques est nécessaire pour comprendre les circonstances écologiques qui ont permis aux loups de survivre aux extinctions de la mégafaune du Pléistocène supérieur. De plus, comme les chiens ont été domestiqués à partir de loups gris du Pléistocène supérieur (Larson et al., 2012), un aperçu détaillé de la démographie des loups au cours de cette période fournirait un contexte essentiel pour reconstruire l'histoire de la domestication des chiens.

Reconstruire les événements démographiques passés uniquement à partir de génomes modernes est un défi car de multiples histoires démographiques peuvent conduire à des modèles génétiques similaires dans les échantillons actuels (Groucutt et al., 2015). Les analyses qui incorporent des séquences d'ADN anciennes peuvent éliminer certaines de ces histoires alternatives en quantifiant les changements dans les différences génétiques des populations à travers le temps. Si les marqueurs nucléaires offrent une plus grande puissance par rapport à l'ADN mitochondrial (ADNmt), ce dernier est plus facilement récupérable et mieux préservé dans les échantillons anciens en raison de son nombre de copies plus élevé par rapport à l'ADN nucléaire, ce qui permet de générer des ensembles de données avec une plus grande couverture géographique et temporelle. En particulier, l'analyse d'échantillons datant d'avant, de pendant et d'après les événements démographiques d'intérêt augmente considérablement le pouvoir de déduire les histoires démographiques passées. En outre, le taux de mutation nucléaire chez les canidés est mal compris, ce qui conduit à de larges plages de dates pour les événements démographiques passés reconstruits à partir de panels de génomes entiers modernes (par exemple, Fan et al., 2016 ; Freedman et al., 2014). Le fait de disposer d'échantillons directement datés sur une large période nous permet d'estimer les taux de mutation avec une plus grande exactitude et précision par rapport aux méthodes alternatives (Drummond, Nicholls, Rodrigo, & Solomon, 2002 ; Rambaut, 2000 ; Rieux et al., 2014).

Les processus démographiques, tels que les expansions et les contractions d'aires de répartition, qui impliquent l'espace ainsi que le temps sont particulièrement difficiles à reconstruire car ils conduisent souvent à des schémas difficiles à interpréter intuitivement (Groucutt et al., 2015). Les hypothèses impliquant des processus spatiaux peuvent être testées formellement à l'aide de modèles de génétique des populations qui représentent explicitement les différents processus démographiques et leur effet sur la variation génétique à travers le temps et l'espace (Eriksson et al., 2012 ; Eriksson & Manica, 2012 ; Posth et al., 2016 ; Raghavan et al., 2015 ; Warmuth et al., 2012). L'intégration formelle du temps et de l'espace dans les cadres de la génétique des populations

permet d'analyser des ensembles de données éparées, un défi courant lorsqu'on traite de l'ADN ancien (Loog et al., 2017).

Ici, nous utilisons un cadre de génétique des populations spatialement explicite pour modéliser une gamme de différentes histoires démographiques de loups à travers l'hémisphère nord qui impliquent des combinaisons de goulets d'étranglement de la population, de turnover et de migrations à longue distance ainsi que de flux génétiques locaux. Afin d'estimer les paramètres du modèle et de tester formellement les hypothèses sur l'origine et la dynamique de population de l'expansion des loups gris au cours du LGM, nous avons rassemblé un ensemble de données substantiel (Figure 1 ; Tableau S1), couvrant les 50 000 dernières années et la largeur géographique de l'hémisphère Nord. Cet ensemble de données est constitué de 90 génomes mitochondriaux entiers de loups modernes et 45 anciens (dont 38 ont été nouvellement séquencés). Dans ce qui suit, nous présentons d'abord une analyse phylogénétique de nos séquences et une **calibration** des taux de mutation mitochondriaux des loups. Nous effectuons ensuite un test d'hypothèse formel en utilisant le Calcul Bayésien Approximatif (ABC) avec nos modèles spatio-temporels explicites. Nous concluons par une discussion sur la manière dont nos résultats sont liés aux études précédentes et sur les implications pour les recherches futures.

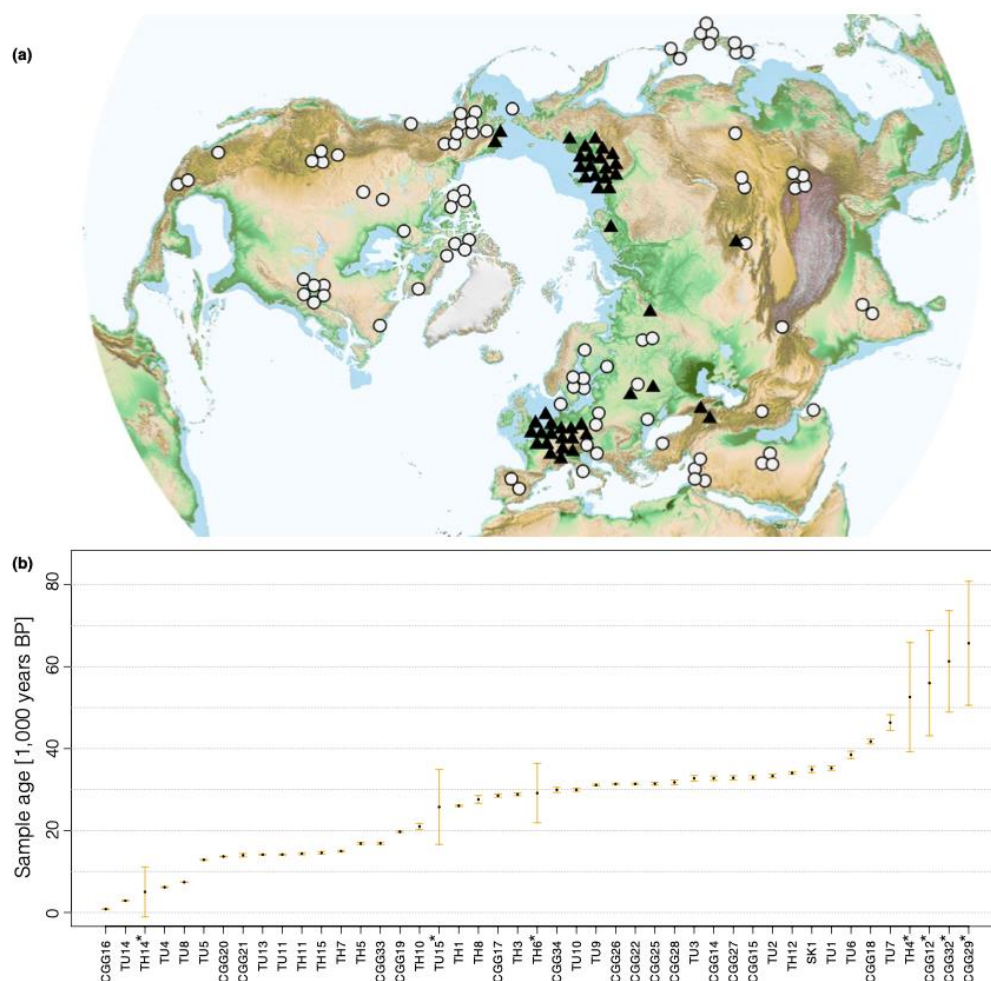


FIGURE 1. Distribution géographique des échantillons modernes (<500 ans, cercles) et anciens (>500 ans, triangles) (a) et distribution temporelle des échantillons anciens (b) utilisés dans les analyses. Les emplacements géographiques des échantillons ont été légèrement ajustés pour plus de clarté (voir le Tableau S1 pour les emplacements exacts des échantillons). *Échantillons datés par datation moléculaire

2 | MATÉRIAUX ET MÉTHODES

2.1 | Préparation des données

Nous avons séquencé les génomes mitochondriaux entiers de 40 échantillons de loups anciens. Les informations sur les échantillons, y compris les emplacements géographiques, les âges estimés et les informations sur le contexte archéologique pour les échantillons anciens, sont fournies dans le Tableau S1 et l'Annexe S1. Sur les 40 échantillons anciens, 24 ont été directement datés au radiocarbone pour cette étude et calibrés à l'aide de la courbe d'étalonnage IntCal13 (voir le Tableau S1 pour les dates de radiocarbone, les fourchettes d'âge calibrées et les numéros de référence du laboratoire de spectrométrie de masse par accélérateur [AMS]). L'extraction de l'ADN, le séquençage et le filtrage de la qualité, ainsi que les protocoles de cartographie utilisés sont décrits dans l'Annexe S2.

Nous avons inclus 16 génomes mitochondriaux anciens de loups précédemment publiés (Tableau S1 et Annexe S2). Pour obtenir un ensemble de données uniforme, nous avons retraité les lectures brutes des échantillons précédemment publiés en utilisant le même pipeline bioinformatique que pour les séquences nouvellement générées.

Nous avons soumis les séquences anciennes alignées à des critères de qualité stricts en termes de modèles de dommages et de données manquantes (Figures S3-S5). **Premièrement**, nous avons exclu toutes les séquences mitochondriales entières pour lesquelles il manquait plus d'un tiers du génome mitochondrial (à l'exclusion de la région de contrôle mitochondriale - voir ci-dessous) avec une couverture minimale de trois fois. **Ensuite**, nous avons exclu toutes les séquences mitochondriales entières anciennes qui contenaient plus de 0,1 % de singletons présentant des signes de désamination typiques de l'ADN ancien (singletons C à T ou A à G). Après le filtrage de la qualité, nous nous sommes retrouvés avec 32 séquences mitochondriales entières nouvellement séquencées et 13 séquences anciennes publiées (Tableau S1).

Nous avons également exclu les séquences provenant de spécimens archéologiques postérieurs à la fin du Pléistocène et identifiés comme étant des chiens (Tableau S1), car toute structure de population significative résultant d'un manque de flux génétique entre les chiens et les loups pourrait violer l'hypothèse d'une population unique de canidés s'accouplant de manière aléatoire. Certains des spécimens pléistocènes utilisés dans les analyses démographiques (TH5, TH12, TH14) ont été considérés comme présentant des caractéristiques communément trouvées chez les chiens modernes et ont donc été suggérés comme représentant des chiens paléolithiques (par exemple, Druzhkova et al., 2013 ; Germonpré, Lázničková-Galetová, Losey, Rääkkönen, & Sablin, 2015 ; Germonpré, Lázničková-Galetová, & Sablin, 2012 ; Germonpré et al., 2009 ; Sablin & Khlopachev, 2002). Ici, nous ne tenons pas compte de ces appels au statut en raison de la controverse qui les entoure (Crockford & Kuzmin, 2012 ; Drake, Coquerelle, & Colombeau, 2015 ; Morey, 2014 ; Perri, 2016), et parce que les premiers chiens auraient été génétiquement similaires aux populations locales de loups dont ils sont issus. Ce raisonnement est soutenu par la proximité de ces échantillons avec d'autres spécimens de loups décrits avec confiance comme des loups dans l'arbre phylogénétique (voir Figure S10).

Enfin, nous avons séquencé six échantillons de loups modernes et ajouté 66 séquences de loups modernes publiées provenant du NCBI, deux séquences de Freedman et al. (2014), 13 séquences de Sinding et al. (2018) et trois séquences de Gopalakrishnan et al. (2018) (Tableau S1). Les données de Sinding et al. (2018) et de Gopalakrishnan et al. (2018) ont été nouvellement assemblées en suivant les mêmes protocoles bioinformatiques que ceux utilisés pour les échantillons de loups modernes nouvellement séquencés (voir Annexe S2). Cela a donné lieu à un ensemble de données

final de 135 séquences complètes du génome mitochondrial du loup, dont 45 étaient anciennes et 90 étaient modernes. Nous avons utilisé l'outil d'alignement CLUSTALW (version 2.1) (Larkin et al., 2007) pour générer un alignement conjoint de tous les génomes.

Pour éviter l'effet potentiellement confondant des mutations récurrentes dans la région de contrôle mitochondriale (Excoffier & Yang, 1999) dans les calculs de différence par paire, nous avons supprimé cette région de toutes les analyses ultérieures. Nous avons ainsi obtenu un alignement de séquences de 15 466 pb de long, dont 1 301 sites (8,4 %) étaient variables. L'ensemble des données alignées est présenté dans l'Annexe S1.

2.2 Analyse phylogénétique

Nous avons calculé le nombre de différences par paires entre tous les échantillons (Figure S6) et généré un arbre de voisinage basé sur les différences par paires (Figure S7). Cet arbre montre un clade constitué d'échantillons provenant exclusivement de la région Tibétaine et du sous-continent Indien qui divergent profondément de tous les échantillons de loups anciens et modernes (voir également Aggarwal, Kivisild, Ramadevi, & Singh, 2007 ; Sharma, Maldonado, Jhala, & Fleischer, 2004). Une étude récente des données du génome entier a montré une histoire complexe des loups d'Eurasie du Sud (Fan et al., 2016) qui dépasse le cadre de notre étude. Alors que leur phylogénie de voisinage regroupait les loups d'Eurasie du Sud avec les loups d'Asie de l'Est et du Nord-Est (Fan et al., 2016 : Figure 3), ils se regroupent en dehors de tous les autres loups gris dans une analyse en composantes principales (Fan et al., 2016 : Figure 4), et montrent également une histoire démographique distincte dans une analyse de coalescence séquentielle markovienne par paires (PSMC) (Fan et al., 2016 : Figure 5). Comme notre étude ne possédait pas suffisamment d'échantillons de l'Himalaya et du sous-continent Indien pour démêler leur démographie complexe, nous avons exclu les échantillons de ces régions et nous nous sommes concentrés sur l'histoire des loups d'Eurasie du Nord et d'Amérique du Nord, pour lesquels nous avons une bonne couverture dans le temps et l'espace.

Nous avons utilisé PARTITIONFINDER (Lanfear, Calcott, Ho, & Guindon, 2012) et BEAST (version 1.8.0) (Drummond, Suchard, Xie, & Rambaut, 2012) pour construire un arbre mitochondrial de loup calibré (avec une horloge globale stricte, voir Annexe S1 pour plus de détails) à partir d'échantillons modernes et anciens directement datés, et pour estimer les taux de mutation pour quatre partitions différentes du génome mitochondrial du loup (voir Tableaux S3 et S4 pour les résultats).

Nous avons utilisé BEAST pour dater moléculairement sept séquences provenant d'échantillons qui n'étaient pas directement datés au radiocarbone (TH4, TH6, TH14, TU15) ou qui avaient été datés à une période au-delà de la limite de fiabilité des datations au radiocarbone (>48 000 ans) (CGG12, CGG29, CGG32). Nous avons estimé l'âge des échantillons en effectuant une analyse BEAST où le taux de mutation a été fixé aux estimations moyennes de l'analyse BEAST précédente et où tous les autres paramètres ont été définis comme décrit dans l'Annexe S1. Nous avons validé cette approche par le biais d'une analyse « leave-one-out » où nous avons séquentiellement retiré un échantillon directement daté et estimé sa date comme décrit ci-dessus. Nous constatons un ajustement étroit ($R^2 = 0,86$) entre les dates radiocarbone et moléculaires (Figure S9). Nous avons combiné les sept échantillons non datés avec les 110 échantillons anciens et modernes de la série précédente et utilisé une antériorité uniforme allant de 0 à 100 000 ans pour estimer l'âge des sept échantillons non datés (voir le Tableau S5 pour les résultats).

Enfin, afin d'estimer le temps de divergence mitochondriale entre les échantillons de loups d'Eurasie du Sud (Tibétains et Indiens) et le reste de nos échantillons de loups, nous avons effectué un passage supplémentaire de BEAST dans lequel nous avons inclus tous les loups gris modernes et anciens ($N = 129$) ainsi que cinq loups Tibétains et un loup Indien, et utilisé des paramètres identiques à ceux décrits ci-dessus. L'âge des échantillons anciens a été défini comme la moyenne de la distribution des dates radiocarbone calibrées (pour les échantillons radiocarbone) ou comme la moyenne de la distribution des âges provenant des analyses de bêtes (pour les échantillons à datation moléculaire).

2.3 Analyse de l'isolement par la distance

Nous avons effectué des analyses d'isolement par la distance (IBD) pour voir dans quelle mesure la variation génétique mitochondriale des loups montre une structure de population. À cette fin, nous avons régressé les distances géographiques par paires entre 84 échantillons de loups modernes (Tableau S1) par rapport à leurs distances génétiques (mitochondriales) par paires. La distance géographique entre toutes les paires d'échantillons a été calculée en kilomètres comme la distance orthodromique à partir de coordonnées géographiques, en utilisant la formule de Haversine (Sinnott, 1984) pour tenir compte de la courbure de la Terre comme suit :

$$G_{ij} = 2r \arcsin \left(\sqrt{\sin^2((\varphi_i - \varphi_j)/2) + \cos(\varphi_i) \cos(\varphi_j) \sin^2((\lambda_i - \lambda_j)/2)} \right) \quad (1)$$

où G est la distance en kilomètres entre les individus i et j ; φ_i et φ_j sont les coordonnées de latitude des individus i et j , respectivement ; λ_i et λ_j sont les coordonnées de longitude des individus i et j , respectivement ; et r est le rayon de la terre en kilomètres. Les distances génétiques par paire ont été calculées comme la proportion de sites qui diffèrent entre chaque paire de séquences (en excluant les bases manquantes), en utilisant la fonction `dist.dna` du paquet R `APE` (Paradis, Claude, & Strimmer, 2004).

2.4 | Définitions des dèmes géographiques

Nous avons représenté l'aire de répartition géographique du loup en **sept dèmes**, définis par des barrières géographiques majeures à travers le temps.

1. Le **dème Européen** est bordé par des eaux libres au nord et à l'ouest (respectivement l'océan Arctique et l'océan Atlantique), par les montagnes de l'Oural à l'Est, et par la Méditerranée, la mer Noire, la mer Caspienne et les montagnes du Caucase au sud.
2. Le **dème du Moyen-Orient** est constitué de la péninsule Arabique, de l'Anatolie et de la Mésopotamie et est bordé par la mer Noire, la mer Caspienne et la mer d'Aral au nord ; l'océan Indien au sud ; la chaîne de montagnes Tien Shen, le plateau Tibétain et l'Himalaya à l'Est ; et la mer Méditerranée à l'ouest.
3. Le **dème d'Eurasie centrale et septentrionale** est constitué du plateau Sibérien et est bordé par l'océan Arctique au nord, par les monts Oural à l'ouest, par le fleuve Lena et les chaînes de montagnes du nord-Est de la Sibérie (chaînes Chersky et Verkhoyansk) à l'Est, et par la chaîne de montagnes Tien Shen, le plateau Tibétain et le désert de Gobi au sud-Est.
4. Le **dème d'Eurasie orientale** est bordé par la chaîne de montagnes Tien Shen, le plateau Tibétain et le désert de Gobi à l'ouest ; par l'océan Pacifique à l'Est ; et par le fleuve Lena et les chaînes de montagnes du nord-Est de la Sibérie (chaînes Chersky et Verkhoyansk) au nord.
5. Le **dème de Béringie** enjambe le détroit de Béring, qui a été un pont terrestre pendant une grande partie du Pléistocène supérieur et de l'Holocène inférieur. Il est bordé à l'ouest par le fleuve Lena et les chaînes de montagnes du nord-Est de la Sibérie (chaînes Chersky et Verkhoyansk), et

au sud et à l'Est par l'étendue des nappes glaciaires de la Cordillère et des Laurentides au cours du LGM.

6. Le **dème de l'Amérique du Nord arctique** est constitué d'une zone du continent nord-Américain à l'Est des Rocheuses et à l'Ouest du Groenland, qui a été recouverte de glace pendant la dernière glaciation et qui est actuellement connue sous le nom d'archipel arctique Canadien.

7. Le **dème de l'Amérique du Nord** est constitué d'une zone du sous-continent nord-Américain jusqu'à et y compris la zone qui était recouverte par les inlandis de la Cordillère et des Laurentides pendant la dernière glaciation (Raghavan et al., 2015).

2.5 | Amovas

Pour quantifier la mesure dans laquelle nos dèmes géographiques capturent la variation génétique dans les données, nous avons effectué des analyses de variance moléculaire (amovas) (Excoffier, Smouse, & Quattro, 1992). Nous avons calculé la distance génétique par paire entre toutes les paires d'échantillons de loups modernes ($n = 84$, Tableau S1) comme décrit ci-dessus (Analyse de l'isolement par la distance) et avons réparti les échantillons, en fonction de leur emplacement géographique, en sept populations correspondant aux dèmes géographiques, comme décrit dans la Section 2.4, Définitions des dèmes géographiques. Nous avons utilisé ces dèmes comme niveau d'analyse et effectué 1 million de permutations en utilisant la fonction amova du paquet R pegas (version 0.10). Nous avons trouvé un fort soutien pour nos dèmes géographiques ($p < 10^{-6}$), 24,4 % de la variance au sein de l'ensemble de données étant expliquée par les dèmes choisis.

2.6 Scénarios démographiques

Nous avons testé un total de 16 combinaisons de scénarios démographiques, issus de quatre différents types de scénarios démographiques (illustrés dans la Figure 4a) :

1. **Modèle statique (l'hypothèse nulle)** - des dèmes voisins échangent des migrants, pas de changements démographiques.
2. **Scénarios de goulot d'étranglement** : les dèmes échangent des migrants comme dans le modèle statique, mais les populations ont une taille différente à différentes périodes. Nous considérons trois périodes de temps : 0-15 000 ans, 15 000-40 000 ans, et >40 000 ans.
3. **Scénarios d'expansion** - les dèmes échangent des migrants comme dans le modèle statique, mais un seul dème (qui a lui-même une population continue dans le temps) connaît une expansion qui commence entre 5 000 et 40 000 ans (à un rythme minimum de 1 000 ans par dème, de sorte que le monde entier pourrait être colonisé en 3 000 ans ou plus). Le dème d'origine a une population continue dans le temps, tandis que les populations indigènes de tous les autres dèmes sont remplacées, ce qui nous permet de tester formellement les hypothèses de continuité et de remplacement dans chacun des dèmes.
4. **Combinaisons** des scénarios 2 et 3.

2.7 | Cadre de coalescence génétique des populations

Nous avons mis en œuvre des modèles de génétique des populations coalescents pour les différents scénarios démographiques afin d'échantillonner les généalogies.

Dans le scénario statique, nous avons simulé des processus de coalescence locale (Kingman, 1982) à l'intérieur de chaque dème (à l'échelle du taux $1/K$ par paire de lignées, où K est le temps moyen jusqu'à l'ancêtre commun le plus récent (TMRCA) dans un dème et est donc proportionnel à la taille effective de la population). En outre, nous avons déplacé les lignées entre les dèmes selon un processus de Poisson avec un taux m par lignée. Pour correspondre à la distribution géographique

et temporelle des données, nous avons représenté chaque échantillon par une lignée du dème et de la date correspondants.

Le scénario du goulot d'étranglement a été mis en œuvre comme le scénario statique, mais avec des valeurs constantes par morceaux pour K en fonction du temps. Nous avons considéré trois périodes de temps, chacune avec sa propre valeur de K ($K1$, $K2$ et $K3$), motivées par les preuves archéologiques et génétiques des changements de population de loups décrits dans le texte principal. La première période s'étend du présent à l'Holocène précoce, il y a 0-15 000 ans. La deuxième période s'étend de l'Holocène précoce au Pléistocène récent et couvre le LGM, il y a 15 000 à 40 000 ans. Enfin, la troisième période couvre le Pléistocène tardif et au-delà, c'est-à-dire il y a 40 000 ans et plus.

Les scénarios d'expansion de la population étaient basés sur le modèle statique mais avec un modèle d'expansion de la population ajouté avec des effets fondateurs et le remplacement des populations locales (nous désignons les populations non encore remplacées par l'expansion comme « indigènes »). À partir du temps T , la population s'est étendue à partir du dème initial et a remplacé les populations voisines. La population du dème d'origine est représentée comme une population continue dans le temps. Après le début de l'expansion, celle-ci s'est poursuivie par étapes fixes de ΔT (dans le temps). À chaque étape, les populations colonisées remplacent les populations indigènes voisines (si un dème indigène borde plus d'un dème colonisé, ces dèmes contribuent de manière égale à la colonisation du dème indigène). Dans le cadre de la **coalescence** (qui simule des généalogies à rebours dans le temps), les événements de colonisation correspondent à des migrations forcées du dème **indigène** vers le dème **source**. S'il y avait plus d'un dème source, la source de chaque lignée était choisie au hasard avec une probabilité égale. Enfin, les effets fondateurs pendant la colonisation d'un dème indigène ont été mis en œuvre sous la forme d'un goulot d'étranglement local et instantané de la population dans le dème (après l'expansion), avec une sévérité mise à l'échelle pour donner une probabilité fixe x d'un événement de coalescence pour chaque paire de lignées dans le dème pendant le goulot d'étranglement (Eriksson & Mehlis, 2004) ($x = 1$ correspond à une perte complète de la diversité génétique dans le goulot d'étranglement, et $x = 0$ correspond à aucune réduction de la diversité génétique).

Enfin, le scénario combiné d'expansion de la population et de goulots d'étranglement a été mis en œuvre en rendant le paramètre K de la taille de la population dans le modèle d'expansion de la population dépendant du temps comme dans le modèle de goulots d'étranglement.

2.8 | Analyse du calcul bayésien approximatif...

2.9 | Tracés cartographiques

La carte de fond utilisée dans les Figures 1(a) et 3(a), montrant les régions climatiques sur les masses terrestres, a été générée en téléchargeant le fichier `color_etopo1_ice_low.jpg` d'ETOPO1 (Amante & Eakins, 2016), un modèle de relief global d'une minute d'arc de la surface de la Terre qui intègre la topographie terrestre et la bathymétrie des océans, et en masquant les régions où les profondeurs de la mer sont supérieures à 120 m.

3 | RÉSULTATS

3.1 | Structure de la population du loup gris dans l'hémisphère nord l'hémisphère nord

Motivés par la structure de la population observée dans les études du génome entier des loups modernes (Fan et al., 2016), nous avons testé le degré de la structure génétique spatiale parmi les échantillons de loups modernes de notre ensemble de données, et avons trouvé un fort modèle de DIB génétique à travers l'Eurasie ($\rho = 0,3$, $p < 0,0001$; voir la Figure S8). Ignorer cette structure de population (c'est-à-dire modéliser les loups comme une seule population panmictique) peut conduire à des résultats artefactuels (Mazet, Rodríguez, & Chikhi, 2015 ; Mazet, Rodríguez, Grusea, Boitard, & Chikhi, 2016). L'utilisation de **modèles spatialement structurés**, dans lesquels la migration est limitée aux populations adjacentes, est une approche commune pour traiter de telles situations (Eriksson et al., 2012 ; Eriksson & Manica, 2012 ; Kimura & Weiss, 1964 ; Wegmann et al., 2010).

Pour capturer la structure géographique observée dans notre ensemble de données, nous avons divisé l'hémisphère nord en sept régions, à peu près similaires en termes de superficie (Figure 3a). Les limites de ces régions sont définies par des caractéristiques géographiques, notamment des chaînes de montagnes, des mers et des déserts (voir Matériaux et Méthodes), qui sont susceptibles de réduire le flux génétique (Geffen, Anderson, & Wayne, 2004 ; Lucchini, Galov, & Randi, 2004) et de fournir un équilibre optimal entre la résolution et la puissance étant donné la distribution des échantillons disponibles pour les analyses. Pour quantifier la façon dont ce schéma représente la structure de la population de loups modernes, nous avons utilisé une amova pour séparer la variance génétique dans et entre les régions. Nos régions capturent 24,4 % de la variation génétique parmi nos échantillons modernes (amova, $p < 0,001$). Ce chiffre est considérablement plus élevé que les ~10 % de variance dérivant de la simple DIB, et soutient l'hypothèse selon laquelle les caractéristiques géographiques (grands fleuves, déserts et chaînes de montagnes) définissent la structure de la population de loups contemporains dans l'hémisphère nord et constituent donc des obstacles au flux génétique (mais où la force de ces obstacles peut varier).

3.2 Analyse phylogénétique bayésienne

Toutes les séquences anciennes incluses dans l'étude ont été soumises à des critères de qualité rigoureux en ce qui concerne la couverture et les modèles de dommages. Sur les 45 échantillons anciens, 38 avaient des dates radiocarbone directes bien résolues. Nous avons joint ces séquences anciennes à 90 séquences mitogéniques modernes et utilisé BEAST (Drummond et al., 2012) pour estimer un taux de mutation mitochondrial du loup. **En appliquant le taux de mutation déduit, nous avons pu dater moléculairement les sept séquences anciennes restantes** (Matériaux et méthodes). Nous avons validé cette approche par une analyse de type « leave-one-out » (Matériaux et Méthodes) en utilisant toutes les séquences anciennes directement datées et nous avons constaté un ajustement très étroit ($R^2 = 0,86$) entre le radiocarbone et les dates moléculaires estimées et aucun biais systématique dans nos dates estimées par voie moléculaire (Figure S9), ce qui justifie l'inclusion de ces séquences et des dates inférées dans les analyses spatialement explicites.

Notre analyse phylogénétique bayésienne suggère que l'ACRM de toutes les séquences mitochondriales de loups d'Eurasie du Nord et d'Amérique datent d'environ 40 000 ans, tandis que l'ACRM des séquences anciennes et modernes combinées date d'environ 90 000 ans (intervalle de densité postérieure la plus élevée [DHP] à 95 % : 82 000-99 000 ans) (Figure 2a, voir les Figures S11 et S12 pour les valeurs de support de nœud et les intervalles de crédibilité). Un clade divergent à la racine de cet arbre se compose exclusivement d'échantillons anciens d'Europe et du Moyen-Orient qui n'a pas contribué à la diversité mitochondriale actuelle dans nos données (voir également Thalmann et al., 2013).

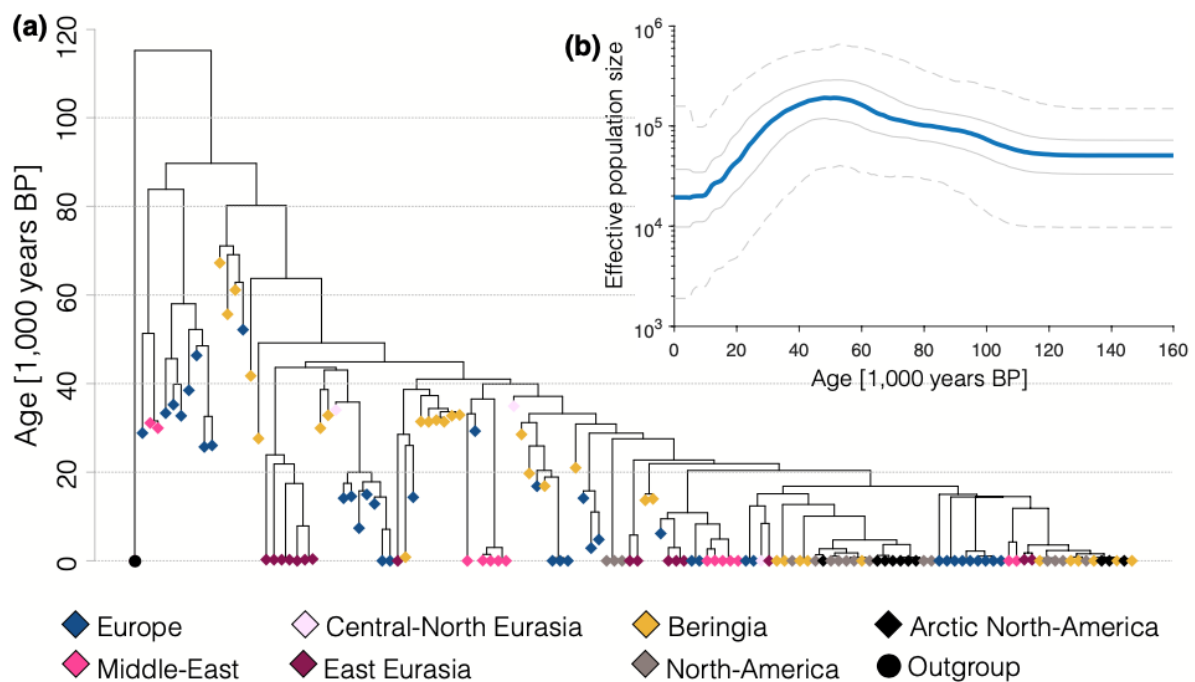


FIGURE 2. (a) Arbre BEAST calibré de tous les échantillons utilisés dans les analyses spatiales (diamants), colorés par région géographique. Le cercle représente un outgroup (loup Indien moderne, non utilisé dans les analyses). (b) La taille effective de la population à travers le temps à partir de l'analyse BEAST (graphique de la ligne d'horizon bayésienne). La ligne bleue continue représente l'estimation médiane et les lignes grises représentent l'intervalle interquartile (lignes soldées) et les intervalles à 95% (lignes pointillées)

Le reste de l'arbre est constitué d'un clade monophylétique composé d'échantillons anciens et modernes provenant de l'ensemble de l'hémisphère nord, qui présente un schéma de bifurcations rapides des lignées génétiques centrées sur 25 000 ans. Pour quantifier davantage ce schéma temporel, nous avons utilisé une analyse bayésienne de la ligne d'horizon (Figure 2b) qui montre une taille de population génétique effective relativement petite et stable entre ~20 000 ans et le présent et une diminution de la taille de population effective entre ~40 000 et 20 000 ans. Ce schéma est cohérent avec le scénario suggéré dans les études sur le génome entier (par exemple Fan et al., 2016 ; Freedman et al., 2014) où les loups avaient une population stable (et probablement structurée géographiquement) à travers l'hémisphère nord jusqu'à un moment situé entre 20 000 et 30 000 ans, lorsque la population a connu un goulot d'étranglement qui a sévèrement réduit la variation génétique, suivi d'une expansion rapide de la population.

Les échantillons à la base de ce clade proviennent principalement de la Beringie, ce qui indique une expansion possible à partir du nord-Est de l'Eurasie ou des Amériques. Cependant, étant donné la distribution temporelle et géographique inégale de nos échantillons, et la stochasticité d'un seul marqueur génétique (Nielsen & Beaumont, 2009), il est important de tester explicitement la mesure dans laquelle ce modèle peut se produire par hasard dans d'autres scénarios démographiques plausibles.

3.3 | Reconstruction spatio-temporelle de la démographie passée du loup gris

Après avoir établi la relation phylogénétique entre nos échantillons et la structure de la population dans l'hémisphère nord, nous avons testé la capacité de différents scénarios démographiques

explicités à expliquer le modèle phylogénétique observé, tout en tenant compte de la localisation géographique et de l'âge de chaque échantillon. À cette fin, nous avons représenté chacune des régions de la Figure 3(a) comme une population dans un réseau de populations reliées par des flux génétiques (Figure 3b). Nous avons utilisé le cadre de la génétique des populations coalescente pour modéliser l'évolution génétique dans ce réseau, dans lequel chaque dème constitue une population se mélangeant librement et s'accouplant de façon aléatoire. La taille effective de la population des dèmes, ainsi que le mouvement des individus entre les dèmes, sont contrôlés par des paramètres couvrant des valeurs qui représentent des histoires démographiques différentes.

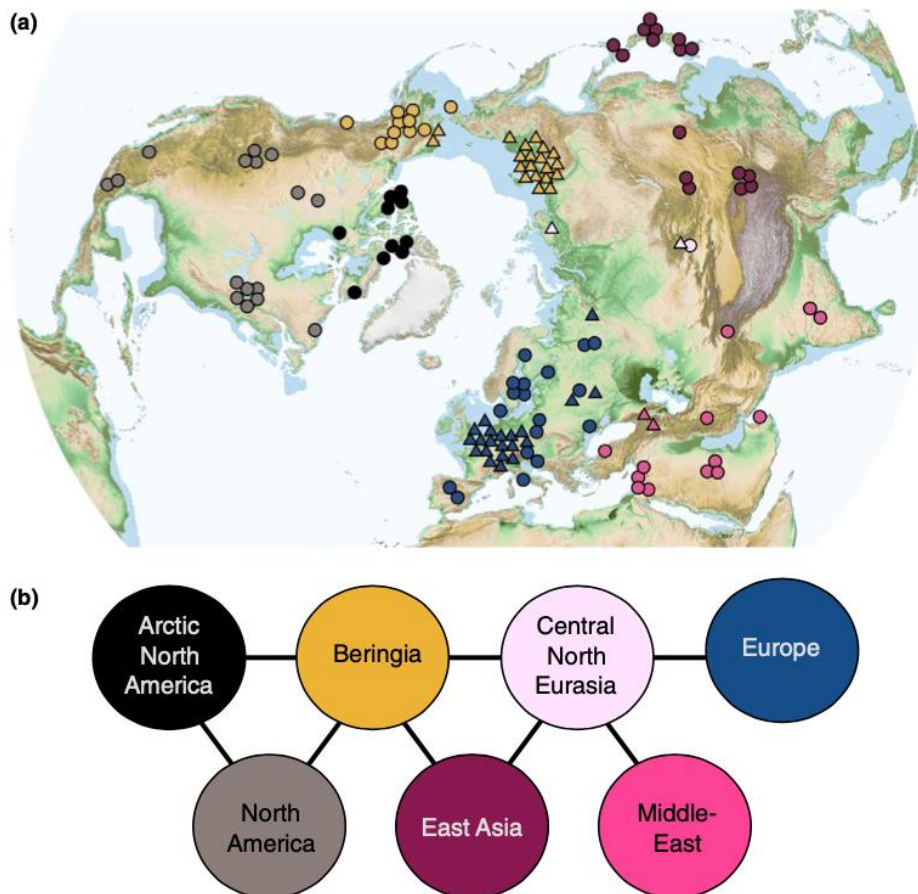


FIGURE 3 (a) Emplacements des échantillons et régions géographiques, les limites étant indiquées par des lignes pointillées. Le bleu foncé indique les niveaux de la mer assez peu profonds pour (a) Emplacements des échantillons et régions géographiques, les limites étant indiquées par des lignes pointillées. Le bleu foncé indique les niveaux de la mer suffisamment peu profonds pour être terrestres au cours du dernier maximum glaciaire (profondeur de la mer < 120 m)

En utilisant ce cadre, nous avons considéré un large éventail de scénarios démographiques explicites différents (illustrés dans la Figure 3a, voir Matériaux et Méthodes pour les détails de la mise en œuvre dans le cadre de coalescence). Le premier scénario consistait en une taille de population constante et un mouvement uniforme entre dèmes voisins. **Cela nous a permis de tester l'hypothèse nulle selon laquelle la dérive au sein d'une population structurée peut à elle seule expliquer tous les schémas observés dans l'arbre mitochondrial.** Nous avons ensuite envisagé deux autres processus démographiques susceptibles d'expliquer les schémas observés : (a) une séquence temporelle de deux changements de taille de la population affectant simultanément tous les dèmes (permettant ainsi un goulot d'étranglement) ; et (b) une expansion de l'un des sept dèmes. Dans les scénarios d'expansion, le dème d'origine avait une population continue dans le temps, tandis que dans les autres dèmes, la population indigène était remplacée séquentiellement par la population en

expansion. Le scénario 2 a été répété pour les sept origines d'expansion possibles, ce qui nous a permis de tester les hypothèses de continuité et de remplacement dans chacun des sept dèmes. Nous avons considéré chaque événement démographique isolément ainsi que leur effet combiné (ce qui a donné un total de 16 scénarios) et nous avons utilisé ABC pour calculer la probabilité de chaque scénario et estimer les valeurs des paramètres (voir Matériaux et Méthodes pour plus de détails).

Le scénario nul et le scénario de changement de la taille de la population uniquement dans tous les dèmes ont tous deux été fortement rejetés ($BF \leq 0,1$, Figure 4b ; Tableau S6), illustrant la puissance de la combinaison d'un grand ensemble de données d'échantillons anciens avec la modélisation statistique. Les scénarios qui combinaient une expansion et un remplacement avec un changement de la taille de la population (goulot d'étranglement) étaient mieux soutenus que les scénarios correspondants (c'est-à-dire avec la même origine d'expansion) avec une taille de population constante (Figure 4b).

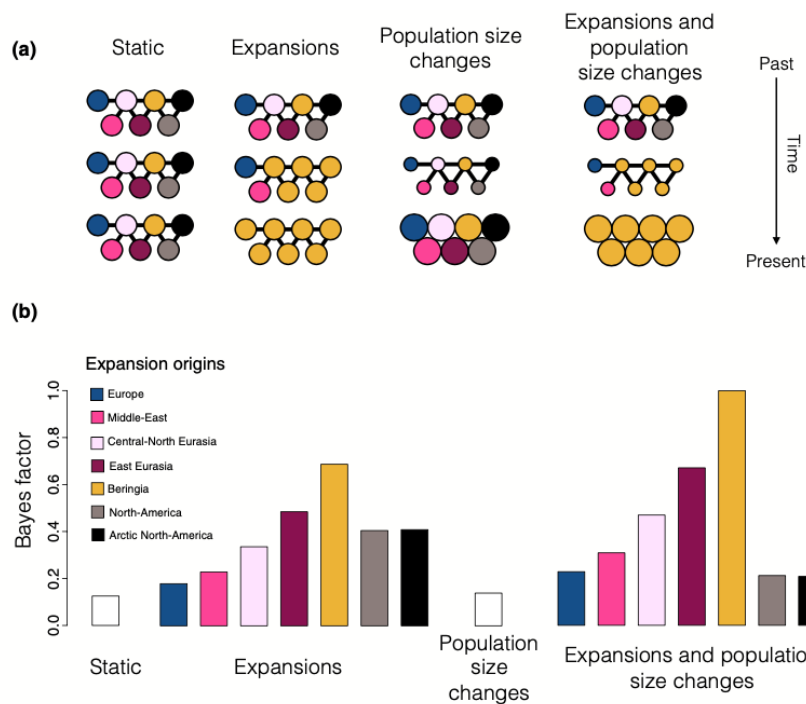


FIGURE 4. Analyse explicite dans l'espace et dans le temps. (a) Illustration des différents scénarios, avec des cercles représentant un dème chacun pour les sept régions géographiques différentes (voir panneau b pour la légende des couleurs et le texte pour la description complète des scénarios). Les lignes pleines représentent la connectivité de la population. Le scénario statique (à l'extrême gauche) montre des populations stables dans le temps. Le scénario d'expansion (au milieu à gauche) montre comment un dème (ici jaune) s'étend et remplace séquentiellement les populations de tous les autres dèmes (de haut en bas). Le scénario de changement de la taille de la population (au milieu à droite) illustre comment la taille de la population dans les dèmes peut changer au fil du temps (une grande ou une petite taille de la population est représentée respectivement par un grand ou un petit cercle). Nous montrons également un scénario combiné (à l'extrême droite) d'expansion et de changement de la taille de la population. (b) Probabilité de chaque scénario démographique par rapport au scénario le plus probable, représentée par des facteurs de Bayes, estimés à l'aide d'analyses de calcul bayésien approximatif (voir le texte pour plus de détails). Pour les scénarios d'expansion (y compris l'expansion et les changements de taille de la population combinés), nous attribuons un code couleur à chaque barre en fonction de l'origine de l'expansion (voir la légende des couleurs)

Le scénario le mieux étayé (Figure 5) était caractérisé par la combinaison d'une expansion rapide des loups hors du dème Béringien il y a environ 25 000 ans (intervalle de confiance [IC] à 95 % : 33 000-14 000 ans) avec un goulot d'étranglement de la population entre 15 000 et 40 000 ans, et un flux génétique limité entre les dèmes voisins (voir le Tableau S7 et la Figure S13 pour les distributions postérieures de tous les paramètres du modèle). Nous avons également trouvé un soutien relativement fort pour un scénario qui décrit une expansion du loup hors du dème d'Eurasie

de l'Est (BF 0,7) avec des paramètres presque identiques au scénario le mieux soutenu (Tableau S8 et Figure S14). Cela peut s'expliquer par la proximité géographique des dèmes d'Eurasie orientale et de Béringie et par la similarité génétique des loups de ces régions.

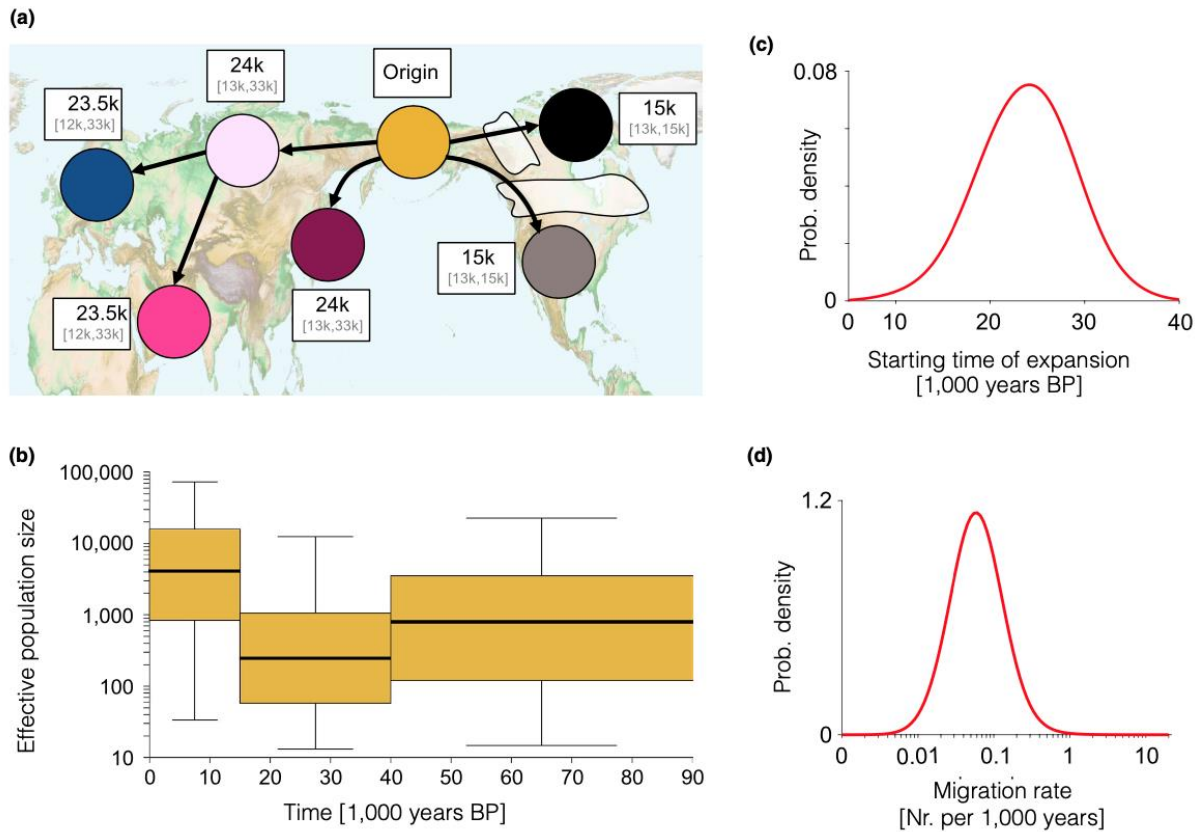


FIGURE 5. Le scénario inféré de la démographie du loup à partir de l'analyse bayésienne utilisant notre modèle spatialement et temporellement explicite (voir Figure 4 et le texte principal). (a) Représentation géographique du scénario d'expansion (hors de la Béringie) avec la médiane et l'intervalle de confiance (IC) à 95 % pour la date de remplacement de la population dans chaque dème, indiqués dans les cases blanches à côté de chaque dème. (b) Taille effective de la population (ligne épaisse, les cases et les moustaches indiquent la médiane, l'intervalle interquartile et l'IC à 95%, respectivement, pour chaque période). (c) Distribution postérieure du taux de migration et (d) moment du début de l'expansion

4 | DISCUSSION

4.1 | Origine géographique de la population ancestrale de loups population

Des études récentes sur le génome entier (Fan et al., 2016 ; Freedman et al., 2014 ; Skoglund et al., 2015) ont révélé que les loups gris modernes (*Canis lupus*) à travers l'Eurasie descendent d'une seule population source. Les résultats de nos analyses combinant des échantillons de loups gris anciens et modernes (Figure 1) avec un cadre de modélisation spatialement et temporellement explicite (Figure 4) suggèrent que ce processus a commencé il y a ~25 000 (IC 95 % : 33 000-14 000) ans, lorsqu'une population de loups de Béringie (ou d'une région d'Asie du Nord-Est géographiquement proche) s'est étendue vers l'extérieur et a remplacé les populations de loups indigènes du Pléistocène à travers l'Eurasie (Figure 5). Ce scénario fournit également un mécanisme expliquant la topologie en étoile des populations de loups modernes observée dans les études du génome entier (Fan et al., 2016 ; Freedman et al., 2014 ; Skoglund et al., 2015) : l'expansion a été fractionnée par des barrières géographiques qui ont restreint le flux génétique ultérieur entre les différentes branches de la population en expansion, ce qui a conduit à la divergence entre les différentes sous-populations observées chez les loups gris contemporains.

Sur le continent Américain, l'expansion Béringienne a été retardée en raison de la présence de calottes glaciaires s'étendant du Groenland au nord de l'océan Pacifique (Figure 5) (Raghavan et al., 2015). Une étude de Koblmüller et al. (2016) suggère que les populations de loups qui existaient au sud de ces inlandsis ont été remplacées par des loups Eurasiens traversant le pont terrestre Béringien. Nos données et analyses soutiennent le remplacement des loups d'Amérique du Nord (suite au retraitement des calottes glaciaires il y a environ 16 000 ans), et notre échantillonnage d'ADN ancien plus étendu, combiné à une modélisation spatialement explicite, nous a permis de réduire l'origine géographique de cette expansion à une zone située entre le fleuve Lena en Russie et le fleuve Mackenzie au Canada, également connue sous le nom de Béringie (Hopkins, Matthews, & Schweger, 1982). Cependant, en raison du manque d'échantillons de loups du Pléistocène antérieurs au retrait des couches de glace dans la région, nous ne sommes actuellement pas en mesure de résoudre l'histoire détaillée des loups d'Amérique du Nord. Par exemple, nous ne pouvons pas rejeter un scénario alternatif selon lequel les loups nord-Américains contemporains sont les descendants d'une population de loups du Pléistocène qui était génétiquement très similaire à la population Béringienne mais qui existait au sud des nappes glaciaires.

Ainsi, malgré un enregistrement fossile continu jusqu'au Pléistocène supérieur, les loups ont connu une histoire démographique complexe impliquant des goulets d'étranglement et des remplacements de population (Figure 5). Notre analyse suggère que la migration à longue distance a joué un rôle important dans la survie des loups lors de la vague d'extinctions de la mégafaune à la fin de la dernière glaciation. Ces résultats permettront à de futures études d'examiner les facteurs climatiques et écologiques locaux spécifiques qui ont permis à la population de loups Béringienne de survivre et de s'étendre dans l'hémisphère nord. De plus, comme les reconstitutions de cette étude ne sont basées que sur un marqueur génétique transmis par la mère, notre modèle n'a donc pu traiter qu'un ensemble de scénarios démographiques simplifiés (continuité partout, ou continuité dans un endroit suivi d'une expansion de remplacement à partir de celui-ci). Lorsque les données sur le génome entier seront disponibles, il sera probablement possible de détecter les contributions des refuges potentiels à l'échelle locale.

4.2 | Implications pour l'évolution de la morphologie du loup gris

Les analyses morphologiques de spécimens de loups ont mis en évidence des différences entre les loups du Pléistocène supérieur et ceux de l'Holocène : Les spécimens du Pléistocène tardif ont été décrits comme craniodentalement plus robustes que les loups gris actuels, ainsi que comme ayant des adaptations spécialisées pour le traitement des carcasses et des os (Baryshnikov, Mol, & Tikhonov, 2009 ; Kuzmina & Sablin, 1993 ; Leonard et al., 2007) associées à la chasse et à la fouille de la mégafaune (Fox-Dobbs, Leonard, & Koch, 2008 ; Germonpré et al., 2017). Les archives archéologiques de l'Holocène précoce n'ont livré qu'un seul échantillon présentant le morphotype du loup pléistocène (en Alaska) (Leonard et al., 2007), ce qui suggère que cet écomorphe robuste avait largement disparu de l'hémisphère Nord lors de la transition Pléistocène-Holocène. Ce changement dans la morphologie du loup coïncide avec un changement dans la composition isotopique du loup (Bocherens, 2015), et la disparition des herbivores mégafauniques et d'autres grands prédateurs tels que les hyènes des cavernes et les lions des cavernes, suggérant un possible changement dans la niche écologique des loups.

Jusqu'à présent, il n'a pas été possible de déterminer si les changements morphologiques résultaient d'un remplacement de population (renouvellement génétique), d'une réponse plastique à un changement de régime alimentaire, ou des deux. Nos résultats suggèrent que la transition

Pléistocène-Holocène s'est accompagnée d'un renouvellement génétique dans la plupart des populations de loups de l'hémisphère Nord, la plupart des populations de loups indigènes ayant subi un remplacement à grande échelle entraînant la perte de toutes les lignées génétiques indigènes du Pléistocène (Figure 5). Des dynamiques de population similaires de discontinuité et de remplacement par des congénères ont été observées chez plusieurs autres grands mammifères du Pléistocène en Europe, notamment les ours des cavernes, les mammoths laineux (Palkopoulou et al., 2013 ; Stuart, Kosintsev, Higham, & Lister, 2004), les cerfs géants (Stuart et al., 2004) et même les humains (Fu et al., 2016 ; Posth et al., 2016).

L'exception géographique à ce schéma de remplacement généralisé est la Béringie, où nous déduisons une continuité démographique entre les populations de loups du Pléistocène supérieur et de l'Holocène (Figure 5). Ce résultat est en contradiction avec une suggestion précédente de rotation génétique en Béringie (Leonard et al., 2007), probablement en raison de différences dans la quantité de données disponibles et la méthodologie analytique utilisée. Leonard et al. (2007) ont utilisé un segment court (427 bases) de la région de contrôle mitochondriale et ont employé une approche phylogéographique descriptive, alors que nos conclusions sont basées sur un ensemble de données élargi en termes de longueur de séquence, de nombre d'échantillons et d'étendue géographique et temporelle (Figure 1) et sur des tests d'hypothèse formels dans un cadre bayésien (Figures 4 et 5).

Par conséquent, le changement morphologique et alimentaire observé chez les loups de Béringie entre le Pléistocène supérieur et l'Holocène (Leonard et al., 2007) ne peut pas être expliqué par le renouvellement de la population, mais nécessite plutôt une **explication alternative** telle que l'adaptation ou les réponses plastiques aux changements environnementaux et écologiques substantiels qui ont eu lieu durant cette période. En effet, les loups gris sont une espèce hautement adaptable. Des études sur les loups gris modernes ont montré que les différences d'habitat, notamment les précipitations, la température, la végétation et la spécialisation des proies, peuvent fortement affecter leur morphologie craniodentaire (Flower & Schreve, 2014 ; Geffen et al., 2004 ; Leonard, 2015 ; O'Keefe, Meachen, Fet, & Brannick, 2013 ; Pilot et al., 2006).

4.3 | Implications pour l'étude de la domestication du loup

Enfin, l'histoire démographique complexe des loups gris d'Eurasie rapportée ici (Figure 5) a également des implications significatives pour l'identification de l'origine géographique de la domestication des loups et de la propagation ultérieure des chiens. Par exemple, notre compréhension limitée de la structure sous-jacente de la population de loups peut expliquer pourquoi les études précédentes ont produit des scénarios géographiques et temporels contradictoires. De nombreuses études antérieures se sont concentrées sur les modèles de variation génétique des chiens domestiques modernes, mais n'ont pas pris en compte la variation génétique potentielle présente dans les populations de loups du Pléistocène supérieur, supposant ainsi implicitement une source de population de loups homogène. En conséquence, la domestication des chiens et les déplacements ultérieurs des chiens par l'homme ont été les seuls processus considérés comme ayant affecté les modèles génétiques observés dans les populations de chiens. Cependant, la domestication et le mélange avec une population de loups structurée auront des conséquences sur les modèles de variation génétique au sein des chiens. À la lumière de l'histoire démographique complexe des loups (et de la structure génétique de la population qui en résulte) reconstituée par notre analyse, plusieurs des schémas géographiques de distribution des haplotypes observés dans des études antérieures, y compris les différences de niveaux de diversité trouvés au

sein des populations locales de chiens (Wang et al, 2016), et la profonde scission phylogénétique entre les chiens d'Eurasie orientale et occidentale (Frantz et al., 2016), pourraient avoir résulté du mélange connu entre les chiens domestiques et les loups gris (Fan et al., 2016 ; Freedman et al., 2014 ; Godinho et al., 2011 ; Verardi, Lucchini, & Randi, 2006). Les analyses futures devraient donc inclure explicitement l'histoire démographique des loups et démontrer que les modèles de variation observés au sein des chiens ne correspondent pas aux attentes qui prennent en compte le mélange avec des populations de loups géographiquement structurées.