

SCIENTIFIC REPORTS



OPEN

Close inbreeding and low genetic diversity in Inner Asian human populations despite geographical exogamy

Nina Marchi¹ , Philippe Menecier¹, Myriam Georges^{1,4}, Sophie Lafosse¹, Tatyana Hegay², Choduraa Dorzhu³, Boris Chichlo¹, Laure Ségurel¹ & Evelyne Heyer¹

When closely related individuals mate, they produce inbred offspring, which often have lower fitness than outbred ones. Geographical exogamy, by favouring matings between distant individuals, is thought to be an inbreeding avoidance mechanism; however, no data has clearly tested this prediction. Here, we took advantage of the diversity of matrimonial systems in humans to explore the impact of geographical exogamy on genetic diversity and inbreeding. We collected ethno-demographic data for 1,344 individuals in 16 populations from two Inner Asian cultural groups with contrasting dispersal behaviours (Turko-Mongols and Indo-Iranians) and genotyped genome-wide single nucleotide polymorphisms in 503 individuals. We estimated the population exogamy rate and confirmed the expected dispersal differences: Turko-Mongols are geographically more exogamous than Indo-Iranians. Unexpectedly, across populations, exogamy patterns correlated neither with the proportion of inbred individuals nor with their genetic diversity. Even more surprisingly, among Turko-Mongols, descendants from exogamous couples were significantly more inbred than descendants from endogamous couples, except for large distances (>40 km). Overall, 37% of the descendants from exogamous couples were closely inbred. This suggests that in Inner Asia, geographical exogamy is neither efficient in increasing genetic diversity nor in avoiding inbreeding, which might be due to kinship endogamy despite the occurrence of dispersal.

At least 10% of humans descend from parents that are related as second cousins or closer, based on registers as well as civil and medical surveys¹. Such consanguineous mating events can occur for different reasons: it can be because of socio-cultural habits, as emblematically described for European royal families² (mating choice inbreeding), or by chance, for example if there are too many related individuals in the population, as found in isolated groups^{3,4} (drift inbreeding). At the individual scale, the offspring of such close relatives are prone to carry a genetic burden⁵. Indeed, their genomes contain multiple chromosomal segments that are identical by descent⁶ (*i.e.*, that are inherited from a common ancestor). These segments, called runs of homozygosity (ROHs), consist of a succession of homozygous sites⁷ that subsequently increase the risk for deleterious recessive mutations to be phenotypically expressed⁸ and reduce the fitness at loci under heterozygote advantage⁹. This reduction in fitness associated with inbreeding can result in reduced fertility^{10,11}, viability^{12,13} and a wide range of other general disorders¹⁴. At the population level, inbreeding decreases the effective population size and thus the population genetic variability¹⁵. A lack of genetic diversity is often associated with a decrease in the adaptive response of a population¹⁶ and can even lead to its extinction^{17,18}. Moreover, in small populations undergoing strong genetic drift, inbreeding can increase the frequency (or even lead to the fixation) of mildly deleterious mutations^{19,20}. However,

¹Eco-anthropologie et Ethnobiologie, UMR 7206 CNRS, MNHN, Univ Paris Diderot, Sorbonne Paris Cité, Sorbonne Universités, 75016, Paris, France. ²Republican Scientific Center of Immunology, Ministry of Public Health, Tashkent, 100060, Uzbekistan. ³Department of biology and ecology, Tuvan State University, Kyzyl, 667000, Russia. ⁴Present address: LM2E-UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes, Institut Universitaire Européen de la Mer, Technopôle Brest-Iroise, Plouzane, 29280, France. Laure Ségurel and Evelyne Heyer jointly supervised this work. Correspondence and requests for materials should be addressed to N.M. (email: nina.marchi@mnhn.fr) or E.H. (email: heyer@mnhn.fr)

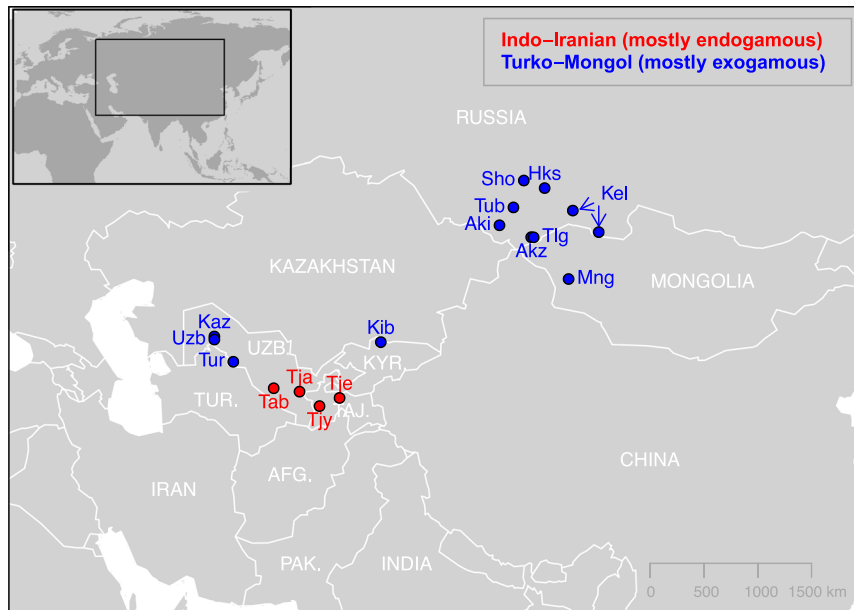


Figure 1. Geographical locations of the 16 Inner Asian populations sampled for this study. On this map generated from *maps*⁸⁰ and *mapdata*^{81 R⁸²} packages, the populations are coloured based on their linguistic affiliation, which also correlates with their matrimonial system. Note that the Kel population is composed of Northern Asian Kyrgyz from two different locations, and that Kaz and Uzb, as well as Akz and Tlg, were sampled at the same location, respectively. UZB.: Uzbekistan, KYR.: Kyrgyzstan, TUR.: Turkmenistan, TAJ.: Tajikistan, AFG.: Afghanistan, PAK.: Pakistan. See Supplementary Table 1B for population codes.

inbreeding can also enable the purge of severely deleterious recessive mutations^{21,22} and eventually improve the population fitness²³.

How do species cope with such inbreeding issues? Several pre- and post-copulatory behaviours are thought to reduce inbreeding in the animal kingdom^{11,24}. Most pre-copulatory behaviours involve kin recognition, which evolved both for avoidance and for cooperation^{25,26}. These can rely on the learning of familiarity, which is called Westermarck's effect in humans^{27,28}, on physical and acoustic cues^{29–32} (e.g., human facial kin recognition³³), or on the expression of some genes such as the human MHC³⁴ or mouse MUP³⁵. Additionally, in some species, there is a delayed sexual maturation in the presence of relatives from the opposite sex^{11,36}. Finally, a more straightforward mechanism, even though it is costly³⁷, is to avoid relatives by geographically dispersing^{11,38,39}. However, it is unclear if this dispersal is aimed originally at avoiding inbreeding or rather at reducing intra-sexual competition⁴⁰. In western gorillas, females appear to specifically avoid related males when dispersing⁴¹, which supports the former hypothesis. In humans, geographical exogamy, *i.e.*, choosing spouses among geographically distant individuals, has been interpreted as a mechanism to avoid marriages between related individuals⁴². One line of evidence supporting this hypothesis comes from the negative relationship observed between exogamy and individual inbreeding in inhabitants from the Orkney Islands of Scotland⁴³. Despite this observation, there is a lack of data, whether in humans or in other species, to better understand the general impact of geographical distance between mates on both mating choice inbreeding and drift inbreeding patterns, as well as on population genetic diversity. This is likely because dispersal data are hard to obtain for most species, and combining such data with genetic information is even more challenging⁴⁴.

To explore the impact of geographical exogamy on genetic diversity at both the population and individual scales, we jointly analysed ethnological, geographical and genome-wide data in 16 populations corresponding to 11 distinct ethnic groups from Inner Asia, *i.e.*, in Uzbekistan, Kyrgyzstan, Tajikistan, West Mongolia and South Siberia. Inner Asia, the region located between the Caspian Sea in the West and Lake Baikal in the East, is particularly interesting in this context, as two groups presenting contrasting cultural traits (notably their language and their matrimonial system) cohabit. The first group, composed of Indo-Iranian-speaking populations, practices mainly geographically endogamous marriages, while the second group, Turkic- and Mongolic-speaking populations (referred to later as Turko-Mongol populations), practices mostly geographically exogamous marriages⁴⁵. Our genomic data allowed us to untangle drift inbreeding, which is due to small population sizes, from mating choice inbreeding, which is due to matrimonial preferences. To our knowledge, this is the first time that such quantitative data have been combined in humans or in any other species.

Results

Dispersal behaviours in Inner Asia. To quantify the amount of dispersal among spouses, we collected ethno-demographic questionnaires including spatial information for 1,344 individuals (a total of 643 couples) in 16 populations belonging to 11 distinct ethnic groups from Inner Asia and corresponding to 4 Indo-Iranians and 12 Turko-Mongols populations (see Fig. 1 and Supplementary Tables 1A,B). We calculated the distance between

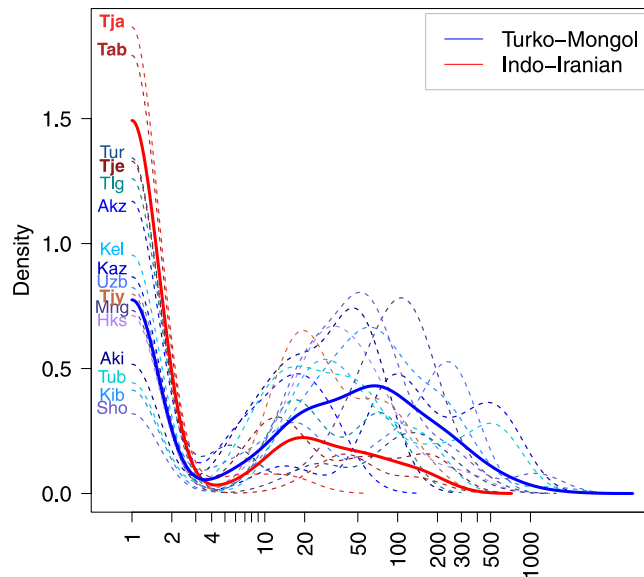


Figure 2. Geographical distances between the birth places of couples in Turko-Mongols and Indo-Iranians. The geographical distances are plotted in log scale (km). Their densities are represented by population (dashed lines) or for the Indo-Iranian and Turko-Mongol groups (solid lines). We represented the average distances within couples per population using a Kernel's density estimate implemented in R with a smoothing bandwidth of 0.2. See Supplementary Table 1B for population codes.

the birth places of spouses for each of the 643 couples (Fig. 2); this distance ranged between 0 km for spouses born in the same village (strictly endogamous couples) to 1,474 km, with a median of 5.6 km. We found that Turko-Mongols choose spouses born at larger distances than Indo-Iranians (a median of 17 km and 0 km, respectively; one-tailed Mann-Whitney U, or MWU, test p -value < 0.001). Based on the local minimum of the distance densities, we set the limit for geographical exogamy at 4 km and found the percentage of exogamous couples per population to be significantly higher in the 12 Turko-Mongol populations (60% on average, from 33% to 84%) than in the four Indo-Iranians (28% on average, from 6% to 60%) (Supplementary Fig. 1A, Supplementary Table 1, one-tailed MWU test p -value = 0.021). When defining exogamy with other thresholds, we still found significant differences between groups at 10, 20 and 50 km (p -value < 0.034) and a similar trend for 30 and 40 km (p -value = 0.051 and 0.057, respectively). As this result could only be due to differences in average exogamous proportions between groups, we also focused exclusively on exogamous couples (>4 km) and found a tendency for larger distances for Turko-Mongol couples than for Indo-Iranians (60 km and 42 km, respectively, one-tailed MWU test p -value = 0.023). Therefore, not only are Turko-Mongol couples geographically more exogamous than Indo-Iranians but they are also composed of spouses born at slightly larger distances.

In addition to the distance within the sampled couples, for each of the 1,344 sampled individuals, we calculated the distance between the birth places of their parents (referred to as the “parental couples”) (Supplementary Fig. 2). As found for the current generation, Turko-Mongols parental couples have higher exogamy rates than those of Indo-Iranians (34% and 11%, respectively; one-tailed MWU test p -value = 0.001) (Supplementary Fig. 1B), and these parental distances are significantly larger for Turko-Mongols (one-tailed MWU test p -value < 0.001). Furthermore, the exogamy rates of both generations are significantly correlated (Spearman's ρ = 0.81; p -value = 0.001), even though we observed lower exogamy rates for the parental couples.

Genetic diversity in Turko-Mongol and Indo-Iranian populations. To explore the differences in genetic diversity between Turko-Mongols and Indo-Iranians, we genotyped a subset of 503 individuals among the 16 populations under study (which corresponds to a median of 27 per population) and computed allele-sharing dissimilarity (ASD) distances⁴⁶, a measure based on the proportion of shared alleles between all pairs of individuals within each population (Supplementary Fig. 3). We found a higher median ASD distance for Indo-Iranians than for Turko-Mongols (on average, 0.280 and 0.269, respectively; MWU test p -value = 0.029). Because genotyping arrays are known to be enriched for European SNPs⁴⁷ and Indo-Iranians are genetically closer to Europeans than Turko-Mongols^{48,49}, the lower diversity observed within Turko-Mongols could be an artefact. To correct for this ascertainment bias, we used another measure of population diversity which is less sensitive to ascertainment bias than site-by-site measures^{50,51}: the averaged haplotypic heterozygosity computed for each autosome over low-recombination rate blocks. For this measure, we similarly found significantly higher mean heterozygosities for Indo-Iranians than for Turko-Mongols (on average across autosomes, 0.71 versus 0.69, respectively; one-tailed MWU test p -value = 0.029) (Fig. 3A, Supplementary Table 1A). Therefore, Turko-Mongol populations overall have a lower genetic diversity than Indo-Iranians, in accordance with their previously reported smaller effective population size^{49,52}. These two cultural groups indeed represent two genetic groups with contrasting demographic histories^{48,49}, as confirmed in this dataset (see Supplementary Fig. 4, Supplementary Table 2).

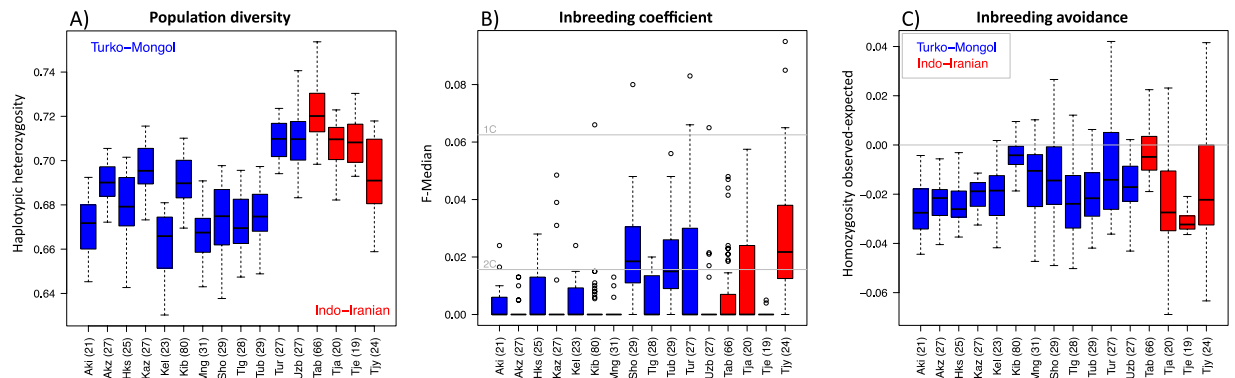


Figure 3. Genetic diversity (A) and inbreeding patterns (B,C) within populations. Grey lines in (B) represent inbreeding values corresponding to second-cousins and first-cousins. The grey line in (C) represents the homozygosity population baseline expected under panmixia. The number of samples per population is indicated between parentheses. See Supplementary Table 1B for population codes.

Inbreeding differences between populations. To investigate inbreeding differences between populations, we used a site-by-site inbreeding coefficient (F-Median)⁵³ (Fig. 3B) included in the FSuite pipeline⁵⁴ and corresponding to the probability of identity by descent at each marker of the genome⁵³. Across populations, 39% of the individuals had positive inbreeding coefficients (from 10 to 90% of each population), with a similar average among Turko-Mongols and Indo-Iranians (39% and 38%, respectively, MWU test p -value = 0.86). Moreover, for each of the 503 genotyped individuals, we inferred, from a likelihood ratio test performed by FSuite, the most likely parental mating type, *i.e.*, inbred mating between first or second cousins, or avunculars, or outbred mating (between less related individuals). Using this likelihood test, we found that 36% of the individuals are inbred across populations with a similar average between Turko-Mongol and Indo-Iranian populations (35% and 36%, respectively; MWU test p -value = 0.86). The majority of inbred individuals have parents that are second-cousins: they represent between 5% and 76% of all individuals, depending on the population (Supplementary Table 1A), while individuals with parents being first-cousins were detected only in nine out of the 16 studied populations and represent up to 21% of individuals, without any significant differences between the two cultural groups (MWU test p -value = 0.86 and 0.41 for second- and first-cousins, respectively). None of the parents have avuncular relationships, but one case of double-first-cousins was inferred. Overall, the Turko-Mongol and Indo-Iranian groups have a similar distribution of these mating type categories: on average, 31% and 27% of second-cousins, 4% and 7% of first-cousins, and 0% versus less than 1% of double-first-cousins (non-significant chi² test p -value with Yates's correction = 0.45), respectively.

We also used runs of homozygosity (ROHs), *i.e.*, regions of the genome where the paternal and maternal copies are identical and which are long enough to be considered as identical by descent. Their sizes and numbers are therefore informative for the nature of inbreeding⁵⁵. We aimed at distinguishing ROHs of an intermediate size, probably resulting from matings between individuals sharing distant ancestry (and therefore mostly due to drift), from long ROHs likely derived from matings between close relatives (and thus due to matrimonial preferences⁴³), even though they could rise from other mechanisms^{56,57}. First, using classical thresholds from the literature^{43,57–59}, we found that Turko-Mongol individuals have statistically more intermediate ROHs (500–1,500 kb) than Indo-Iranians (52.2 versus 36.6; MWU test p -value < 0.001; Supplementary Fig. 5A), and these ROHs represent a larger portion of their genomes (44,097 kb versus 30,613 kb, respectively; MWU test p -value < 10⁻³⁶; Supplementary Fig. 5B). Then, for long ROHs (>1,500 kb), we observed that the Turko-Mongol individuals have more ROHs than Indo-Iranian individuals (10.9 versus 9.0, respectively; MWU test p -value = 10⁻⁶) (Supplementary Fig. 5C), suggesting more matings between closely related individuals in Turko-Mongols. This is not in agreement with our results from FSuite, where we did not detect any significant difference in mating choice inbreeding between groups. However, when we instead defined boundaries per population with a model-based clustering algorithm as conducted in Pemberton *et al.*⁵¹, we obtained different thresholds for the ROH classes: on average, class B (intermediate) ROHs range between 885 and 2,647 kb, while class C (long) ROHs are above 2,647 kb (Supplementary Table 1A). Using these new boundaries, we did not detect differences either for the number or for the total length of class C-ROHs between Turko-Mongols and Indo-Iranians (MWU test p -value = 0.14 and 0.24, respectively, Supplementary Fig. 5C,D), but we still found more and longer class B-ROHs in Turko-Mongols (28.3 versus 20.1, respectively; MWU test p -value < 10⁻²²; and 36,340 versus 26,389 kb, respectively; MWU test p -value = 10⁻¹⁷, Supplementary Fig. 5A,B).

Inbreeding avoidance by geographical dispersal? To investigate inbreeding avoidance, we compared the observed homozygous proportion of each individual genome to the expectation under random mating⁶⁰. In each population, we found most individuals to be less homozygous than expected: overall, the median differences between observed and expected homozygosity are negative and are between -0.032 and -0.004 (average of -0.019), without any significant difference between Indo-Iranians and Turko-Mongols (MWU test p -value = 0.31; Fig. 3C). To test whether this result is restricted to Inner Asia, we analysed 11 worldwide populations from HapMap3 and found that they also have negative median values, except for the Gujarati Indians from Houston (GIH), who have a median of 1.6×10^{-6} (Supplementary Fig. 6).

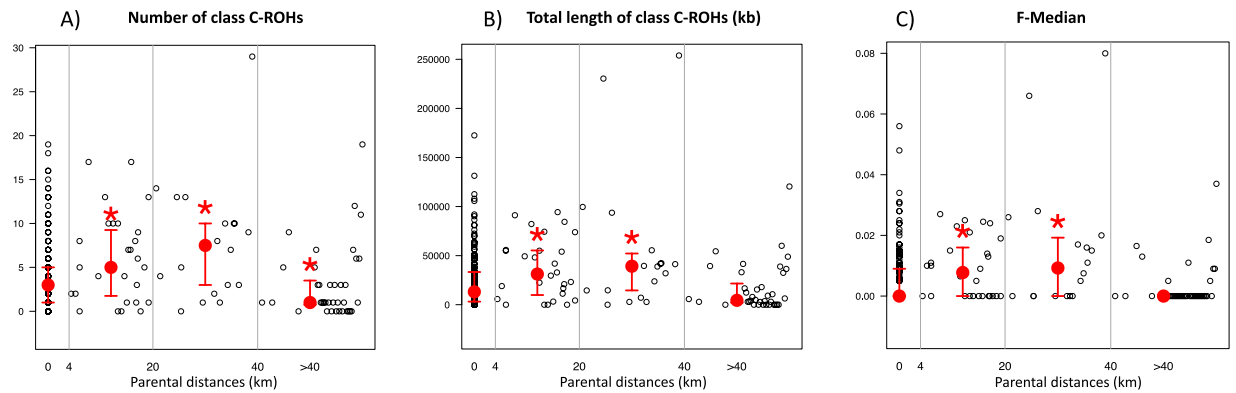


Figure 4. Parental couple distances and close inbreeding patterns for descendants from Turko-Mongols. For each of three estimators of inbreeding, we show in red the median value for four different classes of parental distances (<4 km, 4–20 km, 20–40 km and >40 km) as well as confidence intervals at 25% and 75%. The results of a MWU test comparing the <4 km class to all three others is shown below the plot.

We then tested whether exogamous populations include less inbred individuals than endogamous ones as expected if geographical dispersal is a mechanism of inbreeding avoidance. Surprisingly, in our dataset, the exogamy rate per population (defined with a limit of 4 to 50 km) is significantly correlated neither with the percentage of closely inbred individuals nor with the proportion of first-cousins' descendants (Spearman's correlation test p -value > 0.1 for all populations, only Turko-Mongols or only Indo-Iranians). Similarly, in terms of genetic diversity, the population exogamy rates are not associated with haplotypic heterozygosities (Spearman's correlation test: p -value > 0.29 for Turko-Mongols and p -value = 0.75 for Indo-Iranians). We further verified these results when taking into account potential confounding factors such as the age and sex ratio within each population, various environmental variables (altitude, ecosystem), as well as lifestyle (urban/rural and sedentary/nomadic) (Supplementary Table 1B). Considering the effect of each variable one at a time, we found that only the ecosystem had a significant effect on the percentage of inbred individuals per population while both the sedentary/nomadic, ecosystem and sex-ratio variables had a significant effect on the haplotypic heterozygosity (linear regressions, p -values < 0.05). However, when using multiple linear regression models and ANOVAs taking into account various combinations of these significant variables, we never found the exogamy rate to be significantly correlated with the percentage of inbred individuals per population nor with the haplotypic heterozygosity (p -values > 0.10).

We further explored the correlation between exogamy and inbreeding at the individual scale, *i.e.*, analysing the relationship between the mating choice inbreeding pattern of each individual and the geographical distance between his/her parents' birth places. This was only done in Turko-Mongols, as there were too few descendants from exogamous couples in Indo-Iranians. Interestingly, as seen in Fig. 4, we observed that the relationship between mating choice inbreeding and geographical distances between parents is not linear. Indeed, we found that descendants from exogamous couples born 4 to 40 km apart have significantly higher inbreeding levels than descendants from endogamous couples (≤ 4 km) (whether measured as the number or total length of class C-ROHs or as the F-Median: MWU test p -values < 0.05). Then, inbreeding decreases such that descendants from long-range exogamous couples (>40 km) are as inbred as descendants from endogamous couples for class C-ROH total length and F-Median (MWU test p -values > 0.05) and significantly less inbred for the number of class C-ROHs (MWU test p -value = 0.03). Consistently, we found that the distance between exogamous spouses (in log scale) had an effect on their descendant's inbreeding (for number and total length of class C-ROHs and F-Median coefficient, Spearman's ρ between -0.36 and -0.32 , p -values < 3×10^{-4}) (Supplementary Fig. 7).

Overall, among descendants from exogamous couples (>4 km), 30% have parents that are second-cousins, and 7% are first-cousins, which are proportions that do not differ from those obtained for endogamous couples (26% and 5%) (Supplementary Fig. 8).

Discussion

Two cultural groups, which matrimonial systems are reported to differ, coexist in Inner Asia: Turko-Mongols are described as mainly exogamous while Indo-Iranians are thought to be mainly endogamous⁴⁵. However, it is not always clear if exogamy refers to clan (ethnic) or village (geographical) exogamy. Here, we used a dataset of 16 populations representing 11 different ethnic groups from both cultural groups and we quantified geographical exogamy rates and distances in each population. Using an empirical threshold of 4 km, we confirmed that matrimonial behaviours differ as described in the literature, even though we found some exceptions: three Turko-Mongol populations (out of 14) have less than 50% exogamy, whereas one Indo-Iranian population (out of four) has more than 50% exogamy. Of note, this latter Indo-Iranian exception is a peculiar Tajik population (Tjy) that underwent population displacement in the 1970s from the Yagnob Valley to the Dushanbe area (approximately 60 km apart) in Tajikistan⁶¹. In general, the exogamous rates measured for the parental and current generations are highly correlated, even though we observed a trend towards higher values in the current generation, possibly reflecting changes in matrimonial behaviours through time (e.g., linked to urbanisation) or some recall bias for information about parental couples.

We used these contrasted dispersal behaviours to investigate the level of inbreeding in these populations and to test whether geographical exogamy is a mechanism of inbreeding avoidance. First, we showed that inbreeding resulting from matings between relatives that are at least second-cousins indeed occurs in Inner Asia: on average, 36% of the individuals are inbred. This estimation exceeds the worldwide lower bound estimate of 10.4%¹ but is still within the range of estimations made for neighbouring regions (28.6% for Turko-Mongol and Tajik populations from Xinjiang; 35.7% in South Asia) from data available in http://consang.net/index.php/Global_prevalence. Interestingly, marriages between first-cousins that are prescribed in some Muslim societies from areas close to Inner Asia, such as Middle East⁶² or Pakistan⁶³, only represent a small proportion of the Inner Asian marriages despite their Muslim tradition. Furthermore, most of the mating choice inbred alliances in our dataset are between second-cousins, and no avuncular relationship was observed. As inbreeding information is lacking for Inner Asia, our data complete the global picture of mating choice inbreeding prevalence. Interestingly, among the 16 studied populations, we found a large variability of mating choice inbreeding level (from 5% to 83%) and of mating type proportions. In particular, three populations (Sho, Tub and Tjy) include more than 50% of inbred individuals (79%, 83% and 79%, respectively). Of all the factors we tested (altitude, ecosystem, rural/urban, nomadic/sedentary lifestyle, age or sex-ratio within populations), only the ecosystem had a significant effect on the rate of inbreeding, with the forest environment (*i.e.*, the Sho and Tub populations) being significantly associated with a higher inbreeding than the other environments. Further data would be needed to explore these relationships. Moreover, the population effective population size, as approximated by haplotypic heterozygosity, was not correlated with the level of close inbreeding observed in the populations, based on the proportions of i) inbred individuals per population, ii) first-cousins' descendants, and on iii) the population median F-Median (Spearman's correlation test *p-values* > 0.2).

From a methodological point of view, our study highlights the necessity to define for each population its own ROH boundaries rather than using arbitrary thresholds originally defined for European populations (*i.e.* between 500 and 1,500 kb and over 1,500 kb, respectively). Indeed, we detected some inconsistencies between the two approaches: the classes B and C-ROHs ranges only partly overlap with the arbitrary thresholds ranges, and the limit between class B and C is about two times larger than the 1,500 kb threshold. Consequently, the results obtained for ROHs longer than 1,500 kb are contradictory with those obtained for class C-ROHs and with the results obtained based on FSuite.

Another result of our study is that the observed inbreeding levels in all Inner Asian populations deviate from the expectation under random mating, with each population being composed of less inbred offspring than expected by chance. We found the same result in nearly all populations from the HapMap3 worldwide dataset, which suggests that this is a general pattern. Such a deficit of inbreeding could be due to the avoidance of close kin matings (*i.e.*, brother-sister) that are proscribed in all societies⁶⁴ or to other pre/post-copulatory mechanisms, and this remains an open question.

An additional important result of our study is that geographical distances are not negatively correlated with inbreeding, as could have been expected under an isolation-by-distance model⁶⁵. Interestingly, a recent study based on a large genealogical dataset, collected across Western Europe and North America, and including birth places information, similarly found an absence of correlation between relatedness and the distance between couples, for the cohorts born before 1850⁶⁶. Our analyses within present-day Turko-Mongols reveal more specifically that the structure of the relationship between geographical distance and mating choice inbreeding is not linear, but rather tends to be bell-shaped, and thus cannot be correctly assessed with a single correlation test. Indeed, descendants from parents born 4 to 40 km apart are more inbred than descendants from endogamous couples (≤ 4 km) or from long-range exogamous ones (> 40 km). As a consequence, close inbreeding exists despite geographical exogamy, and about a third of descendants from exogamous couples are inbred.

These results, in addition to those obtained by⁶⁶, highlight the importance of using geographic distances rather than exogamy rates to characterize the impact of exogamy on inbreeding, as already described when studying patrilocal⁶⁷. Indeed, when we compare mating choice inbreeding patterns for descendants from exogamous and endogamous couples defined for thresholds of 4, 10, 20 and 30 km, we find no significant differences (for number and total length of class C-ROHs and F-Median coefficient: MWU test *p-values* > 0.1). We only detect significantly lower values in descendants from exogamous couples for larger distances above 40 and 50 km (*p-values* < 0.03).

Such inbred matings of individuals born at intermediate distances can be explained by two scenarios. First, spouses could be chosen outside the village unit but still within a quite genetically homogeneous unit^{68,69}. However, under this hypothesis, we would expect inbreeding patterns to be similar between descendants of endogamous and intermediate exogamous couples, yet we observe a significant excess of mating choice inbreeding in the latter (4 to 40 km apart). A second scenario is that geographical exogamous marriages preferentially involve spouses from the same family (kinship endogamy). It is indeed known from previous ethnological work that in patrilineal societies where individuals cannot mate with related spouses from the paternal line, they often choose spouses among relatives from the maternal side⁷⁰. To properly examine this hypothesis in the particular case of Turko-Mongols, further analyses and sampling are required. Indeed, we would need to analyse genetic data for more exogamous couples as well as for individuals coming from the populations of origin of both spouses. Thereby, we would be able to explore spouses' kinship by comparison with other potential partners and genetic distances between the populations. Furthermore, based on the kinship estimated from the X chromosome, we could investigate if related partners belong to the same maternal line, as assumed for these patrilineal societies, which would be of great interest to understand matrimonial strategies.

Our results also challenge the intuition that exogamy necessarily increases the genetic diversity within a population and therefore reduces drift inbreeding. Indeed, we found that Turko-Mongol populations have a lower genetic diversity (as measured by the mean haplotypic heterozygosity) and more intermediate ROHs associated with drift inbreeding than those of Indo-Iranians despite higher exogamous rates. However, because these groups

have different demographic histories (see Supplementary Fig. 4) and effective population sizes⁵², we chose to focus on each group separately. Within each group, we also found that across populations, the exogamy rates are not correlated with haplotypic heterozygosities. Moreover, at the individual scale, the total length or number of class B-ROHs was not significantly different between descendants from endogamous and exogamous Turko-Mongol couples (MWU test *p-values* = 0.7 and 0.9, respectively). This result is contrary to what was observed in the Orkney Islands of Scotland⁴³; thus, this correlation might differ depending on the cultural context.

Overall, this research sheds light on mating choice preferences: we showed that two thirds of partners that have not dispersed did mate with unrelated individuals, and that drift and mating choice inbreeding is variable, even among close-by populations. We also provide new insights into the relationship between dispersal and inbreeding in humans, based on genetic data, and demonstrate that geographical exogamy is not necessarily negatively associated with mating choice inbreeding, but rather can have a more complex non-linear relationship. Contrary to the common situation in many animals, this finding suggests that Inner Asian human populations who practise exogamy at small geographical scales might be focused on alliance strategies that result in kinship endogamy. To test whether these patterns depend on the cultural context, these questions should be addressed in other human populations, for example, Canadian populations⁷¹ or European isolates⁷² for whom genealogical detailed data and genomic data are already available. In such populations where phenotypic data are also available, inbreeding patterns can further be used to map disease-associated mutations^{73,74}. More generally, it would be of great interest to elucidate the variability of mating choice among cultures by systematically investigating inbreeding, which is generally disregarded in human demographic studies^{75,76}.

Materials and Methods

Population samples. We collected data during several field expeditions conducted in Inner Asia between 2001 and 2012. We defined populations as groups of individuals living in a similar area and belonging to the same ethnic group, based on the self-reported spoken native language. We obtained geographical and ethnological information (birth place and home language) for 1,344 individuals (a total of 643 couples) and their parents in 16 Inner Asian populations, *i.e.* 12 Turko-Mongol populations and four Indo-Iranian populations (Fig. 1, Supplementary Table 1B). Among these, we collected DNA from blood or saliva in 907 individuals. Written informed consent was obtained for all participants, and our sampling and analyses followed ethical principles approved by the Museum National d'Histoire Naturelle (France), Ministry of Health (Altai Republic), Tuvan State University (Republic of Tuva), Khovd University (Mongolia) and Uzbek Academy of Sciences (Uzbekistan).

Geographical exogamy. We inferred geographical exogamy from the geographical distance between the birth places of spouses (see Supplementary Materials and Methods). These distances can be calculated between each husband and wife (*i.e.* current generation) or between the parents of the husband on the one hand, and those of the wife on the other hand (*i.e.* parental generation). We defined exogamous couples as those with spouses born at more than 4 km (average population local minima based on Fig. 2), and we also tested for other definitions of exogamy with arbitrary thresholds at 10, 20, 30, 40, and 50 km.

Genetic diversity. Among the 907 individuals with blood or saliva sampled, we extracted DNA and genotyped 576 of them on five different Illumina arrays (either 660W-Quad, OmniExpress, Omni1-Quad, Omni2.5, or Omni5Exome). We chose the individuals randomly to reach an average of approximately 30 individuals per population. Among them, 503 passed quality-control filters (see Supplementary Materials and Methods and Supplementary Table 3) and were unrelated (*i.e.*, we excluded 36 individuals found to be more related than first-cousins), which resulted in a median of 27 individuals per population (66% of males and 34% of females, Supplementary Table 1). After merging the arrays, we obtained data for 253,532 SNPs, including 105,858 independent SNPs ($r^2 < 0.5$). The SNPs data has been deposited at the European Genome-phenome Archive (EGA), under accession number EGAS00001002951. We computed allele-sharing dissimilarity (ASD) distances⁷⁷ between individuals using the software *asd*⁴⁶ (see Supplementary Materials and Methods). We also calculated the autosomal haplotypic heterozygosity over low-recombination rated blocks for each population based on⁷⁸ (see Supplementary Materials and Methods).

Inbreeding coefficients. For each individual, we estimated the inbreeding coefficient (F-Median) which is the median value of multiple estimates of the probability of identity by descent at each marker of the genome⁵³, with FSuite v1.0.3⁵⁴. We estimated it by generating 100 subsets of independent SNPs (>0.5 cM)⁵⁵, and computing the median of these 100 coefficients. Furthermore, a likelihood ratio test was performed to assign each individual as inbred or outbred. For each individual, based on both F-Median and A-Median, it was possible to estimate the probability to be the descendant of avunculars (AV), double first-cousins (2x1C), first-cousins (1C), second-cousins (2C), or unrelated individuals (OUT; defined as less related than second-cousins). We kept the most likely inbred mating type for each individual. We also calculated the genomic excess of homozygosity⁶⁰ relative to an expected baseline of homozygosity for each of the Inner Asian populations and for the 11 populations from the HapMap3 worldwide dataset using plink1.9 (see Supplementary Materials and Methods). Furthermore, we identified runs of homozygosity, called ROHs³, using plink1.9 (see Supplementary Materials and Methods). First, to be able to compare with other studies from the literature^{43,57–59}, we categorized ROHs based on classical thresholds: 500–1,500 kb for intermediate ROHs, *versus* $> 1,500$ kb for long ROHs. However, as Pemberton *et al.*³ showed that these ROH thresholds can be variable between populations, we also used population-specific categories defined from their ROH size classification method (*i.e.* based on three-component Gaussian fitting of the ROH length distribution, using Mclust from the *mclust* package (v.5⁷⁹) in R) resulting in what we call intermediate (class B-ROHs) and long ROHs (class C-ROHs). For each of the four ROH classes, we computed the number of ROHs observed within an individual genome and their total length.

Data Availability. The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

References

- Bittles, A. Consanguinity and its relevance to clinical genetics. *Clinical Genetics* **60**, 89–98 (2008).
- Alvarez, G., Ceballos, F. C. & Quinteiro, C. The Role of Inbreeding in the Extinction of a European Royal Dynasty. *PLoS ONE* **4**, 1–7 (2009).
- Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics* **91**, 275–292 (2012).
- Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E. & Leutenegger, A.-L. High level of inbreeding in final phase of 1000 Genomes Project. *Scientific Reports* **5**, 1–7 (2015).
- Woods, C. G. *et al.* Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *American journal of human genetics* **78**, 889–896 (2006).
- Li, L.-H. *et al.* Long contiguous stretches of homozygosity in the human genome. *Human Mutation* **27**, 1115–1121 (2006).
- Mitton, J. B. In *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives* 17–41 (University of Chicago Press, 1993).
- Gao, Z., Waggoner, D., Stephens, M., Ober, C. & Przeworski, M. An Estimate of the Average Number of Recessive Lethal Mutations Carried by Humans. *Genetics* **199**, 1243–1254 (2015).
- Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783–796 (2009).
- Robert, A., Toupance, B., Tremblay, M. & Heyer, E. Impact of inbreeding on fertility in a pre-industrial population. *European Journal of Human Genetics* **17**, 673–681 (2009).
- Pusey, A. & Wolf, M. Inbreeding avoidance in animals. *Trends in Ecology & Evolution* **11**, 201–206 (1996).
- Fareed, M., Ahmad, M. K., Azeem Anwar, M. & Afzal, M. Impact of consanguineous marriages and degrees of inbreeding on fertility, child mortality, secondary sex ratio, selection intensity and genetic load: a cross-sectional study from Northern India. *Pediatric Research*. <https://doi.org/10.1038/pr.2016.177> (2016).
- Pedersen, J. The influence of consanguineous marriage on infant and child mortality among palestinians in the West Bank and Gaza, Jordan, Lebanon and Syria. *Community Genetics* **5**, 178–181 (2002).
- Alvarez, G., Quinteiro, C. & Ceballos, F. In *Advances in the Study of Genetic Disorders* 21–41 (InTech, 2011).
- Charlesworth, D. Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **358**, 1051–1070 (2003).
- Spielman, D., Brook, B. W. & Briscoe, D. A. & Frankham, R. Does inbreeding and loss of genetic diversity reduce disease resistance? *Conservation Genetics* **5**, 439–448 (2004).
- Frankham, R. Genetics and extinction. *Biological Conservation* **126**, 131–140 (2005).
- Saccheri, I. *et al.* Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**, 491–494 (1998).
- Sacks, O. *The Island of the Colorblind*. (A.A. Knopf, 1997).
- Ubbink, G. J., van de Broek, J., Hazewinkel, H. A. & Rothuizen, J. Cluster analysis of the genetic heterogeneity and disease distributions in purebred dog populations. *The Veterinary record* **142**, 209–13 (1998).
- Keller, L. & Waller, D. M. Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**, 230–241 (2002).
- Xue, Y. *et al.* Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).
- Moreno, E., Pérez-González, J., Carranza, J. & Moya-Laraño, J. Better Fitness in Captive Cuvier's Gazelle despite Inbreeding Increase: Evidence of Purging? *PLOS ONE* **10**(1–15), e0145111 (2015).
- Firman, R. C. & Simmons, L. W. Gametic interactions promote inbreeding avoidance in house mice. *Ecology Letters* **18**, 937–943 (2015).
- Holmes, W. G. & Sherman, P. W. The Ontogeny of Kin Recognition in Two Species of Ground Squirrels. *American Zoologist* **22**, 491–517 (1982).
- Hepper, P. G. Kin Recognition: Functions and Mechanisms a Review. *Biological Reviews* **61**, 63–93 (1986).
- Westermarck, E. *The history of human marriage*. (Allerton Book Company, 1921).
- Wolf, A. P. Childhood Association and Sexual Attraction: A Further Test of the Westermarck Hypothesis. *American Anthropologist* **72**, 503–515 (1970).
- Hauber, M. E. & Sherman, P. W. Self-referent phenotype matching: Theoretical considerations and empirical evidence. *Trends in Neurosciences* **24**, 609–616 (2001).
- Pfefferle, D., Kazem, A. J. N., Brockhausen, R. R., Ruiz-Lambides, A. V. & Widdig, A. Monkeys Spontaneously Discriminate Their Unfamiliar Paternal Kin under Natural Conditions Using Facial Cues. *Current Biology* **24**, 1806–1810 (2014).
- Levréro, F. *et al.* Social shaping of voices does not impair phenotype matching of kinship in mandrills. *Nature Communications* **6**, 1–7 (2015).
- Crepy, M. A. & Casal, J. J. Kin recognition by self-referent phenotype matching in plants. *New Phytologist* **209**, 15–16 (2016).
- DeBruine, L. M., Jones, B. C., Little, A. C. & Perrett, D. I. Social perception of facial resemblance in humans. *Archives of Sexual Behavior* **37**, 64–77 (2008).
- Laurent, R. & Chaix, R. MHC-dependent mate choice in humans: Why genomic patterns from the HapMap European American dataset support the hypothesis. *BioEssays* **34**, 267–271 (2012).
- Green, J. P. *et al.* The Genetic Basis of Kin Recognition in a Cooperatively Breeding Mammal. *Current Biology* **25**, 2631–2641 (2015).
- Hoogland, J. L. Prairie Dogs Avoid Extreme Inbreeding. *Science* **215**, 1639–1641 (1982).
- Ronce, O. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics* **38**, 231–253 (2007).
- Lambin, X. Natal Philopatry, Competition for Resources, and Inbreeding Avoidance in Townsend's Voles (*Microtus townsendii*). *Ecology* **75**, 224–235 (1994).
- Pusey, A. E. & Packer, C. The Evolution of Sex-Biased Dispersal in Lions. *Behaviour* **101**, 275–310 (1987).
- Pusey, A. In *Pedophilia* 201–220. https://doi.org/10.1007/978-1-4613-9682-6_8 (Springer New York, 1990).
- Bradley, B. J., Doran-Sheehy, D. M. & Vigilant, L. Potential for female kin associations in wild western gorillas despite female dispersal. *Proceedings. Biological sciences/The Royal Society* **274**, 2179–85 (2007).
- Fix, A. G. *Migration and colonization in human microevolution*. at [http://dx.doi.org/10.1002/1520-6300\(200102/03\)13:2%3C280::AID-AJHB1040%3E3.0.CO;2-6](http://dx.doi.org/10.1002/1520-6300(200102/03)13:2%3C280::AID-AJHB1040%3E3.0.CO;2-6) (Cambridge University Press, 1999).
- McQuillan, R. *et al.* Runs of Homozygosity in European Populations. *The American Journal of Human Genetics* **83**, 359–372 (2008).
- Ims, R. a. & Yoccoz, N. G. In *Metapopulation Biology* 247–265. <https://doi.org/10.1016/B978-012323445-2/50015-8> (Elsevier, 1997).
- Krader, L. Peoples of Central Asia, 2nd edition. *Indiana University Publications, Bloomington* **26** (1966).
- Szpiech, Z. A. asd computer program. <http://szpiech.com/perform.html> (2011).
- Stoneking, M. Single nucleotide polymorphisms: From the evolutionary past. *Nature* **409**, 821–822 (2001).
- Martínez-Cruz, B. *et al.* In the heartland of Eurasia: the multilocus genetic landscape of Central Asian populations. *European Journal of Human Genetics* **19**, 216–223 (2011).

49. Marchi, N. *et al.* Sex-specific genetic diversity is shaped by cultural factors in Inner Asian human populations. *American Journal of Physical Anthropology* **162**, 627–640 (2017).
50. Conrad, D. F. *et al.* A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature genetics* **38**, 1251–1260 (2006).
51. Pemberton, T. J. & Rosenberg, N. A. Population-Genetic Influences on Genomic Estimates of the Inbreeding Coefficient: A Global Perspective. *Human Heredity* **77**, 37–48 (2014).
52. Palstra, F. P., Heyer, E. & Austerlitz, F. Statistical Inference on Genetic Data Reveals the Complex Demographic History of Human Populations in Central Asia. *Molecular Biology and Evolution* **32**, 1411–1424 (2015).
53. Leutenegger, A.-L. *et al.* Estimation of the Inbreeding Coefficient through Use of Genomic Data. *The American Journal of Human Genetics* **73**, 516–523 (2003).
54. Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E. & Leutenegger, A.-L. FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics* **30**, 1940–1941 (2014).
55. Leutenegger, A.-L., Sahbatou, M., Gazal, S., Cann, H. & Génin, E. Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? *European Journal of Human Genetics* **19**, 583–587 (2011).
56. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789–795 (2006).
57. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nature reviews Genetics*. <https://doi.org/10.1038/nrg.2017.109> (2018).
58. Joshi, P. K. *et al.* Directional dominance on stature and cognition in diverse human populations. *Nature* **523**, 459–62 (2015).
59. Kirin, M. *et al.* Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLoS ONE* **5**(1–7), e13996 (2010).
60. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
61. Gunà, A. N. In *Cahiers d'Asie centrale (11/12)* 161–178 (EDISUD, 2004).
62. Al-Gazali, L., Hamamy, H. & Al-Arrayad, S. Genetic disorders in the Arab world. *BMJ (Clinical research ed.)* **333**, 831–834 (2006).
63. Shami, S. A., Schmitt, L. H. & Bittles, A. H. Consanguinity related prenatal and postnatal mortality of the populations of seven Pakistani Punjab cities. *Journal of Medical Genetics* **26**, 267–271 (1989).
64. Thompson, E. A. & Roberts, D. F. Kinship structure and heterozygosity on Tristan da Cunha. *American journal of human genetics* **32**, 445–52 (1980).
65. Malécot, G. *Les Mathématiques de l'Hérédité*. (1948).
66. Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of relatives. **9309**, 1–12 (2018).
67. Marks, S. J., Levy, H., Martinez-Cadenas, C., Montinaro, F. & Capelli, C. Migration distance rather than migration rate explains genetic diversity in human patrilocal groups. *Molecular Ecology* **21**, 4958–4969 (2012).
68. Boyce, A. J., Küchemann, C. F. & Harrison, G. A. Neighbourhood knowledge and the distribution of marriage distances. *Annals of Human Genetics* **30**, 335–338 (1967).
69. Wright, S. Isolation by distance under diverse systems of mating. *Genetics* **31**, 39–59 (1946).
70. Murdock, G. P. *Ethnographic Atlas: A Summary*. *Ethnology* **6**, 109–236 (1967).
71. Roy-Gagnon, M. H. *et al.* Genomic and genealogical investigation of the French Canadian founder population structure. *Human Genetics* **129**, 521–531 (2011).
72. Colonna, V. *et al.* Comparing population structure as inferred from genealogical versus genetic information. *European Journal of Human Genetics* **17**, 1635–1641 (2009).
73. Génin, E. & Todorov, A. A. In *Encyclopedia of Life Sciences* 1–5, <https://doi.org/10.1038/ngp.els.0005407> (John Wiley & Sons, Ltd, 2006).
74. Redden, D. T. & Allison, D. B. The Effect of Assortative Mating upon Genetic Association Studies: Spurious Associations and Population Substructure in the Absence of Admixture. *Behavior Genetics* **36**, 678–686 (2006).
75. Larmuseau, M. H. D., Van Geystelen, A., van Oven, M. & Decorte, R. Genetic genealogy comes of age: Perspectives on the use of deep-rooted pedigrees in human population genetics. *American Journal of Physical Anthropology* **150**, 505–511 (2013).
76. Sebro, R. & Risch, N. J. A brief note on the resemblance between relatives in the presence of population stratification. *Heredity* **108**, 563–568 (2012).
77. Mountain, J. L. & Cavalli-Sforza, L. L. Multilocus genotypes, a tree of individuals, and human evolutionary history. *American journal of human genetics* **61**, 705–18 (1997).
78. Verdu, P. *et al.* Patterns of Admixture and Population Structure in Native Populations of Northwest North America. *PLoS Genetics* **10**(1–17), e1004530 (2014).
79. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal* **8**, 289–317 (2016).
80. Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. maps: Draw Geographical Maps. at <https://cran.r-project.org/package=maps> (2017).
81. Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. mapdata: Extra Map Databases, at <https://cran.r-project.org/package=mapdata> (2016).
82. R Core Team. R: A Language and Environment for Statistical Computing (2015).

Acknowledgements

We wish to thank W. Carpentier (Plate-forme Post-Génomique de la Pitié-Salpêtrière), B. Regnault and L. Lemée (Institut Pasteur – Genopole) for the genotyping processing; L. Berman for developing the code to compute geographical distances; A.-L. Leutenegger and S. Gazal for precious help with FSuite; and F. Austerlitz, R. Chaix, R. Laurent, G. Ly, and P. Verdu for constructive discussions; and E. Verzele. The project was funded by the ANR NUTGENEVOL (07-BLAN-0064); ANR Altérité culturelle (10-ESVS-0010); CNRS Programme international de collaboration scientifique.

Author Contributions

E.H., N.M., and L.S. interpreted the results and wrote the manuscript. B.C., T.D., T.H., E.H., P.M. and L.S. participated in the sampling. M.G. and S.L. handled the DNA samples. N.M. analysed the data. E.H. and L.S. designed the study.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27047-3>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018