HAL
open science

Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations

Patricia Balaresque, Nicolas Poulet, Sylvain Cussat-Blanc, Patrice Gerard, Lluis Quintana-Murci, Evelyne Heyer, Mark A. Jobling

To cite this version:

Patricia Balaresque, Nicolas Poulet, Sylvain Cussat-Blanc, Patrice Gerard, Lluis Quintana-Murci, et al.. Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. European Journal of Human Genetics, 2015, 23 (10), pp.1413-1422. 10.1038/ejhg.2014.285 . hal-01280347

HAL Id: hal-01280347

https://hal.science/hal-01280347

Submitted on 29 Feb 2016

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

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

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 14821

To link to this article : DOI :10.1038/ejhg.2014.285
URL : http://dx.doi.org/10.1038/ejhg.2014.285

To cite this version : Balaresque, Patricia and Poulet, Nicolas and Cussat-Blanc, Sylvain and Gerard, Patrice and Quintana-Murci, Lluis and Heyer, Evelyne and A Jobling, Mark *Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations.* (2015) European Journal of Human Genetics, vol. 23. pp. 1413-1422. ISSN 1018-4813

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations

Patricia Balaresque*[,1,2], Nicolas Poulet[3,7], Sylvain Cussat-Blanc[4,7], Patrice Gerard[1], Lluis Quintana-Murci[5], Evelyne Heyer[6,8] and Mark A Jobling*[,2,8]

High-frequency microsatellite haplotypes of the male-specific Y-chromosome can signal past episodes of high reproductive success of particular men and their patrilineal descendants. Previously, two examples of such successful Y-lineages have been described in Asia, both associated with Altaic-speaking pastoral nomadic societies, and putatively linked to dynasties descending, respectively, from Genghis Khan and Giocangga. Here we surveyed a total of 5321 Y-chromosomes from 127 Asian populations, including novel Y-SNP and microsatellite data on 461 Central Asian males, to ask whether additional lineage expansions could be identified. Based on the most frequent eight-microsatellite haplotypes, we objectively defined 11 descent clusters (DCs), each within a specific haplogroup, that represent likely past instances of high male reproductive success, including the two previously identified cases. Analysis of the geographical patterns and ages of these DCs and their associated cultural characteristics showed that the most successful lineages are found both among sedentary agriculturalists and pastoral nomads, and expanded between 2100 BCE and 1100 CE. However, those with recent origins in the historical period are almost exclusively found in Altaic-speaking pastoral nomadic populations, which may reflect a shift in political organisation in pastoralist economies and a greater ease of transmission of Y-chromosomes through time and space facilitated by the use of horses.

INTRODUCTION

Reproductive success, widely used as a measure of fitness, is described as the genetic contribution of an individual to future generations.[1] Human behavioural ecologists use various fertility-specific measures to directly estimate reproductive success in extant populations (eg, Strassmann and Gillespie[2]); however, indirect estimates of past reproductive success can also be obtained through evolutionary population genetic approaches, which demonstrate that its cultural transmission can be extremely effective.[3,4]

High variance of male reproductive success is detectable from genetic data because it leads to many Y-chromosomal lineages becoming extinct through drift and others expanding markedly. For any increase to lead to a high population frequency, two factors are needed: biological, with a need for men to be fertile, and cultural, with a continued transmission of reproductive success over generations. Indeed, previously, strong signals of successful transmission of Y-lineages have been associated with recent social selection, and were explained by inherited social status, with two cases in Asia and one in the British Isles. The best-known instance is the finding that ~ 0.5% of the world's Y-chromosomes belongs to a single Asian patrilineage,[5] descending from a common ancestor in historical times, and suggested to be due to the imperial dynasty founded by Genghis Khan (died 1227). The lineage is characterised by a high-frequency Y-microsatellite haplotype and a set of close mutational neighbours—a so-called 'star-cluster'. In its structure and haplotype diversity, it resembles the recent descent clusters (DCs) observed within rare British surnames,[6] despite its presence in populations spread over an enormous geographical range and speaking many different languages. Subsequently, two further examples of high-frequency DCs were described and were associated with the Qing dynasty descendants of Giocangga (died 1582) in Asia,[7] and the Irish early medieval 'Uí Néill' dynasty in Europe.[8] These three examples might suggest an association between high reproductive success (detectable at the population level) and both wealth and socio-political power, as is observed in some present-day populations (eg, the Tsimané of Bolivia[9]).

In this study, we ask whether the two signals of continued transmission of success over generations detected in mainland Asia are the only examples of likely recent social selection, or if similar patterns can be detected among the Y-chromosomes of this continent, known for the numerous expansive polities that emerged there during

[1]UMR 5288, Faculté de Médecine Purpan, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), CNRS/Université Paul Sabatier, Toulouse, France; [2]Department of Genetics, University of Leicester, Leicester, UK; [3]Onema, Direction de l'Action Scientifique et Technique, Toulouse, France; [4]UMR5505- CNRS, Université de Toulouse, Toulouse, France; [5]CNRS URA3012, Unit of Human Evolutionary Genetics, Institut Pasteur, Paris, France; [6]Eco-Anthropologie et Ethnobiologie, UMR 7206 CNRS, MNHN, Université Paris Diderot, Sorbonne Paris Cité, Sorbonne Universités, Paris, France

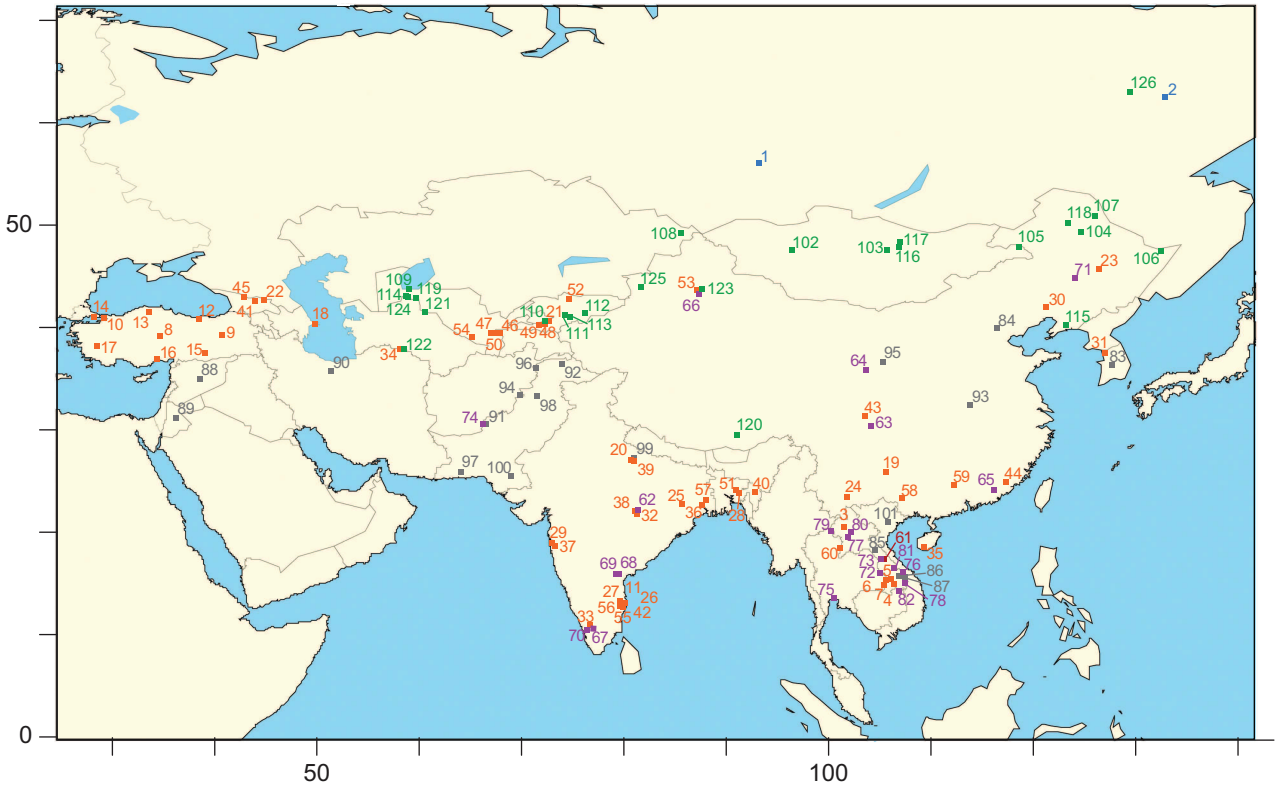
[7]These authors contributed equally to this work.

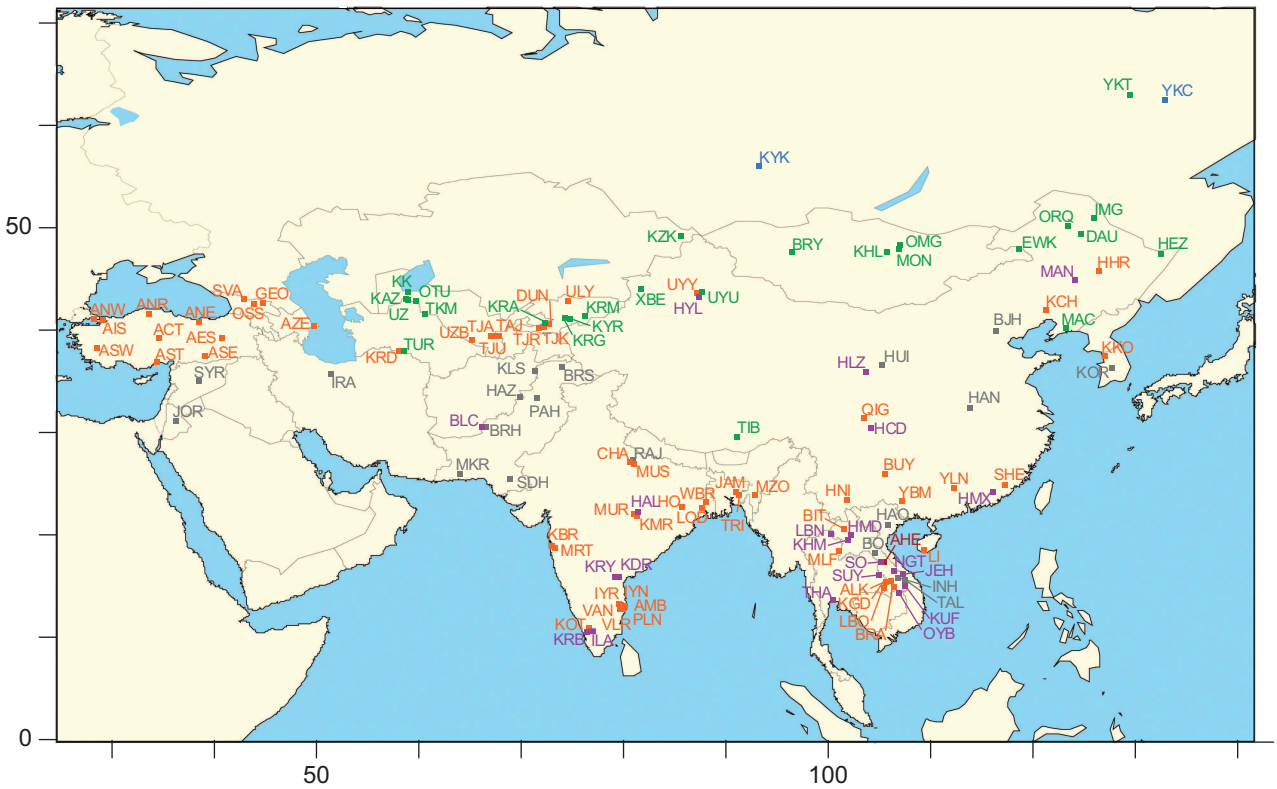
[8]These authors contributed equally to this work.

*Correspondence: Dr P Balaresque, UMR 5288, Faculté de Médecine Purpan, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AIMS), CNRS/Université Paul Sabatier, 37 allées Jules Guesde, 31073 Toulouse, France. Tel: +33 5 61 14 55 04; Fax: +33 5 61 14 55 04; E-mail: patricia.balaresque@univ-tlse3.fr

or Professor MA Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK. Tel: +44 116 252 3427; Fax: +44 116 252 3378; E-mail: maj4@le.ac.uk

the Bronze Age and later (beginning 5000–4000 YBP). Pinpointing the historical figures associated with any such signals of Y-lineage transmission would require their identification via ancient DNA testing, and/or a comparison with certified living descendants and will not be attempted here. Instead, we aim to understand whether efficient transmission is linked to rules that apply in populations with specific histories of subsistence or culture, or can be detected in other groups with diverse cultural features. We also ask whether any expansions detected converge to the same time period or reflect distinct temporal episodes.





To address these questions, we generated novel Y-microsatellite and Y-SNP data and combined these with published data in order to carry out a systematic analysis of the most frequent haplotypes and their associated DCs. We chose a lineage-based, or interpopulation approach, rather than an intrapopulation approach,[10] as we reasoned that dominant lineages are likely to migrate.[11] We considered the geographical pattern of identified expansions, their ages and the cultural factors (language and mode of subsistence) characteristic of each population in which they were detected. In total, the 5321 chromosomes analysed belong to 127 populations distributed from the Middle East to Korea. Our analysis focuses upon the 15 most frequent haplotypes that include two haplotypes reflecting the previously recognised 'Khan' and 'Giocangga' expansions.[5,7] Our analysis shows that the most successful Y-chromosomes in Asia form part of expansions that began between 2100 BCE and 1100 CE, found both among sedentary agriculturalists and pastoral nomads. The ages of these expansions are considered in their cultural contexts.

MATERIALS AND METHODS

DNA samples

DNA samples (461) from Central Asia were collected by the authors with appropriate informed consent.[12] An additional 4860 samples form part of sets published previously[13–21] with the exclusion of population samples containing fewer than 15 individuals. A total of 5321 males from 127 populations were analysed (Supplementary Table 1 and Figure 1).

Y-chromosome haplotyping and nomenclature

In the 461 Central Asian samples up to 31 binary markers were typed hierarchically by using the SNaPshot protocol on an ABI3100 capillary electrophoresis apparatus (Applied Biosystems Inc., Foster City, CA, USA). Primers were based on ones published previously[22–24] with additional primers based on published sequences.[25] The binary markers define haplogroups that are represented in a maximum parsimony tree[25] (Supplementary Figure 1). For comparison among data sets, some simplification of resolution was required, but the original haplogroup information was also retained and correspondence between previous nomenclatures was established.

Eight Y-specific microsatellites (DYS19, DYS389I, DYS389b (equivalent to DYS389II-DYS389I), DYS390, DYS391, DYS392, DYS393 and DYS439) were typed in two previously described multiplexes[26,27] by using an ABI3100 apparatus and GeneMapper v 4.0 (Applied Biosystems Inc.). When DYS19 was observed as two alleles, we retained for analysis the allele found at higher frequency in the population. Allele nomenclature was as described[27] with appropriate adjustments in some published data sets.

DC definition and display

Having identified frequent haplotypes, we applied a rule to define a DC centred upon each, using a script ('Cluster Generator'; Supplementary Information) to extract these DCs from the database. This selects individuals from the data set, starting from a given haplotype and including all haplotypes that are continuously traceable to a core haplotype via a pathway of increasing frequency, a topology consistent with a Poisson-distributed single-step mutational model.[8] An additional 'single-step+10% two-steps' model was tested, but had the effect of incorporating unrelated chromosomes belonging to distinct haplogroups. We therefore applied only the most conservative single-step model. For each population, we reported the proportion of each DC detected, and, as a measure of DC diversity, its average within-population microsatellite variance. The latter measure was used only for cases in which the number of DC chromosomes per population was >7 in order to avoid strong variance bias.[28,29] Use of seven or six microsatellites instead of eight identifies the same DCs (data not shown).

Frequencies of DCs and local microsatellite variances were calculated, and both frequency and highest variance value (source) were displayed using Surfer 8.02 (Golden Software, LLC, Golden, CO, USA) by the gridding method. Latitudes and longitudes were based on sampling centres.

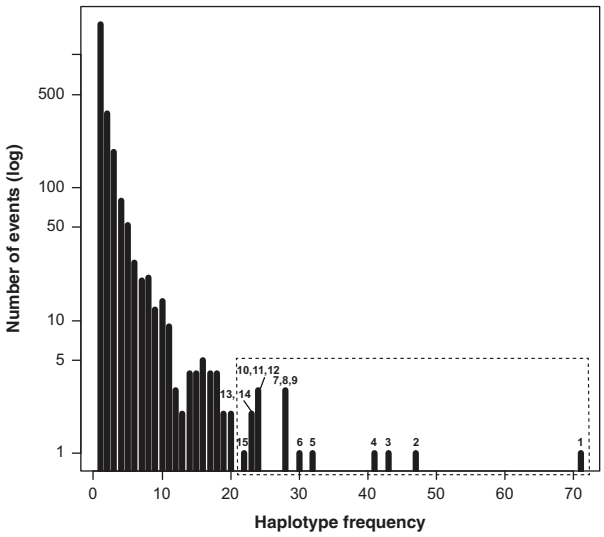


Figure 2 Microsatellite haplotype frequency distribution. The 2552 distinct microsatellite haplotypes among 5321 individuals are plotted by frequency. The 15 frequent haplotypes studied in detail here are numbered from 1 to 15, and contained in the dotted rectangle.

Figure 1 Locations of populations included in the analysis. In the upper panel, numbers represent a numerical code given to each population (Supplementary Table 1) and in the lower panel three-letter abbreviations of population names are given. AHE: Aheu; ALK: Alak; AMB: Ambalakarar; ACT: Anatolia Centre; AES: Anatolia East; AIS: Anatolia Istanbul; ANR: Anatolia North; ANE: Anatolia NE; ANW: Anatolia NW; AST: Anatolia South; ASE: Anatolia SE; ASW: Anatolia SW; AZE: Azeri; BLC: Balochi; BJH: Beijin-Han; BIT: Bit; BO: Bo; BRH: Brahui; BRA: Brau; BRS: Burusho; BRY: Buryat; BUY: Buyi; CHA: Chamar; DAU: Daur; DUN: Dungans; EWK: Ewenki; GEO: Georgian; HAL: Halba; HAN: Han; HCD: Han (Chendgu); HHR: Han (Harbin); HLZ: Han (Lanzhou); HMX: Han (Meixian); HYL: Han (Yili); HNI: Hani; HAO: Hanoi; HAZ: Hazara; HEZ: Hezhe; HMD: Hmong Daw; HO: Ho; HUI: Hui; INH: Inh; IMG: Inner Mongolian; IRA: Iran; ILA: Irula; IYN: Iyengar; IYR: Iyer; JAM: Jamatia; JAV: Javanese; JEH: Jeh; JOR: Jordan; KLS: Kalash; KMR: Kamar; OTU: Karakalpaks (On Tört Uruw); KK: Karakalpaks (Qongrat); KGD: Kataang; KUF: Katu; KAZ: Kazak (Karakalpakia); KZK: Kazak; KHL: Khalkh; KHM: Khmu; KBR: Koknasth Brahmin; KRY: Konda Reddy; KKO: Korean; KOR: Korean; KCH: Korean (China); KOT: Kota; KDR: Koya Dora; KYK: Krasnoyarsk Kurgan; KRD: Kurds; KRB: Kurumba; KYR: Kyrgyz; KRA: Kyrgyz (Andijan); KRM: Kyrgyz (Narin); KRG: Kyrgyz (Narin); LBN: Lamet; LBO: Laven; LI: Li; LOD: Lodha; MKR: Makrani; MLF: Mal; MAC: Manchu; MAN: Manchurian; MRT: Maratha; MZO: Mizo; MON: Mongolian; MUR: Muria; MUS: Muslim; NGT: Ngeq (Nkriang); ORQ: Oroqen; OSS: Ossetian; OMG: Outer Mongolian; UZ: Ouzbeks (Karakalpakia); OYB: Oy; PLN: Pallan; PAH: Pathan; QIG: Qiang; RAJ: Raiput; SHE: She; SDH: Sindhi; SO: So; SUY: Suy; SVA: Svan; SYR: Syria; TAJ: Tajik; TJK: Tajik (Ferghana); TJR: Tajik (Ferghana); TJA: Tajik (Samarkand); TJU: Tajik (Samarkand); TAL: Talieng; THA: Thai; TIB: Tibet; TRI: Tripuri; TUR: Turkmen; TKM: Turkmen (Karakalpakia); ULY: Uyghur; UYU: Uygur (Urumqi); UYY: Uygur (Yili); UZB: Uzbek; VAN: Vanniyan; VLR: Vellalan; WBR: W Bengali Brahman; XBE: Xibe; YKC: Central Yakutia; YKT: Yakut; YBM: Yao (Bama); YLN: Yao (Liannan). Different colours correspond to the mode of subsistence characteristic of each population as follows: orange—agricultural; green—pastoral nomadic; red—hunter–gatherer; purple—multiple modes; grey—information unavailable.

TMRCA estimation and multivariate statistical analysis

TMRCA estimates were generated using BATWING[30] using an exponential growth model, an average mutation rate of 0.0028 per locus per generation for microsatellites[31,32] and a generation time of 30 years, compatible with population-based estimates for males.[33] BATWING was run on samples comprising DCs to reduce computation time. Tests showed that running BATWING on populations and extracting DC TMRCAs led to similar results (data not shown).

Statistical tests were done using R.[34] PCA was carried out on the frequency of DCs per population (arcsine square root transformed) to test relative contributions of DCs and potential associations between expansions. To test for shared cultural features (language and subsistence methods), a co-inertia analysis[35] was carried out on the 95 populations for which cultural information was available. This examines two synthetic variables, one in each data set (here, frequency of DC and cultural features), with maximum covariance. The Rv-coefficient (a multivariate extension of the Pearson correlation coefficient) was calculated and tested by means of a Monte–Carlo method (1000 random permutations of the rows of each data set). PCA and co-inertia analyses were done using the ADE package.[36]

RESULTS

Identification of unusually frequent haplotypes and associated DCs

Through haplotyping of Central Asian samples (Supplementary Table 2) and a literature survey, we collected 5321 eight-locus Y-chromosomal microsatellite haplotypes belonging to 127 Asian populations (Figure 1 and Supplementary Table 1). We ranked haplotypes by frequency (Figure 2), reasoning that particularly frequent haplotypes should represent potential cores of clusters indicating past expansions of paternal lineages. The sample contained 2552 distinct haplotypes, 67% of which were unique, and 15 haplotypes (0.6%) were each present >20 times (and up to 71 times), indicating probable examples of successful transmissions (Figure 2). We focused on these 15 haplotypes and studied their spatial distributions and ages to illuminate the history of high reproductive success of their respective common ancestors.

Frequent haplotypes are expected to be accompanied by neighbouring haplotypes within the same Y-SNP haplogroup that arise from them via mutation, thereby forming DCs.[6] These DCs were uniformly and objectively extracted from the database using 'Cluster Generator'. When applied to the 15 most frequent haplotypes, this led to some common haplotypes merging into clusters centred on other common haplotypes, and thus to a total of 11 DCs, whose characteristics are summarised in Table 1.

Characteristics of DCs

A total of 2000 (37.8%) males carry Y-chromosomes belonging to these DCs, and they are remarkably widespread, with only three of the 127 populations analysed lacking any DC chromosomes; DC2 is also detected in one of the two ancient DNA population samples in our database (Krasnoyarsk_Kurgan[17]). Overall DC proportions in each population vary greatly from 5 to 96.7% (Supplementary Table 3).

As DC4 is restricted to Korea,[18] it was not considered for interpopulation analysis. For the remaining 10 DCs, we explored three characteristics for each: its geographical frequency distribution, the geographical distribution of its mean microsatellite variance to indicate its likely expansion source and its TMRCA to suggest the age of expansion. Frequency and variance are represented on maps (Figure 3 and Supplementary Figure 2), and are summarised in Table 2 and Figure 4, which also gives TMRCA estimates. Note that the dating method used, BATWING, provides estimates with large confidence intervals.

Our results show heterogeneity in DC features for all variables analysed: (i) *Geographical pattern*: two kinds of pattern are observed: six DCs display high frequencies spread over large geographic areas (DCs 1, 2, 6, 8, 10 and 14), and four DCs are more restricted to specific areas (DCs 3, 5, 8 and 12). (ii) *Origin*: three main source locations emerge: Silk Road/East Mongolia (DCs 1, 8 and 10), Middle East/Central Asia (DCs 2, 3 and 5) and East India/South East Asia (DCs 2, 6, 11, 12 and 14). The potential 'source' populations are given in Table 2. (iii) *Age*: mean TMRCA estimates range from 900 to 4100 YA. In considering these results, we focus on two periods: the 'historical' period (700–1060 CE) and the protohistorical period (2100–300 BCE), corresponding, respectively, to pre-Imperial

Table 1 Features of the fifteen primary descent clusters

DC	n	*Simplified haplogroup*[a]	*Highest resolution haplogroup*	*Core microsatellite haplotype*								*Link with other DC?*
				DYS19	*DYS389I*	*DYS389b*	*DYS390*	*DYS391*	*DYS392*	*DYS393*	*DYS439*	
DC1	71	C	C3 (xC3c)	16	13	16	25	10	11	13	10	No
DC2	47	R1a	R1a1	16	14	18	25	11	11	13	10	No
DC3	43	K(xN1c1,P)	L	14	12	16	22	10	14	11	13	No
DC4	41	K(xN1c1,P)	O2b	16	14	15	23	10	13	13	12	No
DC5	32	J	J	14	13	17	23	11	11	12	11	No
DC6	30	K(xN1c1,P)	O3a3c	14	12	16	24	10	14	12	12	No
DC7	28	J	J	14	13	17	23	10	11	12	11	DC5[b]
DC8	28	C	C3c	16	13	16	24	9	11	13	11	No
DC9	28	J	J	14	13	16	23	10	11	12	11	DC5[b]
DC10	24	K(xN1c1,P)	N	14	14	16	23	10	14	14	10	No
DC11	24	Y(xC,D,E,J,K)	H1a	15	13	16	22	10	11	12	12	No
DC12	24	K(xN1c1,P)	O2a	15	13	16	24	10	13	14	11	No
DC13	23	J	J1	14	13	16	23	11	11	12	11	DC5[b]
DC14	23	K(xN1c1,P)	O2a	15	13	16	25	11	13	14	12	No
DC15	22	R1a	R1a1	16	13	18	25	11	11	13	10	DC2 [b]

Abbreviation: DC, descent cluster.
[a]Simplified to allow combination of published data sets.
[b]These DCs were incorporated into the analysis of the linked DC.

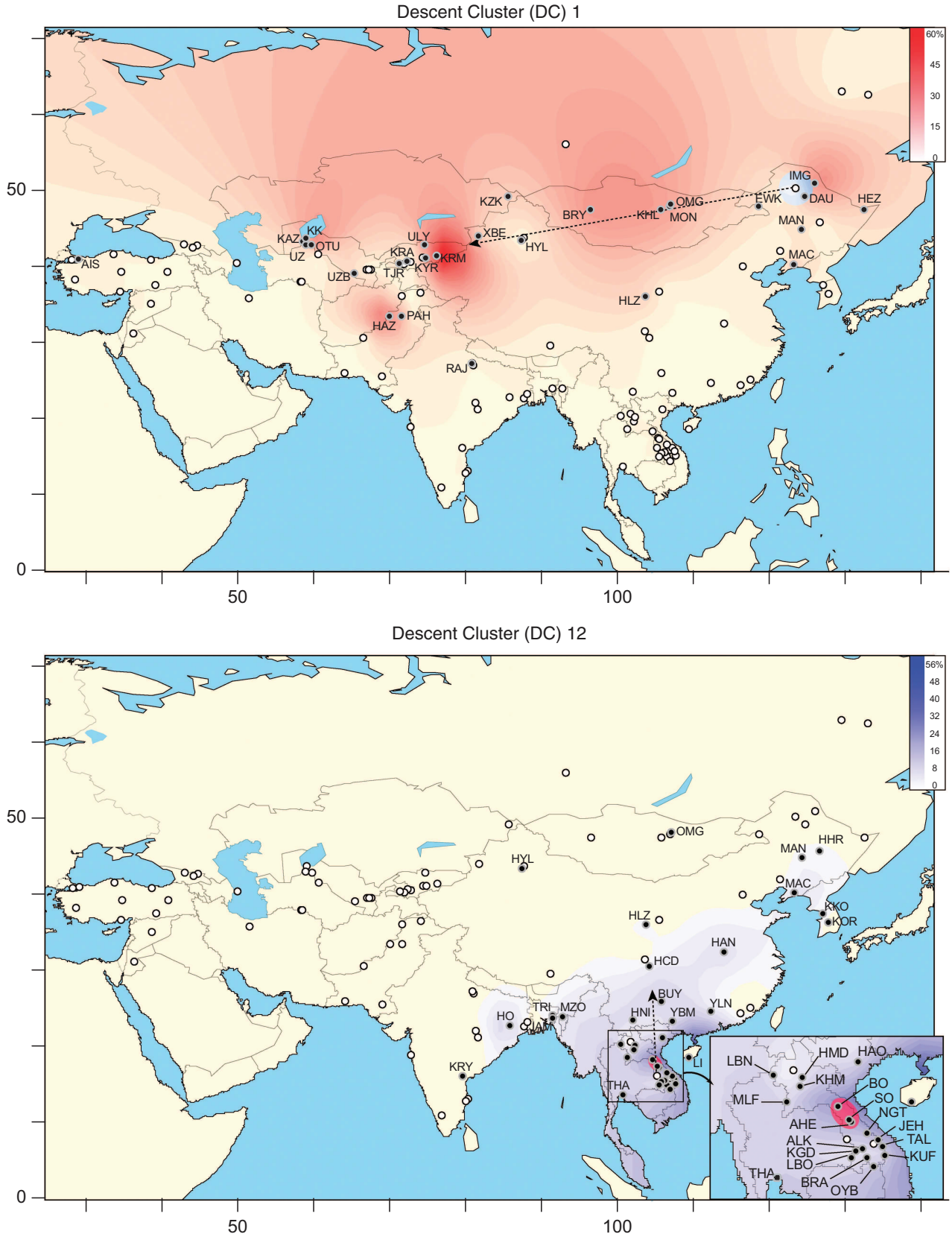


Figure 3 Geographical distribution of frequency and microsatellite variance for two DCs. The examples here are for DC1 and DC12 (see Supplementary Figure 2 for the remainder). Open circles indicate populations lacking the DC and filled circles indicate populations possessing it. Increasing frequency is indicated by darkening colour as indicated in the scale to the top right. The population with maximum microsatellite variance for the DC is indicated by a coloured circle and the direction of the expansion based on DC variance is indicated by arrows.

Table 2 TMRCA estimates for the 11 DCs and summary of geographical distributions and likely origins

DC	Core ht (n)	n in DC	Max steps from core	hg	TMRCA mean/y (95% CI)	Growth rate per generation (alpha)	Origin	Max variance (pop)	Period[a]	Cultural period
1[b]	71	142	3	C3 (xC3c)	951 (212–3826)	0.344	Silk Road (NE. China/Mongolia)	0.131 (Inner Mongolian)	H: 1060 CE	Khitan Empire-Liao Dynasty/Mongol Empire
2	47	271	6	R1a1	3284 (222–22922)	0.129	N.India/Tibet/Central Asia	0.321 (Uyghur)	P: 1300 BCE	Indus Valley Civilisation (mid-late Bronze Age)
3	43	95	4	L	929 (129–5219)	0.386	Fertile Crescent	0.154 (Iran)	H: 1100 CE	Islamic Golden Age; Pre Islam/Post Islam
4	41	109	3	O2b	1721 (184–10054)	0.230	Korea	0.179 (Korean)	H: 300 CE	Proto-Three Kingdoms / Several States Period
5	32	505	6	J	2730 (576–10373)	0.111	Fertile Crescent	0.315 (Anatolia NE)	P: 700 BCE	Neo-Assyrian, Neo-Babylonian Empire (late Bronze Age)
6	30	211	4	O3a3c	2887 (208–20202)	0.152	Tibet	0.279 (Tibet)	P: 900 BCE	Vedic Period - Iron Age
8[c]	28	59	3	C3/C3c	1325 (57–14756)	0.373	Silk Road (NE China/Mongolia)	0.104 (Oroqen)	H: 700 CE	Rouran / Uyghur Khaganate/ Khitan Empire-Liao Dynasty
10	24	26	3	N	1158 (116–8353)	0.323	Silk Road (NE China/Mongolia)	0.058 (Hezhe)	H: 850 CE	Rouran / Uyghur Khaganate/ Khitan Empire-Liao Dynasty
11	24	95	4	H1a	2296 (73–24529)	0.255	India/Central Asia	0.111 (Muria)	P: 300 BCE	Maurya Empire
12	24	152	6	O2a	3540 (369–19967)	0.117	Laos	0.243 (So)	P: 1500 BCE	Early-mid Bronze Age
14	23	335	6	O2a	4125 (326–29437)	0.09	Laos	0.324 (Alak)	P: 2100 BCE	Early-mid Bronze Age

[a]H, historical; P, protohistorical; *n*, number of chromosomes. [b]'Khan' haplotype cluster. [c]'Giocangga' haplotype cluster.

'Ancient' China,[37] and the Bronze Age/Iron Age in Levantine/Inner Asia.

Seeking common features among expansions

We performed a PCA on DC frequencies per population (Figure 5) to reveal the relative contributions of the 10 DCs and to test any potential associations between expansions. The two first axes of the PCA contribute, respectively, 27% and 21% of the total variance, with the main contributions of DCs 2, 5, 12 and 14 along axis 1, and those of DC1, 8 and 10 along axis 2. DCs 12 and 14 are systematically associated, as are DCs 2 and 5, and DCs 1, 8 and 10, suggesting shared characteristics between these three sets of expansions (Figure 5).

To test whether individuals included in these 10 expansions share any particular cultural features, such as language or subsistence methods, as is the case for previously published examples,[5,7] we performed a co-inertia test (Figure 6). This revealed the preferential association of DCs 12 and 14 with Sino-Tibetan and Austro-Asiatic languages and a 'multiple' mode of subsistence, of DC2s 2 and 5 with Indo-European languages and 'agricultural' subsistence, and of DCs 1, 8 and 10 with Altaic languages and pastoral nomadism (Figure 6). Genetic and cultural patterns (mode of subsistence and language) are significantly correlated (Rv = 0.47, *P*-value < 0.0001, $N = 95$), suggesting that several expansions share the same cultural characteristics.

The TMRCA estimates are consistent with six expansions beginning in protohistorical times (DCs 2, 5, 6, 11, 12 and 14), and four beginning in historical times (DCs 1, 3, 8 and 10; Table 2). DCs detected in the different periods differ (Figure 7) in regard to mode of subsistence ($\chi^2 = 345.59$; df = 2; $P < 0.0001$) and language ($\chi^2 = 211.67$; df = 4; $P < 0.0001$). This suggests a shift in population dynamics between the two periods, with an increase of successful Y-lineage expansions in Altai-speaking pastoral nomadic populations in more recent times. The historical-period expansions show the highest growth rates (Table 2).

Associating protohistorical and historical expansions with subsistence mode and language

Six of the 10 DCs analysed originated during the Bronze Age or earlier (2100–300 BCE), and correspond to 30% of the individuals in our data set. The 'origins' of these expansions lie within three main regions: South East Asia, including Laos (DCs 12 and 14), Tibet/East India (DCs 6 and 11) and Central Asia/Fertile Crescent (DC2 and DC5). The co-inertia analysis shows that these expansions are found mainly among agricultural populations, speaking Indo-European and Austro-Asiatic languages. DCs 5, 12 and 14 have their putative sources centred both on the Fertile Crescent and South East Asia—places known as centres of agricultural innovation;[38] these DCs could be, then, by-products of the civilisations that underwent a shift to agriculture during the Bronze Age (Figure 4). The origin of DC2 is located in Central Asia and its TMRCA estimate is 1300 BCE; this is coherent with the ancient DNA Y-chromosomes from the Middle Bronze Age (Andronovo) detected in DC2.[17]

Four expansions date to the historical period (700–1100 CE). The most recent (DC3; 1100 CE), originating in the Near East and extending to the South East Indian coast, is mainly detected in agricultural populations. It could be linked to the rapid expansion of Muslim power from the Middle East, across Central Asia and to the borders of China and India, after the establishment of a unified polity in the Arabian Peninsula by Muhammad in the 7th century and under the subsequent Caliphates.

The three other expansions (DCs 1, 8 and 10; 700–1060 CE) encompass 79% of individuals involved in historical expansions and

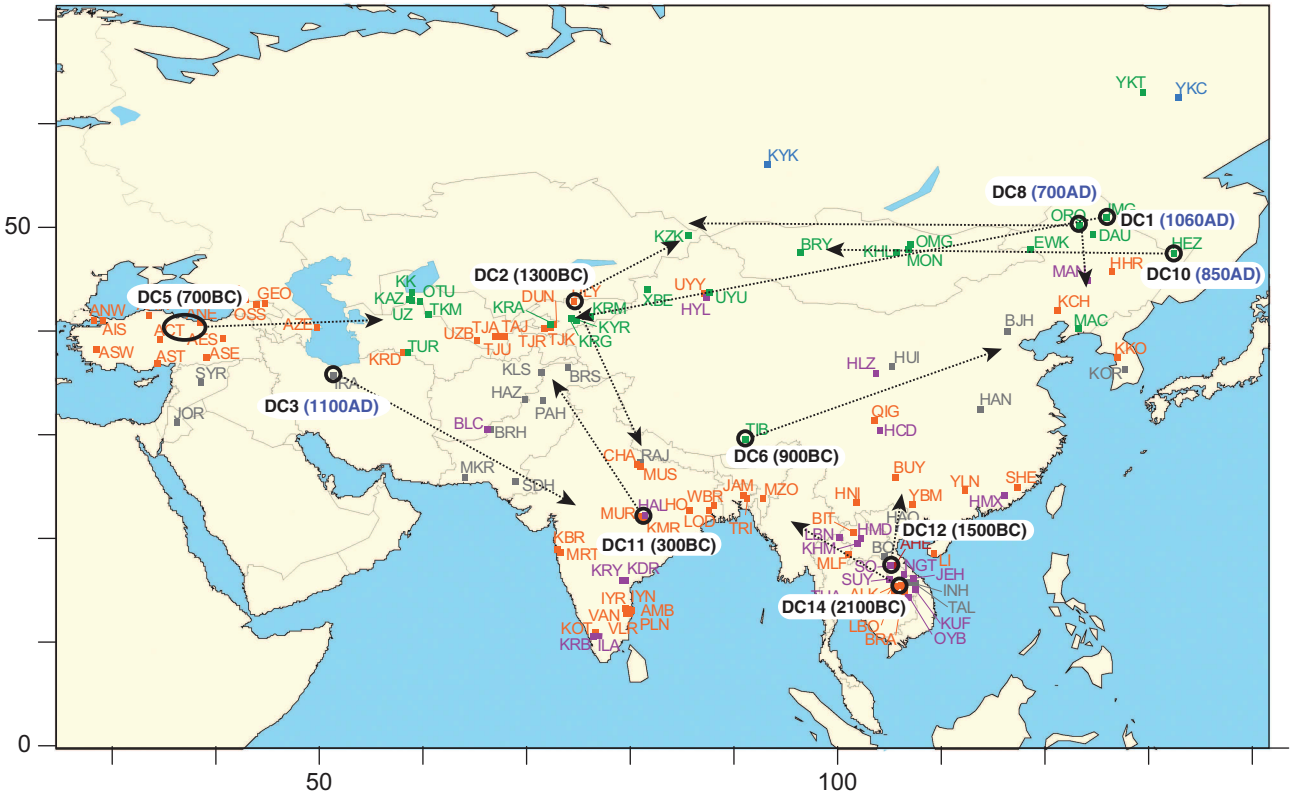


Figure 4 Schematic synthesis of DC features illustrated geographically. Black open circles indicate potential location origins and arrows indicate directionality. Mean TMRCA is given for each DC.

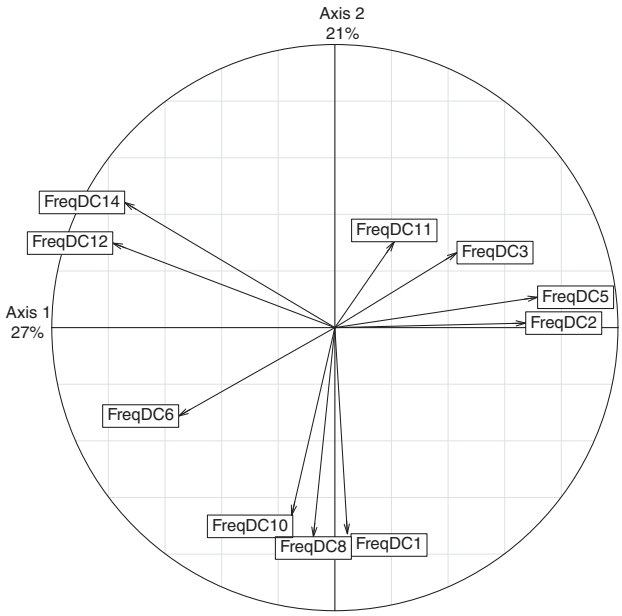


Figure 5 PCA showing the relative contribution of each DC on axes 1 and 2.

are almost exclusively detected in Altaic-speaking pastoral nomadic populations. The territory is large and follows the Silk Road corridor. The signal of expansion spreads from East to West (from Mongolia to the Caspian Sea), as DC1 has its source in Inner Mongolia (hgC3 [xC3c]), DC8 in the Oroqen (hgC3c) and DC10 in the Hezhe (hgN1). Given their different sources and associated haplogroups, these DCs are likely to represent three distinct expansions. These expansions show high variance values not only at their sources but also in other locations (Supplementary Table 3); this could be explained by a rapid migration of founders and parallel transmissions to generate diversity in these different places, all descending from a unique ancestor.

The two previously recognised Asian star-clusters ('Khan' and 'Giocangga'[5,7]) are identified here, and correspond to DCs 1 and 8, respectively. Their TMRCAs are estimated at 1090 CE for DC1, almost identical to the published estimate of ~1000 YA,[5] and at 700 CE for DC8, older than the published estimate of 590±340 YA.[7] The 'Khan' DC remains the most striking signal of an Asian expansion lineage, representing 2.7% of the entire data set, and the highest number of identical Y-chromosomes ($N=71$). Interestingly, the westward directions of expansions DC8 and DC10, their potential sources in northeast China, their geographic extents from China to Karakalpakia, and also the Altai-speaking populations associated with them, could also indicate involvement of the Imperial or elite lineages associated with the Khitan Empire. Abaoji, Emperor Taizu of Liao and the Great Khan of the Khitans (died 926 CE), who officially designated his eldest son as his successor, maintained a pattern of seasonal movements typical of pastoral societies, which could also explain the geographic distribution of these lineages.

DISCUSSION

Highly represented Y-lineages as a proxy for high reproductive success

Reproductive success, widely used as a measure of fitness, is the genetic contribution of one individual to future generations; it implies

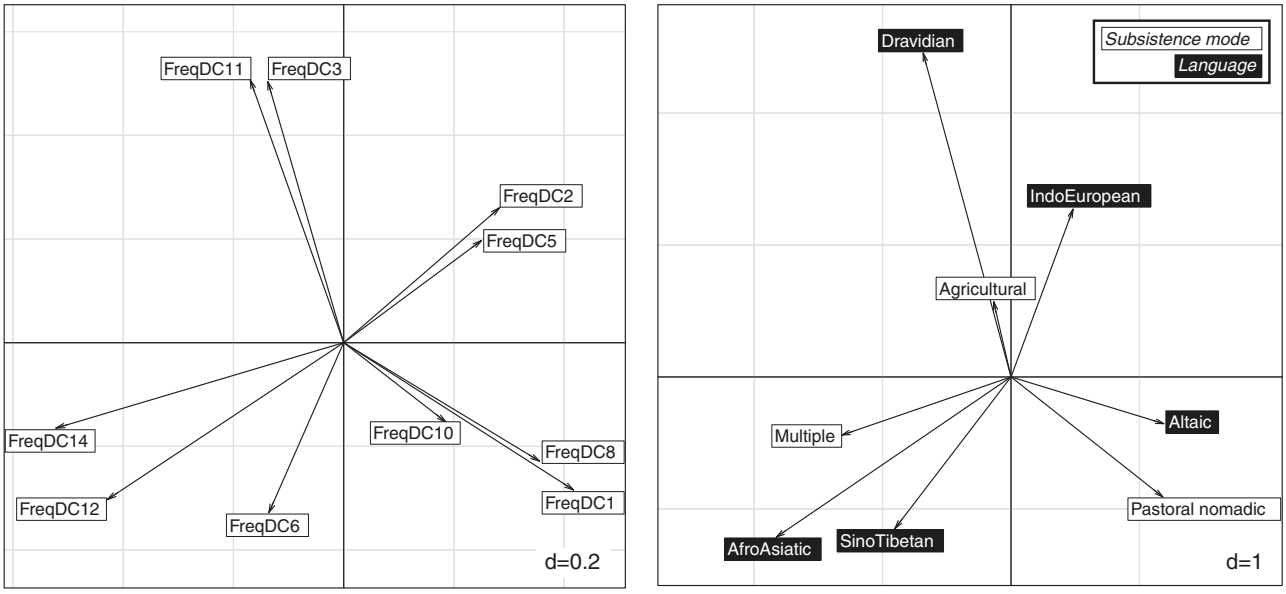


Figure 6 Co-inertia analyses showing preferential associations between DCs and cultural factors.

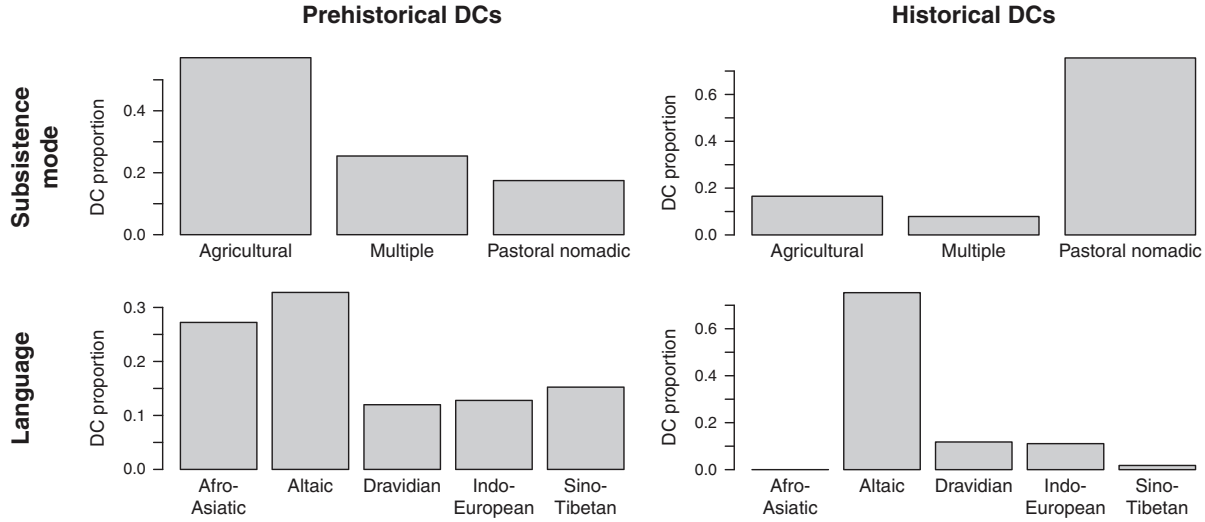


Figure 7 Proportion of DCs in protohistorical and historical periods in populations divided by different cultural factors.

that offspring are plentiful and that they survive. In humans, offspring number and survival can be strongly influenced by culture.[4] Examples of long-term high reproductive success have been revealed by unusually frequent Asian Y-chromosome haplotypes suggested to descend from two individuals, Genghis Khan[5] and Giocangga,[7] and to have spread in past generations via social selection (because of associated power and prestige) during historical times. In this study, we used a similar approach, analysing 5321 Y-microsatellite haplotypes belonging to 127 Asian populations to ask whether other examples of continued Y-lineage transmission are observed.

A favourable post-Bronze Age socio-political context for successful Y-lineage transmissions

We have detected 11 highly represented Y-chromosome lineages among the 5321 males analysed; altogether the DCs derived from 11 founding chromosomes represent a large proportion (38%) of the Y-chromosomes analysed, attesting to the importance of efficient Y-lineage transmission in the history of the continent. The presence of DCs with TMRCAs from 2100 BCE to 1100 CE suggests that the socio-cultural context, from the Bronze Age up to more recent historical periods, has been sufficiently favourable to ensure a continuous transmission of these different Y-lineages. Indeed, several admixture events have deeply affected the Asian gene pool during this period, involving migrant groups coming from North East and South East Asia.[39] It has been argued[40] that the development and impact of complex polities beginning with the Bronze Age ~2500–100 BCE in Central Asia were responsible for drastic changes in population structure, movements and organisation. These polities and early states emerged from local traditions of mobility and multiresource pastoralism, including agriculture and distributed hierarchy and administration,[41,42] with a royal or imperial form of leadership combining elements of kinship and political office.[42] Archaeological

evidence from the late Bronze and early Iron Ages (1400–400 BCE) confirms the existence of control hierarchies associated with wealth differences and socio-political power, even before the emergence of political entities.[42] The development of these polities was accompanied by the emergence of stratified societies, recognised elite lineages and differences in power and status between individuals.

High reproductive success is often associated with high social status, 'prestigious' men having higher intramarital fertility, lower offspring mortality[9] and access to a greater than average number of wives.[43] This suggests that the link between status and fitness detected in modern populations (eg, Nettle and Pollet,[44] Snyder *et al*[45] and Heyer *et al*[46]) also existed in ancient populations. For the TMRCA period from 2100 BCE to 1100 CE, the DCs are equally distributed between sedentary agriculturalist (36.8%) and pastoral nomadic populations (36.7%), indicating that uninterrupted transmission of Y-lineages over many generations could occur in both groups. Cultural transmission of fertility, reproductive success and prestige, seen in contemporaneous populations,[47] could also explain the persistence of these frequent Y-lineages over time. The high variance of DC proportions among populations indicates that the 'memes' responsible for cultural transmission can vary greatly among populations, even when populations share similar cultural characteristics.

A shift from sedentary agriculturalist to pastoral nomadic populations for successful Y-lineages

Our study highlights a difference in population composition between protohistorical and historical periods. In protohistorical times, DCs are detected mainly in agriculturalist but also in multiresource and pastoralist populations, whereas in historical times they are predominantly seen in pastoralist populations. This shift towards pastoralists only suggests a modification of social organisation in Asia and a higher probability to generate reproductively successful lineages in pastoral-derived societies and states. The development of complex polities began with the Bronze Age in Central Asia ~2500–100 BCE.[40] Agriculture arrived ~6000 BCE (Djeitun, Turkmenistan[48,49]) and dispersed during the Bronze Age (3000–2000 BCE) in coexistence with hunter–gatherer communities. The pastoral nomadic lifestyle, linked to horse domestication,[50] emerged in Central Asia ~5000 BCE and gained importance during the late Bronze Age. The pastoral nomadic tribes eventually came to dominate the Ponto–Caspian steppes during the first millennium BCE and entered historical records as Scythians (Herodotus) or Sakas (Persian accounts).[51]

The over-representation of historical-period DCs in pastoral nomadic Altaic-speaking populations is compatible with the development of new forms of political and social organisation in pastoralist economies. Among present-day pastoral nomads, patrilineal descent rules lead to a cultural transmission of reproductive success (Heyer *et al*, submitted) in male lineages.

New social systems and economic adaptations emerged after horse domestication. Horse-riding greatly enhanced both east–west connections and north–south trade between Siberia and southerly regions, and allowed new techniques of warfare, a key element explaining the successes of mobile pastoralists in their conflicts with more sedentary societies.[41] A series of expansive polities emerged in Inner Asia by 200 BCE: 15 steppe polities beginning with Xiongnu (Khunnu) ~200 BCE and concluding with the Zunghars in the mid-18th century have been described, including the Mongol and the Qing empires/dynasties.[42,52] The DCs detected along the Silk Road corridor originated from 700 to 1060 CE and may be associated with dynasties that coexisted with five major empires that dominated the Ponto–Caspian steppes from the 7th to the 13th centuries: the Khitan (Great Liao), Tangut Xia, Jurchin, Kara-Khitan and the Mongol Empires.[42] In order to identify the prestigious DC founders among potential candidates, the expansion strategies of these different dynasties, and also the mechanisms of co-inheritance of Y-lineages and prestige via marriage rules should be carefully investigated. As an example, sororal polygyny (in which a man can marry two or more sisters) followed by a shift to the Han Chinese system of taking one wife and one or more concubines, as occurred among the Liao elite throughout the length of the Liao dynasty, are likely to promote the efficient transmission of prestigious lineages among males.

Further investigations including ancient DNA approaches, and next-generation sequencing of modern expansion lineage Y-chromosomes could be undertaken to refine the history of the expansion lineages we have observed, and perhaps to identify the prestigious and powerful pastoralist founders associated with them.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all DNA donors; Matt Hurles, Emma Parkin and Denise Carvalho-Silva for primers; and Philippe Mennencier for help with language classification. This work was supported by the Centre National de la Recherche Scientifique (CNRS) ATIP programme (to EH), by the CNRS interdisciplinary programme 'Origines de l'Homme du Language et des Langues', and by the European Regional Development Fund, the Conseil Régional Midi-Pyrénées and the PRES (Pôle de Recherche et d'Enseignement Supérieur) Toulouse (to PB). MAJ was supported by a Wellcome Trust Senior Research Fellowship in Basic Biomedical Science (grant nos. 057559 and 087576), and PB by the Wellcome Trust and the CNRS.

1 Zietsch BP, Kuja-Halkola R, Walum H, Verweij KJ: Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. *Proc Natl Acad Sci USA* 2014; **111**: 1032–1036.
2 Strassmann BI, Gillespie B: How to measure reproductive success? *Am J Hum Biol* 2003; **15**: 361–369.
3 Austerlitz F, Heyer E: Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proc Natl Acad Sci USA* 1998; **95**: 15140–15144.
4 Heyer E, Sibert A, Austerlitz F: Cultural transmission of fitness: genes take the fast lane. *Trends Genet* 2005; **21**: 234–239.
5 Zerjal T, Xue Y, Bertorelle G *et al*: The genetic legacy of the Mongols. *Am J Hum Genet* 2003; **72**: 717–721.
6 King TE, Jobling MA: Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. *Mol Biol Evol* 2009; **26**: 1093–1102.
7 Xue Y, Zerjal T, Bao W *et al*: Recent spread of a Y-chromosomal lineage in northern China and Mongolia. *Am J Hum Genet* 2005; **77**: 1112–1116.
8 Moore LT, McEvoy B, Cape E, Simms K, Bradley DG: A Y-chromosome signature of hegemony in Gaelic ireland. *Am J Hum Genet* 2006; **78**: 334–338.
9 von Rueden C, Gurven M, Kaplan H: Why do men seek status? Fitness payoffs to dominance and prestige. *Proc Biol Sci* 2011; **278**: 2223–2232.
10 Lansing JS, Watkins JC, Hallmark B *et al*: Male dominance rarely skews the frequency distribution of Y chromosome haplotypes in human populations. *Proc Natl Acad Sci USA* 2008; **105**: 11645–11650.
11 Clarke AL, Low BS: Testing evolutionary hypotheses with demographic data. *Popul Dev Rev* 2001; **27**: 633–660.
12 Heyer E, Balaresque P, Jobling MA *et al*: Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genet* 2009; **10**: 49.
13 Cai X, Qin Z, Wen B *et al*: Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 2011; **6**: e24282.
14 Cinnioglu C, King R, Kivisild T *et al*: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
15 Crubezy E, Amory S, Keyser C *et al*: Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC Evol Biol* 2010; **10**: 25.
16 El-Sibai M, Platt DE, Haber M *et al*: Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Ann Hum Genet* 2009; **73**: 568–581.

17 Keyser C, Bouakaze C, Crubezy E *et al*: Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet* 2009; **126**: 395–410.
18 Kim SH, Kim KC, Shin DJ *et al*: High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. *Invest Genet* 2011; **2**: 10.
19 Sengupta S, Zhivotovsky LA, King R *et al*: Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 2006; **78**: 202–221.
20 Xue Y, Zerjal T, Bao W *et al*: Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 2006; **172**: 2431–2439.
21 Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C: A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* 2002; **71**: 466–482.
22 Bosch E, Calafell F, González-Neira A *et al*: Male and female lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 2006; **70**: 459–487.
23 Hurles ME, Sykes BC, Jobling MA, Forster P: The dual origin of the Malagasy in island southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am J Hum Genet* 2005; **76**: 894–901.
24 Paracchini S, Arredi B, Chalk R, Tyler-Smith C: Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucl Acids Res* 2002; **30**: e27.
25 Karafet TM, Mendez FL, Meilerman M, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
26 Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF: A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int* 2002; **129**: 10–24.
27 Parkin EJ, Kraayenbrink T, van Driem GL, Tshering K, de Knijff P, Jobling MA: 26-locus Y-STR typing in a Bhutanese population sample. *Forensic Sci Int* 2006; **161**: 1–7.
28 Balaresque P, Bowden GR, Adams SM *et al*: A predominantly Neolithic origin for European paternal lineages. *PLoS Biol* 2010; **8**: e1000285.
29 Kayser M, Krawczak M, Excoffier L *et al*: An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 2001; **68**: 990–1018.
30 Wilson IJ, Balding DJ: Genealogical inference from microsatellite data. *Genetics* 1998; **150**: 499–510.
31 Kayser M, Roewer L, Hedman M *et al*: Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; **66**: 1580–1588.
32 Burgarella C, Navascues M: Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet* 2011; **19**: 70–75.
33 Fenner JN: Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 2005; **128**: 415–423.
34 R Core Team *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. 2014.
35 Dray S, Chessel D, Thioulouse J: Co-inertia analysis and the linking of ecological tables. *Ecology* 2003; **84**: 3078–3089.
36 Dray S, Dufour AB: The ade4 package: implementing the duality diagram for ecologists. *J Statist Software* 2007; **22**: 1–20.
37 Loewe M, Shaughnessy EL: Introduction; in Loewe M, Shaughnessy EL (eds): *The Cambridge History of Ancient China: from the Origins of Civilization to 221 BC*. Cambridge: Cambridge University Press, 1999, pp 1–36.
38 Jobling MA, Hollox EJ, Hurles ME, Kivisild T, Tyler-Smith C: Agricultural expansions; in *Human Evolutionary Genetics*. New York and London: Garland Science, 2013, 2nd edn, pp 362–407.
39 Hellenthal G, Busby GB, Band G *et al*: A genetic atlas of human admixture history. *Science* 2014; **343**: 747–751.
40 Frachetti MD: Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr Anthropol* 2012; **53**: 2–38.
41 Rogers JD: The contingencies of state formation in Eastern Inner Asia. *Asian Perspect* 2007; **46**: 249–274.
42 Rogers JD: Inner Asian States and Empires: theories and synthesis. *J Archaeol Res* 2012; **20**: 205–256.
43 Schubert G, Stoneking CJ, Arandjelovic M *et al*: Male-mediated gene flow in patrilocal primates. *PLoS One* 2011; **6**: e21514.
44 Nettle D, Pollet TV: Natural selection on male wealth in humans. *Am Nat* 2008; **172**: 658–666.
45 Snyder JK, Kirkpatrick LA, Barrett HC: The dominance dilemma: Do women really prefer dominant mates? *Pers Relat* 2008; **15**: 425–444.
46 Heyer E, Chaix R, Pavard S, Austerlitz F: Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 2012; **21**: 597–612.
47 Kolk M: Multigenerational transmission of family size in contemporary Sweden. *Popul Stud* 2014; **68**: 111–129.
48 Dani AH, Masson VM: *History of Civilizations of Central Asia, Volume 1*. Paris: UNESCO Publishing, 1992.
49 Harris DR, Gosden C, Charles MP: Jeitun: Recent excavations at an Early Neolithic site in southern Turkmenistan. *Proc Prehist Soc* 1996; **62**: 423–442.
50 Warmuth V, Eriksson A, Bower MA *et al*: Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc Natl Acad Sci USA* 2012; **109**: 8202–8206.
51 Harmatta J: *History of Civilizations of Central Asia, Volume II. The Development of Sedentary and Nomadic Civilizations: 700 B.C. to A.D. 250*. Paris: UNESCO Publishing, 1994.
52 Gumilyov LN: *Ancient Turks*. Moscow: Institute of Ethnology and Anthropology of the Academy of Sciences of USSR, 1967.