ORIGINAL ARTICLE

Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians

Wei-Hua Shou[1], En-Fa Qiao[1], Chuan-Yu Wei[1], Yong-Li Dong[1], Si-Jie Tan[1], Hong Shi[2], Wen-Ru Tang[1] and Chun-Jie Xiao[1]

Northwest China is closely adjacent to Central Asia, an intermediate region of the Eurasian continent. Moreover, the Silk Road through the northwest of China once had a vital role in the east–west intercommunications. Nevertheless, little has been known about the genetic makeup of populations in this region. We collected 503 male samples from 14 ethnic groups in the northwest of China, and surveyed 29 Y-chromosomal biallelic markers and 8 short tandem repeats (STRs) loci to reconstruct the paternal architecture. Our results illustrated obvious genetic difference among these ethnic groups, and in general their genetic background is more similar with Central Asians than with East Asians. The ancestors of present northwestern populations were the admixture of early East Asians peopling northwestward and later Central Asians immigrating eastward. This population mixture was dated to occur within the past 10 000 years. The J2-M172 lineages likely entered China during the eastward migration of Central Asians. The influence from West Eurasia through gene flows on the extant ethnic groups in Northwest China was relatively weak.

INTRODUCTION

Patterns of Y-chromosome diversity provide a unique perspective into aspects of the origins and composition of populations.[1,2] Its restricted paternal inheritance, smaller effective population size and specific clustering of Y variants qualify Y chromosome as an indispensable tool to provide substantial genetic evidences.[3,4] In addition, increasing and stable binary markers were being discovered in the past decade, and consequently, the topology and nomenclatures of Y-genealogical tree were being established successively.[5] These modifications empower us to discriminate haplogroups with higher resolution and less ambiguity. The comprehensive approach, that is, combining the high-resolution Y-single-nucleotide polymorphisms and the more rapidly evolving microsatellite markers, would shed more light on the origins and the complex history of populations.[6]

Central Asia serves as a geographic conjunction between East Asia and West Asia and East Europe, lying between Siberia in the north and South Asia subcontinent in the south. Within the non-African context, Central Asia shows a high level of both genetic and ethnic diversity,

indicating that the settlement of this region was a complex process. Two competing hypotheses have been raised concerning the origin of Central Asians. One hypothesis suggests that Central Asians could represent an early incubator of Eurasian variation, whereas the other proposes that the current rich genetic diversity of Central Asians could result from recent admixture between western and eastern Eurasian populations. Y-chromosomal data have been interpreted as the indication that Central Asia was a major source of population migration events.[7,8] However, studies on mitochondrial DNA found that considerable western and eastern Eurasian haplogroups overlapped in present-day Central Asian populations. European and East Asian mitochondrial DNA lineages could be clearly demarcated, suggesting recent peopling.[9–11]

Northwest China closely neighbors Central Asia, and particularly Xinjiang Uygur Autonomous Region starts to extend into Central Asia. In addition, the Silk Road through the northwest of China, a trans-Eurasian trade route established in the second century BC, had a vital role in the east–west multifarious intercommunications, and

[1]Key Laboratory of Bioresources Conservation and Utilization and Human Genetics Center, Yunnan University, Kunming, PR China and [2]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology and Kunming Primate Research Centre, Chinese Academy of Sciences, Kunming, PR China
Correspondence: Dr C-J Xiao, Key Laboratory of Bioresources Conservation and Utilization and Human Genetics Center, Yunnan University, 2 North Cuihu Road, Kunming, Yunnan 650091, PR China.
E-mail: cjxiao@public.km.yn.cn

thus attracted more attention to human migration in this region. Having different religious faith, cultures and life customs, a number of ethnic groups inhabit in northwestern China, and presumably had experienced complicated history.[12,13] Although some of them settled in this region since a long time, several groups were formed recently. The abundance of human genetic resources in this region and the lack of knowledge about population structure, especially the absence of a detailed Y-chromosomal dissection, motivated our endeavors to dissect paternally genetic architecture of northwestern populations using Y-chromosomal variation data. In addition, we anticipated providing some valuable information about the ethnogenesis of Central Asians plus a combination of archaeology- and language-based research results by others.

To elucidate the origin and the population-forming processes of the human populations in Northwest China and to explore their paternal genetic structure, we performed detailed analyses of the extant local ethnic groups.

MATERIALS AND METHODS

Population samples

A total of 503 male Y chromosomes from geographically and linguistically representative 14 ethnic groups were assayed. All individuals for this study are unrelated and belong to the same ethnic groups for at least three generations, whose blood samples were collected with appropriate ethical approval and informed consent. Information about ethnic groups in the study is listed in Table 1, and sampling locations are spotted in Figure 1. Genomic DNA was extracted from whole blood using normal phenol−chloroform method and was stored in 10 mM Tris-1 mM EDTA (TE solution) at −80 °C for further test.

Y Haplogrouping and terminology

The selection of Y-chromosome single-nucleotide polymorphism markers and the nomenclature of haplogroup on Y genealogical tree referred to the International Society of Genetic Genealogy (Y-DNA Haplogroup Tree 2007, http://isogg.org/tree/ISOGG_YDNA_SNP_Index07.html), and the previous studies.[2,14,15] A set of 29 informative binary polymorphic sites was included in our survey, defining Y haplogroups from C to R (Figure 2). Detailed

Table 1 Population genotyped in this study

Population	Census size	Language affiliation	Sampling location[a]
Tu	241 000	Altaic Family, Mongolian Group	1
Xibo	40 000[b]	Altaic Family, Manchu-Tungusic Group	2
Mongolian	5 810 000	Altaic Family, Mongolian Group	3
Tataer	4900	Altaic Family, Turkic Group	4
Uygur	9 870 000	Altaic Family, Turkic Group	5
Yugu	14 000	Altaic Family, Turkic or Mongolian Group	6
Kirghiz	160 000	Altaic Family, Turkic Group	7
Kazakh	1 250 000	Altaic Family, Turkic Group	8
Tajike	42 000	Indo-European Family, Iranian Group	9
Russ	16 000	Indo-European Family, Slavic Group	10
Dongxiang	514 000	Altaic Family, Mongolian Group	11
Ozbek	14 000	Altaic Family, Turkic Group	12
Bao'an	16 000	Altaic Family, Mongolian Group	13
Salar	104 000	Altaic Family, Turkic Group	14

[a]The number codes are spotted in Figure 1.
[b]The population size in Xinjiang Uygur Autonomous Region.
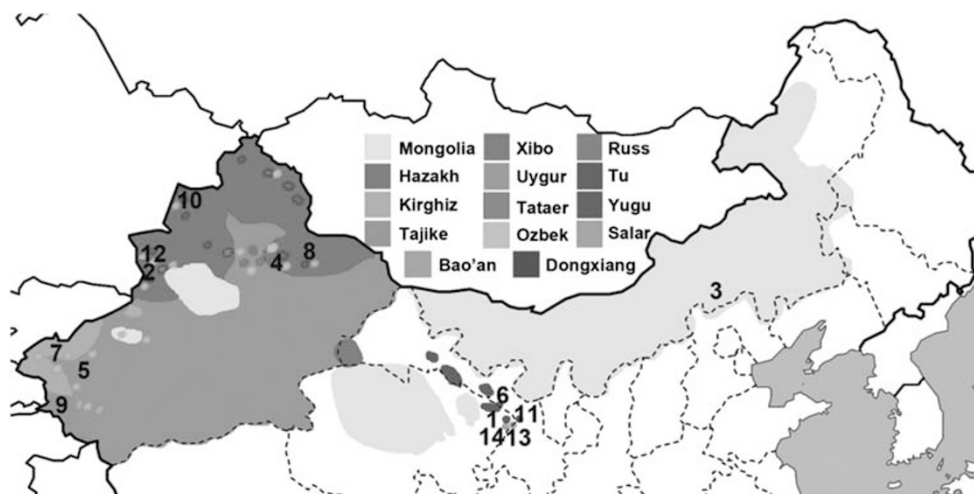
Figure 1 Map of sampling locations of 14 ethnic groups in this study and the geographic distribution of these ethnic nationalities. Areas pigmented with colors represent the distribution of nationalities, and the number codes (consistent with those in Table 1) denote the sampling sites. A full color version of this figure is available at the *Journal of Human Genetics* journal online.
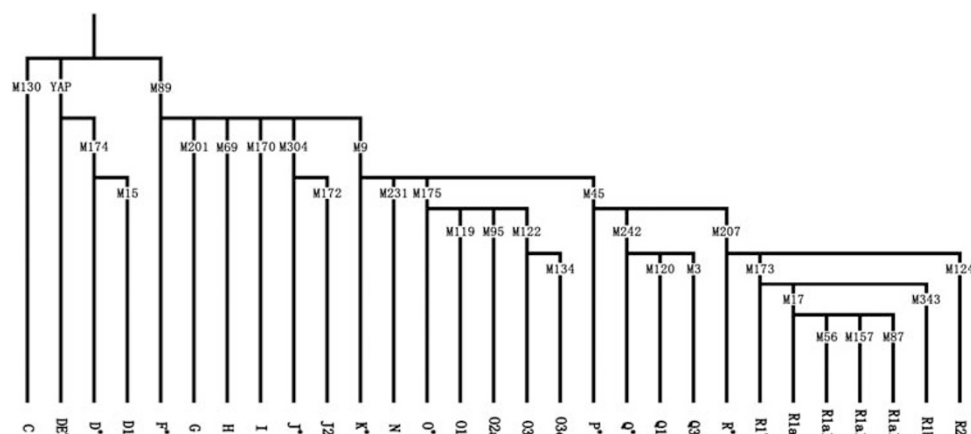
Figure 2 The phylogenetic relationships and nomenclature of Y-chromosome binary haplogroups surveyed in this study.

discriminating of cladistic nodes internal to P-M45 was performed using more derived markers (M3, M17, M56, M87, M120, M124, M157, M173, M207, M242 and M343) given the high frequent occurrence of M45 in Central Asians. We adopted the hierarchical typing strategy[3] to screen these markers using PCR-restriction fragment length polymorphism or sequencing methods. M175, a 5-bp indel polymorphism, was genotyped by GENESCAN on ABI 377genetic analyzer (Applied Biosystems, Foster City, CA, USA). The purified PCR products were sequenced on ABI 377genetic analyzer, using ABI PRISM BigDye Terminator V3.1 Sequencing Kit.

Y-STR typing
A total of eight multiallelic short tandem repeat (STR) loci were genotyped for subjects of haplogroup R1 and J, including two trinucleotide-repeat polymorphisms (DYS388 and DYS392), and six tetranucleotide-repeat polymorphisms (DYS19, DYS389I, DYS389II, DYS390, DYS391 and DYS392). These eight loci were analyzed by PCR using published primers described by Kayser et al.[16] and the STR website (http://www.yhrd.org). The twin polymorphisms, DYS389I and DYS389II, were amplified with single forward primer labeled by fluorescent dye and two respective reverse primers. PCR products were directly run on ABI 377 sequencer, with ABI GS500 TAMRA as the internal lane standard. The GENESCAN and GENOTYPER software packages were used to collect the data and to discriminate allele counting. Y-STRs alleles were called according to the number of repeat units they carried.

Data analysis
The frequencies of Y haplogroups were computed. The multidimensional scaling (MDS) analysis was conducted using SPSS 11.5 software after the generation of F_{st} genetic distance matrix by Arlequin 3.0.[17] Results of the MDS were presented by the two-dimensional plots. The genetic structure of populations was dissected by the analysis of molecular variance (AMOVA) approach,[18] still using Arlequin software. Additional data of Daur, Ewenki, Hezhe, Manchu, Oroqen, Korean, Sichuan Han (SC Han), Guangdong Han (GD Han), Yao, Buyi, Xinjiang Han (XJ Han) and Gansu Han (GS Han) from Xue et al.,[19] Tibetan from Gayden et al.,[20] and Turkmen, Kyrgyz, Tajik, Uzbek and Kazak (distinguished from Kirghiz, Tajike, Ozbek and Kazakh in the present study) from Wells et al.[7] were included in the MDS and AMOVA analysis as referential populations.

Median-joining (MJ) networks[21] of Y-STR haplotypes were constructed for R1a1-M17 and J2-M172 using NETWORK 4.2.0.1 (http://www.fluxus-engineering.com). Epsilon was set as zero. Published Y-STR data for 121 M17-carrying Russians from South Russia region[22] and 236 western Eurasians belonging to J2-M172[23] were also included, in the light of the putative origins of these two haplogroups. For the integration of data between reported and present, MJ networks for R1a1-M17 haplotypes were based on seven STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393), and for

J2-M17 based on six STR loci (DYS19, DYS388, DYS389I, DYS390, DYS391 and DYS392). Y-STR data are listed in Supplementary data file.

The time to the most recent common ancestor was estimated based on Y-chromosome microsatellite variation using the linear expansion,[24,25] BATWING (Bayesian analysis of trees with internal node generation)[26] and single-nucleotide polymorphism-STR coalescence methods.[27] All three methods assumed an average Y-STR mutation rate of 0.00069 per locus per 25 years.[28] The model of exponential growth from initial constant population size at time beta was set for BATWING analysis, and prior distributions in calculations were specified for alpha as a gamma (1.01, 1), for beta as a uniform (0, 15) and for N as a gamma (1, 0.0001). The upper bound for the coalescent time was determined with the assumption $V_0=0$,[29] and the lower bound with the assumption $V_0=V_a$ (V_a is the within-population variance in ancestral populations[30]).

RESULTS
Diversity of Y-single-nucleotide polymorphism markers
A total of 24 haplogroups was detected, including paragroup C, D, F*, G, H, I, J, K*, N, O, P*, P and R among 14 ethnic groups from Northwest China. However, polymorphisms on four markers (Q3-M3, R1a1a-M56, R1a1b-M157 and R1a1c-M87) were not observed. The frequency distribution of haplogroups is shown in Table 2.

The patrilineal gene pool of populations in this area was indicative of high haplogroup diversity (0.7602 ± 0.0546 on average) in general, which was similar to the previously observed value in Central Asians,[7] and being highest in Bao'an population (0.8946 ± 0.0305). As to the distribution of haplogroups, paragroup O and P were the two most predominant lineages. The clade R1a1-M17 within subhaplogroup R had an extensive distribution among populations, and at significant frequency in several groups (68.9% in Kirghiz population, 60.6% in Tataer, 54.3% in Dongxiang, 45.2% in Tajike and 40% in Salar), but three derivatives of R1a1 (R1a1a-M56, R1a1b-M157 and R1a1c-M87) were not observed. Among the individual clades within paragroup O, O1-M119, O2a-M95 and O3-M122 were detected in our samples, and the major constituent of haplogroup O was O3-M122 lineages. Intriguingly, 47% males in Russ population belonged to O3-M122. N-M231 occurred at a moderate frequency in multiple ethnic groups. Haplogroup J was only observed in several groups from the extreme northwest of China. Moreover, 32 individuals out of 37 J-related samples were allocated into J2-M172 (30.4% occurred in Ozbek). G-M201, H-M69 and I-M170 sporadically occurred in several populations. Individuals carrying I-M170 were mostly from Tataer population. An extensive, but uneven, distribution of haplogroup

Table 2 Haplogroup distribution and Y-chromosome diversity in 14 northwestern populations

Population	Tu	Xibo	Mongolian	Tataer	Uygur	Yugu	Kirghiz	Kazakh	Tajike	Russ	Dongxiang	Ozbek	Bao'an	Salar	Total
No. of subjects	50	32	50	33	50	32	45	41	31	19	35	23	27	35	503
C-M130	14	4	20	0	2	1	4	24	1	2	1	0	0	0	73
D*-M174	3	0	1	0	1	14	0	1	0	1	0	1	4	0	26
D1-M15	1	0	1	0	0	0	0	0	0	0	0	0	0	0	2
F*-M89	1	0	0	0	0	2	0	2	1	0	0	0	0	0	6
G-M201	0	0	0	0	0	0	0	1	2	0	0	2	0	0	5
H-M69	0	0	0	0	4	0	0	0	0	0	0	0	0	0	4
I-M170	0	0	0	11	0	0	0	0	0	1	0	0	0	0	12
J*-M304	0	1	0	0	0	0	0	1	0	1	1	1	0	0	5
J2-M172	1	0	0	0	17	2	0	0	5	0	0	7	0	0	32
K*-M9	1	3	2	0	9	2	1	1	1	0	0	6	6	0	32
N-M231	5	5	3	0	1	2	2	4	0	0	2	2	1	4	31
O*-M175	3	7	2	0	3	1	2	0	0	2	4	0	2	3	29
O1-M119	3	4	0	0	0	0	0	0	0	1	0	0	0	1	9
O2a-M95	1	0	1	0	0	0	0	0	0	1	0	0	5	1	9
O3-M122	12	2	8	0	2	4	0	0	0	7	0	0	3	7	45
O3e-M134	1	5	8	2	0	1	2	3	0	2	0	1	1	0	26
P*-M45	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2
Q*-M242	1	0	0	0	0	0	1	0	0	0	8	0	0	0	10
Q1-M120	0	0	1	0	2	0	0	0	0	1	0	0	0	0	4
R*-M207	0	0	0	0	1	0	0	1	5	0	0	0	0	0	7
R1*-M173	0	0	0	0	0	3	1	0	0	0	0	0	0	0	4
R1a1-M17	1	1	2	20	8	0	31	2	14	0	19	3	2	14	117
R1b-M343	2	0	1	0	0	0	0	1	0	0	0	0	1	5	10
R2-M124	0	0	0	0	0	0	0	0	1	0	0	0	2	0	3
Haplogroup diversity	0.856 ± 0.032	0.885 ± 0.024	0.794 ± 0.045	0.534 ± 0.061	0.827 ± 0.035	0.79 ± 0.066	0.521 ± 0.089	0.65 ± 0.082	0.759 ± 0.065	0.86 ± 0.07	0.654 ± 0.07	0.838 ± 0.048	0.895 ± 0.031	0.78 ± 0.048	

C-M130 was observed among the majority of ethnic groups under investigation, occurring in Kazakhs and Mongolians at a climax frequency of 58.5 and 40%, respectively. Haplogroup R2-M124 is infrequent, but informative in Central Asia, Anatolia, Kurdistan, and particularly in Pakistan and India.[7,14,31–33] In this study, there were merely three representatives of R2-M124 observed (one from Kazakh and the other two from Bao'an). All of the YAP[+] associated individuals were designated into lineage D for the observation of the diagnostic marker M174. This may be to a large extent attributed to the long-term influence from Tibetan with the notable prevalence of D lineages.[34,35] The phylogeography of D was rather irregular in this region, and even significant discrepancy was among geographically neighboring groups. This pattern was concordant with the relic distribution of D-M174 delineated in a latest study by Shi et al.,[36] which proposed M174 lineages were representative of the ancient northward colonizers into Asia.

Clustering analysis by MDS plot

Figure 3 presents MDS plots based on F_{st} genetic distance matrix calculated by the frequency of haplogroups. The MDS plot of 32 populations (Figure 3a) showed the between-population differentiating and clustering pattern, reflecting by and large an east–west cline. Nearly all Central Asians clustered together on the left part of the plot. In contrast, the right part was occupied by East Asians (Southeast Asians on the upper side and Northeast Asians on the lower side). A part of 14 populations in this study clustered either with Central Asians or with East Asians. Kazakh, Mongolian, Xibo, Tu, Russ and Yugu were closer to East Asians, and the other nine populations loosely clustered with Central Asians. In general, the paternal genetic structure of ethnic groups in northwestern China was perceptibly similar to Central Asians. The two spots of northwestern Han Chinese (XJ Han and GS Han) were more adjacent to Northeastern Asians on the plot, obviously separated from Central Asians.

When only the haplogroup frequency data from the 14 populations surveyed in this study were subjected to MDS clustering analysis (Figure 3b), little coherence with either language affinity or geographic proximity was evident. To extend this observation, Mantel test was conducted by Arlequin to test correlation between genetic and geographic distances using pairwise F_{st} values and pairwise geographic distances (the geographic distances were drawn from the Great Circle Distance Calculator program [http://www.marinewaypoints.com/learn/greatcircle.shtml]). No correlation was observed ($r=-0.063$, $P=0.675$), further indicating that, in Northwest China, the paternal genetic pattern is inconsistent with isolation-by-distance model, which was commonly used to interpret the difference of genetic structure of populations with geographic scales.[37]

AMOVA dissecting population structure

The rationale behind AMOVA is that, among different grouping hypotheses, a grouping level of populations that accurately reflects their genetic architectures should allocate a higher proportion of the genetic variance between groups, and a lower proportion among populations within groups. Table 3 presents variance components and P-values at 10 grouping levels. Grouping levels 1–3 signified rich divergences among ethnic populations in this study, even among populations with close language affinity and geographic proximity, because of great percentages of the genetic variance among populations within groups. This point was exacerbated by negative values of the
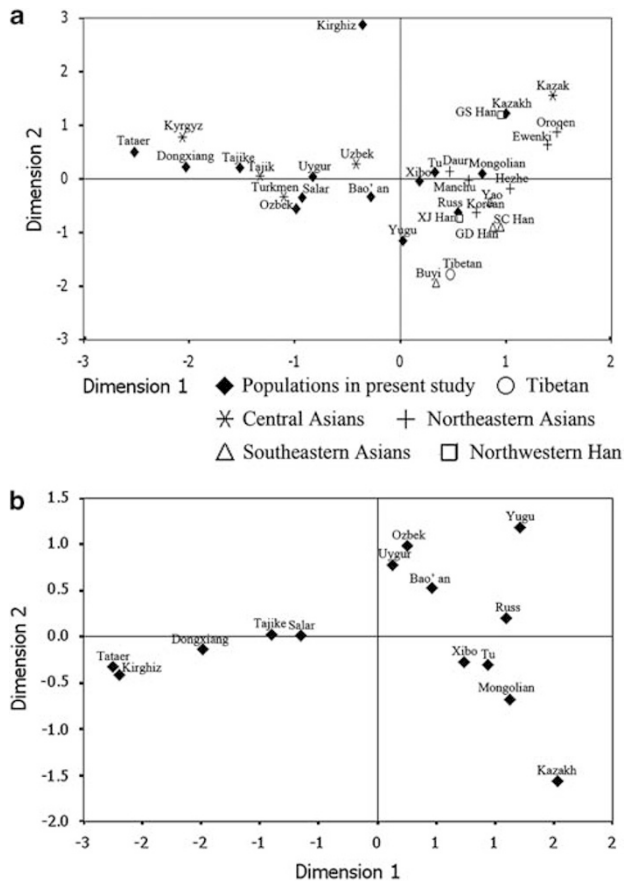
Figure 3 Two-dimensional plots of multidimensional scaling analysis based on F_{st} genetic distances: (**a**) expanded data including referential populations; and (**b**) the data of 14 populations in this study.

genetic variance among groups. Levels 4 and 5 were grouped according to the clustering result on MDS plot (Figure 3b) instead of the affiliation of geography and language. The high proportion of the variance among groups (nearly 14%; $P < 10^{-5}$) suggested the genetic intricacy in these populations, and simultaneously confirmed the validity of MDS analysis. After the referential populations data were incorporated, the variance among groups of 3.05 and 3.57% for grouping levels 6 and 7, respectively, suggested the closer similarity in genetic profile of northwestern Han Chinese with northeastern Asians than with northwestern populations, and furthermore the appreciable differentiation between northwestern populations and Northeastern Asians, Southeastern Asians as well as Central Asians. However, northwestern populations are genetically more similar to Central Asians than East Asians in general, which was corroborated by the significant discrepancy of variance among groups between level 8 (4.83%) and level 9 (negative value). Grouping level 9 according to the MDS plot (Figure 3a) obtained 9.03% variance among groups, which postulated the genetic contributions from East Asians and Central Asians to the gene pool of northwestern populations.

MJ networks and age estimation

MJ networks indicated that neither haplogroup R1a1 nor J2 showed Y-STR haplotypes sharing between northwestern Chinese and the quoted populations (Figure 4). This clear differentiation in Y-STR motifs suggested that these two lineages in Northwest China hardly came directly from their originating resources. MJ networks showed high haplotype diversity within R1a1-M17 and no obvious modal haplotype. The M17 lineages in Northwest China might originate from multiple founders instead of a single early dispersal event, given its prevalence in Central Asia, and then experience rapid demographic expansion and differentiation. Yet, M172 seemingly suffered from strong bottleneck effects after a single early immigrant event.

Table 3 AMOVA results

		Among groups		Among populations within groups		Within populations	
Grouping levels	No. of groups	Variance (%)	P-value	Variance (%)	P-value	Variance (%)	P-value
Data of this study							
1	1			17.01	$<10^{-5}$	82.99	
2	2	−0.17	>0.05	17.11	$<10^{-5}$	83.06	$<10^{-5}$
3	3	−0.02	>0.05	17.02	$<10^{-5}$	83	$<10^{-5}$
4	2	13.68	$<10^{-5}$	9.04	$<10^{-5}$	77.28	$<10^{-5}$
5	3	13.94	$<10^{-5}$	6.38	$<10^{-5}$	79.68	$<10^{-5}$
Data incorporating reference populations							
6	4	3.05	<0.01	14.62	$<10^{-5}$	82.33	$<10^{-5}$
7	4	3.57	<0.01	14.14	$<10^{-5}$	82.28	$<10^{-5}$
8	3	4.83	<0.005	14.98	$<10^{-5}$	80.19	$<10^{-5}$
9	2	−2.59	>0.05	17.13	$<10^{-5}$	85.47	$<10^{-5}$
10	2	9.03	$<10^{-5}$	12.29	$<10^{-5}$	78.68	$<10^{-5}$

Abbreviation: AMOVA, analysis of molecular variance.
Level 1, 14 populations in this study were incorporated as one group; Level 2, 14 populations were divided into two groups based on the geographic proximity: Tu, Mongolian, Yugu, Dongxiang, Bao'an and Salar/Xibo, Tataer, Uygur, Kirghiz, Kazakh, Tajike, Russ and Ozbek; Level 3, 14 populations were categorized into three groups based on the language affiliation: Tu, Mongolian, Yugu, Dongxiang, Bao'an and Xibo/Salar, Tataer, Uygur, Kirghiz, Kazakh and Ozbek/Tajike and Russ; Level 4, 14 populations were divided into two groups according to the clustering pattern on MDS plot: Tataer, Kirghiz, Dongxiang, Tajike and Salar/Ozbek, Uygur, Bao'an, Yugu, Russ, Xibo, Tu, Mongolian and Kazakh; Level 5, 14 populations were divided into three groups according to the clustering pattern on MDS plot: Tataer, Kirghiz, Dongxiang, Tajike and Salar/Ozbek, Uygur, Bao'an and Yugu/Russ, Xibo, Tu, Mongolian and Kazakh; Level 6, 14 populations in this study and 2 Northwestern Han populations/Northeastern Asians/Southeastern Asians/Central Asians; Level 7, 14 populations in this study/Northeastern Asians and 2 Northwestern Han populations/Southeastern Asians/Central Asians; Level 8, 14 populations in present study/Northeastern Asians and 2 Northwestern Han populations/Southeastern Asians; Level 9, 14 populations in this study/Central Asians; Level 10, all populations were divided into two groups according to the clustering pattern on MDS plot: Tataer, Kyrgyz, Dongxiang, Tajike, Tajik, Turkmen, Ozbek, Salar, Uygur, Uzbek, Bao'an and Kirghiz/Yugu, Xibo, Tu, Mongolian, Russ, Kazakh, XJ Han, GS Han, Kazak, Tibetan, GD Han, SC Han, Yao, Buyi, Daur, Manchu, Korean, Hezhe, Ewenki and Oroqen.
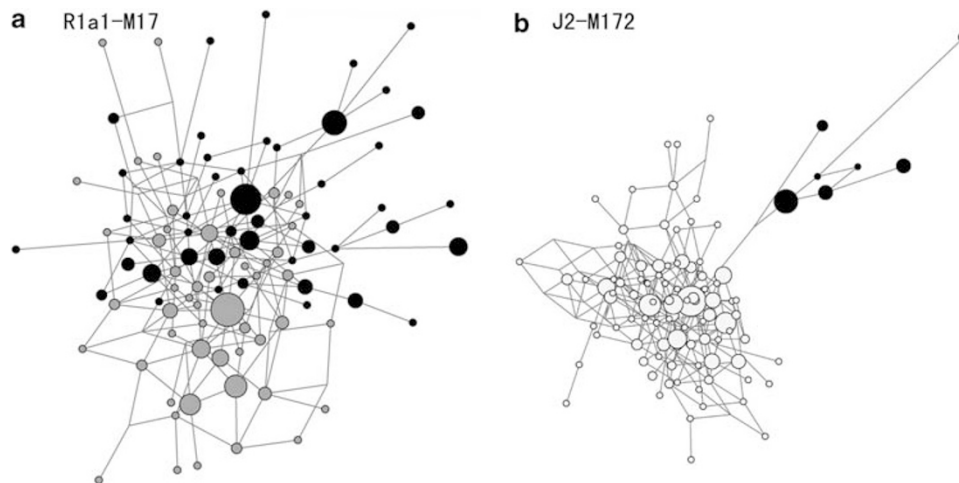
Figure 4 Median-joining networks of Y-STR haplotypes for haplogroups R1a1-M17 (**a**) and J2-M172 (**b**). Haplotypes are represented by circles with area proportional to the number of individuals. Colors indicate the geographic origin. Northwestern Chinese haplotypes are shown in black, Russian haplotypes in grey and western Eurasian haplotypes in white.

Table 4 Estimates of TMRCA for R1a1-M17 and J2-M172 based on Y-chromosome microsatellite variation

	Linear expansion time[a]		BATWING expansion time[b]				Divergence time[c]		
Haplogoup	Haplogroup variance	Time (years ago)	Mean (years ago)	Median (years ago)	0.025 Quantile	0.975 Quantile	Upper Limit (years ago)	Lower Limit (years ago)	Mean ± s.e. (years ago)
R1a1-M17	0.354	12 835	11 250	8086	506	82 480	10 255	4085	7170 ± 2459
J2-M172	0.425	15 413	7347	4998	197	61 015	6756	5065	5910 ± 4435

Abbreviations: BATWING, Bayesian analysis of trees with internal node generation; TMRCA, time to the most recent common ancestor.
[a]One-step mutation model.
[b]BATWING age estimates were under the assumption of exponential growth from initial constant population size.
[c]Divergence time was based on the single-nucleotide polymorphism-STR coalescent method.

The coalescent age estimates of R1a1 and J2 in Northwest China are listed in Table 4. The time calculated using linear expansion and BATWING methods was relatively overestimated than the divergence time. J2-M172 appeared slightly younger than R1a1-M17 except the estimate under linear expansion model. Therefore, the likely ages of these two haplogroups range from 5000 to 10 000 years. Although the precision of regional haplogroup dating can be affected by potential multiple founders, three independent estimating methods suggested similar time ranges. The age estimate indicated that the expansion of lineages in this area is a recent event.

DISCUSSION
In all, 14 ethnic groups from Silk Road region in Northwest China belong to either Turkic- and Mongolian-branch of Altaic Family or Indo-European Family, and most of them are Islamic. Nevertheless, the Y-chromosome data currently obtained suggest that their genetic backgrounds seem quite heterogeneous because of the contributions from multiple separate paternal lineages. In an attempt to explore the composition of paternal gene pool in this region, we recomputed in proportion the contributions of individual lineages to the total gene pool of 14 ethnic groups. Figure 5 explicitly manifests haplogroup 0-M175 and R1a1-M17 accounting for the largest proportion, and C-M130 is also notable.

Two major lineages in paternal gene pool
Parahaplogroup O-M175, the East Asian-specific lineage, is extensively distributed across entire East Asia region, in which its overall
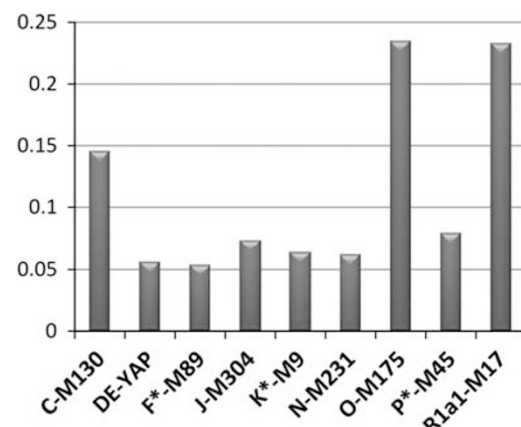
Figure 5 The proportion of different lineages in 14 ethnic groups as a whole. F*-M89 encompasses haplogroup G-M201, H-M69 and I-M170. P*-M45 contains all derived subhaplogroups except R1a1-M17.

frequency (sum of O1-M119, O2-M95 and O3-M122 frequency, above 50%) is uniquely high among East Asian populations.[1,34,35] Among three major haplogroups under O-M175, O3-M122 occurs at the highest frequency, above 40% on average. After systematically screening O3-M122-associated male subjects over nearly whole East Asia, Shi *et al*.[30] strongly argued southern origin of lineage O, which expanded northward approximately 25 000–30 000 years ago. In 14

ethnic groups involved in the present study, the overall frequency of haplogroup 0-M175 was 23.5%, out of which O3-M122 accounted for 14.1%. Present data were consistent with the argument of northward migration, and supported the proposed northwestward entrance of haplogroup O. As a result, O lineages became one of the main sources constituting the northwestern populations.

M173 is putatively regarded as an ancient marker across Eurasian. A vast majority of Y chromosomes belonging to R-M207 can be allocated to R1-M173 lineages, from which two predominant subclades, R1b3-M269 and R1a1-M17, were derived. Substantial reports supported that M269 had been well established throughout Europe since the Paleolithic era, and a spectrum of the frequency was observed at 40–80% with a descending cline from the west to the east.[7,14,38–40] R1b-M343, the ancestry of M269, was just at a detectable frequency in populations involved in this study. R1a1-M17 has an extensive and frequent distribution in Europe, entire Central Asia, Pakistan, northwestern India and West Asia.[7,32,38,41] Haplogroup R1a1-M17 was proposed to originating from South Russia/Ukraine region approximately 10 000 years ago and symbolized the spread of Kurgan culture over Central Asian steppes.[42] The distribution pattern of R1a1-M17 in Europe, in which the frequency increases with eastward gradient,[43–45] is contrary to that of R1ba-M269. In South Asia, a frequency of 15.8% was observed for Indian populations on the whole, and 24.4% for Pakistani.[32] The frequency of R1a1-M17 attained the highest in Central Asia.[8] Thus, R1a1-M17 served as one main expanding source from Central Asian nomads. A significant proportion of genetic components was occupied by R1a1-M17 in populations of Northwest China (Figure 5), which signifies that large masses of Altaic-speaking nomads from Central Asia entered into Northwest China because of multiple founders of R1a1-M17, as inferred from MJ network analysis, except that their arrival time, as estimated by haplogroup dating analysis (Table 4), may be far more recent than that of East Asian O lineages. Our estimation is in large agreement with the history of pastoralism in Central Asia dating back to 6000 years ago.[46]

Consequently, significant proportions occupied by haplogroup O-M175 and R1a1-M17 in paternally genetic composition indicate that the main ancestors of present northwestern populations in China were composed of antecedent East Asians peopling northwestward and later Central Asians immigrating eastward.

Component originating from West Asia

The currently available data maintained Near East as the corroborative homeland of haplogroup J bearing a mark of the Neolithic demic diffusion associated with the development of agriculture.[23,47] Early farming practicers introduced J2-M172 lineages through Levant Corridor into the Europe, where this haplogroup is mainly confined to the Mediterranean coastal areas, southeastern Europe and Anatolia.[14,38,48–51] The occurrence at a noticeable frequency was also observed in Central Asians.[7,8] The eastward expansion of haplogroup J2 to Iraq, Iran and Central Asia was also well documented in the Neolithic archeological records.[52] In this study, M172 was observed in Uygurs, Tajikes and Ozbeks, particularly 30.4% frequency in Ozbeks. Our result showed how far lineage J reached eastward. On the other hand, Islamism was delivered into China 1400 years ago from Arabia and Persia,[12,13] and the western lineages were largely attributable to the encroachment of these Muslims.[53] However, the estimated age of J2-M172 in this region seems more ancient than the history of Islamism. The haplotype clustering in MJ networks clearly demarcated between northwestern populations and western Eurasians. Therefore, the J2-M172 lineages penetrated into China, probably following eastward nomads from Central Asia, instead of directly from West Asia.

Limited gene flow

I-M170, a European-specific haplogroup, was considered as being of Balkan origin.[29,38] I-M170 lineages subsequently dispersed toward Caucasus region and Central Europe.[49,54] In our samples, I-M170 was observed exclusively in Tataer population at 33.3% frequency. G-M201 likely arose in Mesopotamia, and Caucasus and Anatolian populations showed a relatively high frequency.[14,55,56] Semino *et al.*[38] observed that haplogroup G-M201 occurred at 30.1% in Georgian, a Caucasus population. However, the frequency of G-M201 largely reduced in West Asia, India and Pakistan.[31,32,57] Haplogroup G-M201 was detected in as few as five individuals from our three northwestern populations. H-M69 was an emerging marker during the second great human migration from Middle East into Indian subcontinent, with a supposed history of 25 000 years. Moreover, H-M69 lineages more concentrated into clade H1-M52 in Indians, yet were rare and even absent in other populations.[32,40,58–60] We only distinguished four Y chromosomes harboring M69 from the Uygur population. The restricted and occasional distribution of G-M201, H-M69 and I-M170 under trunk F-M89 among northwestern populations of China and Central Asians may result from gene flows mediated by 'Silk Road' from West Eurasia and South Asia. The transcontinental communication complicated the genetic scenario of northwestern populations, but the genetic admixture took a marginal effect.

Haplogroup Q-M242, the sister clade of R-M207 derived from P-M45, is a preponderant lineage throughout Siberia.[61] A subset of Q-M242 lineages crossed over Bering Strait and entered America. During the expansion, a characteristic marker M3 emerged, defining Q3. As reported elsewhere, M3 in Asia was observed only in the extremely northeastern Siberian region of Chukchi Peninsula.[62,63] No Q3-M3 lineage was observed in our samples, and individuals carrying M242 sporadically occurred. Thus, it is hardly convictive to postulate the presence of genetic interaction between northwestern populations and Siberians.

Relay station during the northward migration of N

We typed an equivalent marker M231 for LLY22g polymorphism to distinguish haplogroup N, which was largely underrepresented in previous studies on East Asians. A detailed survey shed more insightful light on the phylogeography of haplogroup N.[64] N-M231 was presumably originated in the southwest of East Asia, and then covered almost entire North Eurasia, except for the little distribution in Central Siberia. N3 is the most frequent subclade under haplogroup N. A higher frequency of M231 was observed in Northeast Asia than in the south.[19] The presence of N3-Tat was also observed in Central Asians at a moderate frequency.[7,8] Our result that N-M231 extensively occurred, although infrequently, in northwestern populations suggests that Northwest China may serve as an intermediate transfer station on the migratory trajectory, where haplogroup N transferred west into Central Asia, and then entered Siberia.

Influence by recent historic events

C-M130 is a predominant haplogroup in northeastern Asians, owing to a higher frequency than southern populations. In particular, it is significantly frequent in Mongolians.[7,19,37] Not a little distribution of haplogroup C was observed in Central Asians and northwestern populations of China. Zerjal *et al.*,[65] through calculating the divergence time of 1000 years, revealed that the Mongol empire expansion left the distribution of C-M130 lineages crossing over a broad region from East to Central Asia. In our samples, the frequent occurrence of C-M130 in the Hazakh other than the Mongolian would be to a large

degree relevant to the historical document that masses of Mongols admixed into the precursors of Hazakh ethnic.[66]

An additional finding that 58% of paternal components in the Russ population belonged into haplogroup O was of great interest. An aberrant distribution of the typical East Asian haplogroup was observed in an atypical East Asian population. Thanks to the availability of the unambiguous ethnic origin, the Russ in China was populated by the immigrants from Russia after the eighteenth century. During the process that they admixed into East Asians, a higher mobility of female variants under the context of the patrilocal society gave rise to the sex-biased admixture,[67] which resulted in assimilating more genetic components from East Asians into the patrilineal gene pool of this ethnic population. Our unpublished data of mitochondrial DNA, which indicated more mitochondrial DNA lineages from European in the Russ population, substantiated this supposition.

Genetic heterogeneity among populations

All ethnic groups in this study are almost descendants of nomadic pastoralists. Notably, the practice of endogamous marriage and seasonal migration is often more prevailing in nomads.[46] Highly mobile and endogamous populations will not show associations between genetic variation and geography, a fact that has been shown for Jewish populations.[68] Multiple analysis results (the distribution of haplogroup, MDS plot and AMOVA analysis), not least the significant variance components among populations within groups ($P<10^{-5}$, second column in Table 3), concordantly showed the significant heterogeneity even among linguistic and geographic neighbors. The genetic heterogeneity may be attributed to two reasons: (1) the difference in isolation and strong cultural boundaries, resulting from different culture and living customs, presented the difference in the degree of genetic admixture; and (2) different populations experienced different population events such as migration, and consequently the current resident regions of these ethnic groups may not really reflect the original geographic peopling patterns of the nomadic ethnics.

Collectively, the above analyses on Y-chromosomal variations revealed the paternal genetic constitutions and the origin of northwestern populations in China, and the gene admixture processes were specified. The early East Asians of northwestward colonization met the later immigrants from Central Asia. J2-M172 was probably introduced into China by Central Asians. The gene flows from West Eurasia diversified the genetic scenarios through Silk Road, the influence of which on local populations, nevertheless, was limited.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Jin, L. & Su, B. Natives or immigrants: modern human origin in East Asia. *Nat. Rev. Genet.* **1,** 126–133 (2000).
2 Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazón, L. M., Foley, R. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65,** 43–62 (2001).
3 Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26,** 358–361 (2000).
4 Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4,** 598–612 (2003).
5 Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome. Res.* **18,** 830–838 (2008).
6 de Knijff, P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* **7,** 1055–1061 (2000).
7 Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J. *et al.* The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl Acad. Sci. USA* **98,** 10244–10249 (2001).
8 Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am. J. Hum. Genet.* **71,** 466–482 (2002).
9 Comas, D., Plaza, S., Wells, R. S., Yuldaseva, N., Lao, O., Calafell, F. *et al.* Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* **12,** 495–504 (2004).
10 Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R. *et al.* Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am. J. Hum. Genet.* **74,** 827–845 (2004).
11 Yao, Y. G., Kong, Q. P., Wang, C. Y., Zhu, C. L. & Zhang, Y. P. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in china. *Mol. Biol. Evol.* **21,** 2265–2280 (2004).
12 Du, R. & Yip, V. F. *Ethnic Groups in China* (Science Press, Beijing and New York,, 1993).
13 Sun, K. *The Research in Culture Changes of the Peoples Speaking Turkic Languages* Ningxia People's Publishing House, 2004, Yinchuan.
14 Cinnioğlu, C., King, R., Kivisild, T., Kalfoğlu, E., Atasoy, S., Cavalleri, G. I. *et al.* Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* **114,** 127–148 (2004).
15 Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome. Res.* **12,** 339–348 (2002).
16 Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S. *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.* **66,** 1580–1588 (2000).
17 Excoffier, L., Laval, L. G. & Schneider, S. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinformatics Online* **1,** 47–50 (2005).
18 Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131,** 479–491 (1992).
19 Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J. *et al.* Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* **172,** 2431–2439 (2006).
20 Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A. *et al.* The Himalayas as a directional barrier to gene flow. *Am. J. Hum. Genet.* **80,** 884–894 (2006).
21 Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16,** 37–48 (1999).
22 Derenko, M., Malyarchuk, B., Denisova, G. A., Wozniak, M., Dambueva, I., Dorzhu, C. *et al.* Contrasting Patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum. Genet.* **118,** 591–604 (2006).
23 Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V. *et al.* Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **74,** 1023–1034 (2004).
24 Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. & Freimer, N. B. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl Acad. Sci. USA* **91,** 3166–3170 (1994).
25 Kittles, R. A., Perola, M., Peltonen, L., Bergen, A. W., Aragon, R. A., Virkkunen, M. *et al.* Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* **62,** 1171–1179 (1998).
26 Wilson, I., Weale, M. & Balding, D. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. R. Stat. Soc.: Ser. A Stat. Soc.* **166,** 158–188 (2003).
27 Zhivotovsky, L. A. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol. Biol. Evol.* **18,** 700–709 (2001).
28 Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T. *et al.* The effective mutation rate at Y chromosome short tandem repeats, with application to human population divergence time. *Am. J. Hum. Genet.* **74,** 50–61 (2004).
29 Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M. *et al.* Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* **75,** 128–137 (2004).
30 Shi, H., Dong, Y. L., Wen, B., Xiao, C. J., Underhill, P. A., Shen, P. D. *et al.* Y-chromosome evidence of southern origin of the East Asian-specific Haplogroup O3-M122. *Am. J. Hum. Genet.* **77,** 408–419 (2005).
31 Nasidze, I., Quinque, D., Ozturk, M., Bendukidze, N. & Stoneking, M. MtDNA and Y-chromosome variation in Kurdish groups. *Ann. Hum. Genet.* **69**(Part 4), 401–412 (2005).
32 Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. T. *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.* **78,** 202–221 (2006).

33 Zerjal, T., Pandya, A., Thangaraj, K., Ling, E. Y., Kearley, J., Bertoneri, S. *et al.* Y-chromosomal insights into the genetic impact of the caste system in India. *Hum. Genet.* **121,** 137–144 (2007).

34 Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X. *et al.* Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am. J. Hum. Genet.* **74,** 856–865 (2004).

35 Su, B., Xiao, J., Underhill, P. A., Deka, R., Zhang, W., Akey, J. *et al.* Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* **65,** 1718–1724 (1999).

36 Shi, H., Zhong, H., Peng, Y., Dong, Y. L., Qi, X. B., Zhang, F. *et al.* Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* **6,** 45 (2008).

37 Karafet, T. M., Xu, L., Du, R., Wang, W., Feng, S., Wells, R. S. *et al.* Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* **69,** 615–628 (2001).

38 Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E. *et al.* The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290,** 1155–1159 (2000).

39 Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A. *et al.* A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* **70,** 1197–1214 (2002).

40 Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J. *et al.* The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72,** 313–332 (2003).

41 Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A. *et al.* Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.* **70,** 1107–1124 (2002).

42 Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S. Q. *et al.* Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am. J. Hum. Genet.* **68,** 537–542 (2001).

43 Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A. *et al.* Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67,** 1526–1543 (2000).

44 Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M. & Oppenheim, A. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am. J. Hum. Genet.* **69,** 1095–1112 (2001).

45 Weale, M. E., Yepiskoposyan, L., Jager, R. F., Hovhannisyan, N., Khudoyan, A., Burbage-Hall, O. *et al.* Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum. Genet.* **109,** 659–674 (2001).

46 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton,, 1994).

47 Semino, O., Passarino, G., Brega, A., Fellous, M. & Santachiara-Benerecetti, A. S. A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* **59,** 964–968 (1996).

48 Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & Bertranpetit, J. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68,** 1019–1029 (2001).

49 Barać, L., Pericić, M., Klarić, I. M., Rootsi, S., Janićijević, B., Kivisild, T. *et al.* Y chromosomal heritage of Croatian population and its island isolates. *Eur. J. Hum. Genet.* **11,** 535–542 (2003).

50 Di Giacomo, F., Luca, F., Popa, L. O., Akar, N., Anagnou, N., Banyko, J. *et al.* Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum. Genet.* **115,** 357–371 (2004).

51 King, R. & Underhill, P. A. Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. *Antiquity* **76,** 707–714 (2002).

52 Cauvin, J. *The Birth of the Gods and the Origins of Agriculture* (Cambridge Univ. Press, Cambridge, UK, 2000).

53 Wang, W., Wise, C., Baric, T., Black, M. L. & Bittles, A. H. The origins and genetic structure of three co-resident Chinese Muslim populations: the Salar, Bo'an and Dongxiang. *Hum. Genet.* **113,** 244–252 (2003).

54 Nasidze, I., Ling, E. Y., Quinque, D., Dupanloup, I., Cordaux, R., Rychkov, S. *et al.* Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann. Hum. Genet.* **68** (Part 3), 205–221 (2004).

55 Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A. *et al.* Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* **28,** 458–472 (2003).

56 Nasidze, I., Sarkisian, T., Kerimov, A. & Stoneking, M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum. Genet.* **112,** 255–261 (2003).

57 Flores, C., Maca-Meyer, N., Larruga, J. M., Cabrera, V. M., Karadsheh, N. & Gonzalez, A. M. Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. *J. Hum. Genet.* **50,** 435–441 (2005).

58 Ramana, G. V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P. A. *et al.* Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* **9,** 695–700 (2001).

59 Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S. *et al.* A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc. Natl Acad. Sci. USA* **103,** 843–848 (2006).

60 Sahoo, S. & Kashyap, V. K. Phylogeography of mitochondrial DNA and Y-chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. *Am. J. Phys. Anthropol.* **131,** 84–97 (2006).

61 Karafet, T. M., Osipova, L. P., Gubina, M. A., Posukh, O. L., Zegura, S. L. & Hammer, M. F. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.* **74,** 761–789 (2002).

62 Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J. *et al.* Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64,** 817–831 (1999).

63 Lell, J. T., Sukernik, R. I., Starikovskaya, Y. B., Su, B., Jin, L., Schurr, T. G. *et al.* The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* **70,** 192–206 (2002).

64 Rootsi, S., Zhivotovsky, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R. *et al.* A counter-clockwise northern route of the Y chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* **15,** 204–211 (2007).

65 Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S. *et al.* The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72,** 717–721 (2003).

66 Bao, E. H., Jiang, Y. & Ya, H. Z. *Cyclopedia of China-Nations* (Cyclopedia of China Press, Beijing,, 1986).

67 Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20,** 278–280 (1998).

68 Hammer, M. F., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T. M. *et al.* Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl Acad. Sci. USA* **97,** 6769–6774 (2000).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)