

Extended Y Chromosome Investigation Suggests Postglacial Migrations of Modern Humans into East Asia via the Northern Route

Hua Zhong,^{†1,2,3} Hong Shi,^{†2} Xue-Bin Qi,² Zi-Yuan Duan,⁴ Ping-Ping Tan,¹ Li Jin,⁵ Bing Su,^{*,2} and Runlin Z. Ma^{*,1}

¹Center for Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

²State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology and Kunming Primate Research Centre, Chinese Academy of Sciences, Kunming, China

³Graduate School, Chinese Academy of Sciences, Beijing, China

⁴Research Resources Center, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

⁵State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences, Fudan University, Shanghai, China

[†]These authors contributed equally to this study.

***Corresponding author:** E-mail: sub@mail.kiz.ac.cn; rlma@genetics.ac.cn.

Associate editor: Lisa Matisoo-Smith

Abstract

Genetic diversity data, from Y chromosome and mitochondrial DNA as well as recent genome-wide autosomal single nucleotide polymorphisms, suggested that mainland Southeast Asia was the major geographic source of East Asian populations. However, these studies also detected Central–South Asia (CSA)- and/or West Eurasia (WE)-related genetic components in East Asia, implying either recent population admixture or ancient migrations via the proposed northern route. To trace the time period and geographic source of these CSA- and WE-related genetic components, we sampled 3,826 males (116 populations from China and 1 population from North Korea) and performed high-resolution genotyping according to the well-resolved Y chromosome phylogeny. Our data, in combination with the published East Asian Y-haplogroup data, show that there are four dominant haplogroups (accounting for 92.87% of the East Asian Y chromosomes), O-M175, D-M174, C-M130 (not including C5-M356), and N-M231, in both southern and northern East Asian populations, which is consistent with the proposed southern route of modern human origin in East Asia. However, there are other haplogroups (6.79% in total) (E-SRY4064, C5-M356, G-M201, H-M69, I-M170, J-P209, L-M20, Q-M242, R-M207, and T-M70) detected primarily in northern East Asian populations and were identified as Central–South Asian and/or West Eurasian origin based on the phylogeographic analysis. In particular, evidence of geographic distribution and Y chromosome short tandem repeat (Y-STR) diversity indicates that haplogroup Q-M242 (the ancestral haplogroup of the native American-specific haplogroup Q1a3a-M3) and R-M207 probably migrated into East Asia via the northern route. The age estimation of Y-STR variation within haplogroups suggests the existence of postglacial (~18 Ka) migrations via the northern route as well as recent (~3 Ka) population admixture. We propose that although the Paleolithic migrations via the southern route played a major role in modern human settlement in East Asia, there are ancient contributions, though limited, from WE, which partly explain the genetic divergence between current southern and northern East Asian populations.

Key words: East Asia, Y chromosome, haplogroup Q, haplogroup R, postglacial migrations, northern route.

Introduction

Recent results from genome-wide autosomal single nucleotide polymorphisms (SNPs) showed that East Asian populations have a clinal structure with haplotype diversity decreasing from south to north (Abdulla et al. 2009). The south-to-north cline was also demonstrated by previous mitochondrial DNA (mtDNA) and Y chromosome studies (Su et al. 1999; Kivisild et al. 2002; Shi et al. 2005; Shi et al. 2008). However, there have been studies showing different patterns in which higher Y chromosome diversity in northern East Asia (NEAS) than in southern East Asia (SEAS) was observed, supporting that the NEAS

populations have a closer genetic relationship with Central Asian populations (Karafet et al. 2001; Xue et al. 2006). Indeed, many reported studies have detected Central–South Asia (CSA)- and/or West Eurasia (WE)-related genetic components in NEAS (Xinjiang was considered as NEAS in this study, although it is often used to represent Central Asian populations), such as Y chromosome (Karafet et al. 2001; Xue et al. 2006) and mtDNA (Yao et al. 2004; Yang et al. 2008). Similarly, autosomal markers also revealed CSA- and WE-related haplotypes in NEAS (Norton et al. 2007; Hellenthal et al. 2008; Abdulla et al. 2009). For example, the whole-genome data showed that

5% of the East Asian haplotypes were found only in CSA (Abdulla et al. 2009).

For these CSA- and WE-related genetic components in East Asia, it was interpreted either as the evidence of ancient migration via the northern route (Underhill et al. 2001) or simply as the reflection of recent population admixture (Xu et al. 2008) associated with trade along the ancient Silk Road (Yao et al. 2004; Zhang et al. 2007; Yang et al. 2008). It is implausible that the genetic input from CSA and WE is completely attributed to the ancient trade because high frequencies of CSA- and WE-related genetic components were found in NEAS populations (e.g., mtDNA >40%; Y chromosome DNA >50% in Xinjiang Uygur) (Yao et al. 2004; Hammer et al. 2006; Xue et al. 2006; Yang et al. 2008). Recently, two archeological populations (approximately 2,500 and 4,000 years before present, respectively) in Xinjiang were found containing 27% (Zhang et al. 2009) and 30% (Li et al. 2010) CSA- and WE-related mtDNA lineages, respectively, indicating that CSA- and WE-related genetic input into NEAS occurred earlier than the proposed relatively recent population admixture. These findings implied that there might be even earlier migration events in NEAS via the northern route.

There are at least two Y chromosome haplogroups that can be used to test ancient human dispersals into East Asia via the northern route. Haplogroup R-M207 occurs frequently in CSA and WE populations. It was detected in East Asia, especially in Uygur with relatively high frequency (Xue et al. 2006). Although the existence of R-M207 in NEAS could be explained by recent population admixture, there has not been systematic examination in East Asia. Another haplogroup is Q-M242, which was detected in CSA (Seielstad et al. 2003; Sharma et al. 2007), WE (Cinnioglu et al. 2004), Siberia (Pakendorf et al. 2006; Pakendorf et al. 2007), East Asia (Su et al. 2000), and the Americas (Bortolini et al. 2003; Zegura et al. 2004), but not in Southeast Asia (except for Vietnam) (Hammer et al. 2006) and Oceania (Karafet et al. 2010). Hence, a systematic screening of haplogroup Q-M242 in East Asia will probably provide new insights into the possible northern route and the early migrations into the Americas (Q1a3a-M3, a subclade of Q-M242, is the Native American-specific haplotype). It has been shown that the dominant haplogroup O-M175, D-M174, C-M130, and N-M231 in East Asia all have a southern origin (Su et al. 1999; Shi et al. 2005; Kumar et al. 2007; Rootsi et al. 2007; Li et al. 2008; Shi et al. 2008; Zhong et al. 2010). However, the remaining haplogroups detected in East Asia are still possibly related to CSA and WE populations.

Most of these CSA- and WE-related haplogroups identified previously in East Asians were not analyzed in detail owing to insufficient sampling and/or the scarcity of Y chromosome markers. Recently, the high-resolution Y chromosome phylogeny has been constructed and detailed phylogeographic information of the Y chromosome haplogroups has been revealed (Semino et al. 2004; Sengupta et al. 2006; Karafet et al. 2008; Tofanelli et al. 2009; Underhill et al. 2009; Balaesque et al. 2010), providing an opportunity to dissect the CSA- and WE-related Y chromosome components in East Asia.

In this study, we investigated Y chromosome population substructure of East Asians by high-resolution genotyping and fine-scale sampling in both NEAS and SEAS, and we performed phylogeographic analysis of CSA- and WE-related Y chromosome haplogroups in East Asia in order to trace the migratory history and geographic source.

Materials and Methods

Samples

A total of 3,826 unrelated male samples (116 populations from China and 1 population from North Korea; [figs. 1 and 2](#)) were recruited with informed consents. The protocol of this study was approved by the Institutional Review Board of Kunming Institute of Zoology, Chinese Academy of Sciences. In order to compare the Y chromosome population structure of East and West Eurasians, a total of 6,308 individuals' Y-haplogroup data from 137 populations ([supplementary tables S1 and S2, Supplementary Material online](#)) were extracted from the literature as reference populations (Cinnioglu et al. 2004; Hammer et al. 2006; Regueiro et al. 2006; Sengupta et al. 2006; Capelli et al. 2007; Gayden et al. 2007; Balanovsky et al. 2008; Gusmao et al. 2008; King et al. 2008; Lopez-Parra et al. 2009; Mirabal et al. 2009). In addition, according to the Y-haplogroup data of East Asians, the Y chromosome short tandem repeat (Y-STR) data ([supplementary table S3, Supplementary Material online](#)) were extracted from the literature (Seielstad et al. 2003; Cinnioglu et al. 2004; Shi et al. 2005; Bosch et al. 2006; Hammer et al. 2006; Sengupta et al. 2006; Capelli et al. 2007; Martinez et al. 2007; Nonaka et al. 2007; Sharma et al. 2007; King et al. 2008; Li et al. 2008; Shi et al. 2008; Tofanelli et al. 2009; Zhong et al. 2010).

Y Chromosome Genotyping

According to the hierarchical genotyping strategy (Underhill et al. 2000; Hammer et al. 2001), M175 was typed first because it occurs at high frequency in East Asians (Shi et al. 2005); all individuals without M175-Del were examined for M130, YAP, and M231; then, the rest of individuals who are ancestral at M175, M130, YAP, and M231 were subject to further subtyping according to the high-resolution Y Chromosomal Haplogroup Tree (Karafet et al. 2008; Underhill et al. 2009). The Y chromosome biallelic markers genotyped are shown in [figure 2](#), which were determined by sequencing and PCR-restriction fragment length polymorphism. To evaluate the diversity of haplogroup F2, G, H, J, L, Q, R, and T, we also typed eight commonly used Y-STR markers: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393 using fluorescence-labeled primers with an ABI 3730 Genetic Analyzer. The Y-STR nomenclature follows the system proposed by Butler et al. (2002). In addition, another Y-STR marker DYS7.2 (Cinnioglu et al. 2004; Sengupta et al. 2006) was typed in the M241-derived individuals.

Data Analysis

Together with the published data ([supplementary tables S1 and S2, Supplementary Material online](#)), the haplogroup



Fig. 1 Geographic locations of the 117 populations sampled in this study.

frequency data of East Asians were used to perform correspondence analysis (CA) using software XLSTAT (<http://www.xlstat.com>) and show the geographical distribution of different haplogroups. Owing to incomplete genotyping of the literature data, subbranches of a haplogroup were integrated into the ancestral markers (haplogroup D, E, C, G, H, I, J, L, N, O, Q, and R) and then the converted frequency data ([supplementary table S1, Supplementary Material online](#)) were used to perform CA.

CSA- and WE-related Y chromosome haplogroups were selected to construct their median-joining networks using Y-STR data ([supplementary table S3, Supplementary Material online](#)), including data from the literature and this study. Median-joining networks were constructed using the program NETWORK 4.5.1.0 (Fluxus Engineering) (Bandelt et al. 1999). When the constructed networks are lack of resolution, they were reconstructed by the median-joining method after processing the data with the reduced median method (reduction threshold = 1.0) or reweighting the STR loci according to their variances (higher weights were assigned to the least variable loci). The age of STR variation within each Y chromosome haplogroup was estimated following the published method (Zhivotovsky 2001; Zhivotovsky et al. 2004; Sengupta et al. 2006). During the analysis of Y-STRs, DYS389II was

named DYS389b after subtracting DYS389I because DYS389II contains the repeat number of DYS389I.

Results

Y Chromosome Haplogroups Detected in East Asia

As shown in [figure 2](#) and [supplementary table S1 \(Supplementary Material online\)](#), there are four dominant Y chromosome haplogroups in East Asians, that is, D-M174 (10.72%), C-M130 (12.43%, not including C5-M356), N-M231 (5.96%), and O-M175 (63.75%), which account for 92.87% of the Y chromosomes in East Asia. Because these four haplogroups have been well studied and shown clear southern origin during Paleolithic time (Su et al. 1999; Shi et al. 2005; Kumar et al. 2007; Rootsi et al. 2007; Li et al. 2008; Shi et al. 2008; Zhong et al. 2010), we did not conduct subtyping except for C-M130 in our reported subtyping data (Zhong et al. 2010).

In the remaining samples, we did not detect the two presumably ancient haplogroup F*-M89 and K*-M9, different from the published data (Wen, Xei, et al. 2004; Xue et al. 2006). We observed the recently reported haplogroup F2-M427 (Sengupta et al. 2006) in the Tibeto-Burman-speaking populations of SEAS (0.3%).



720

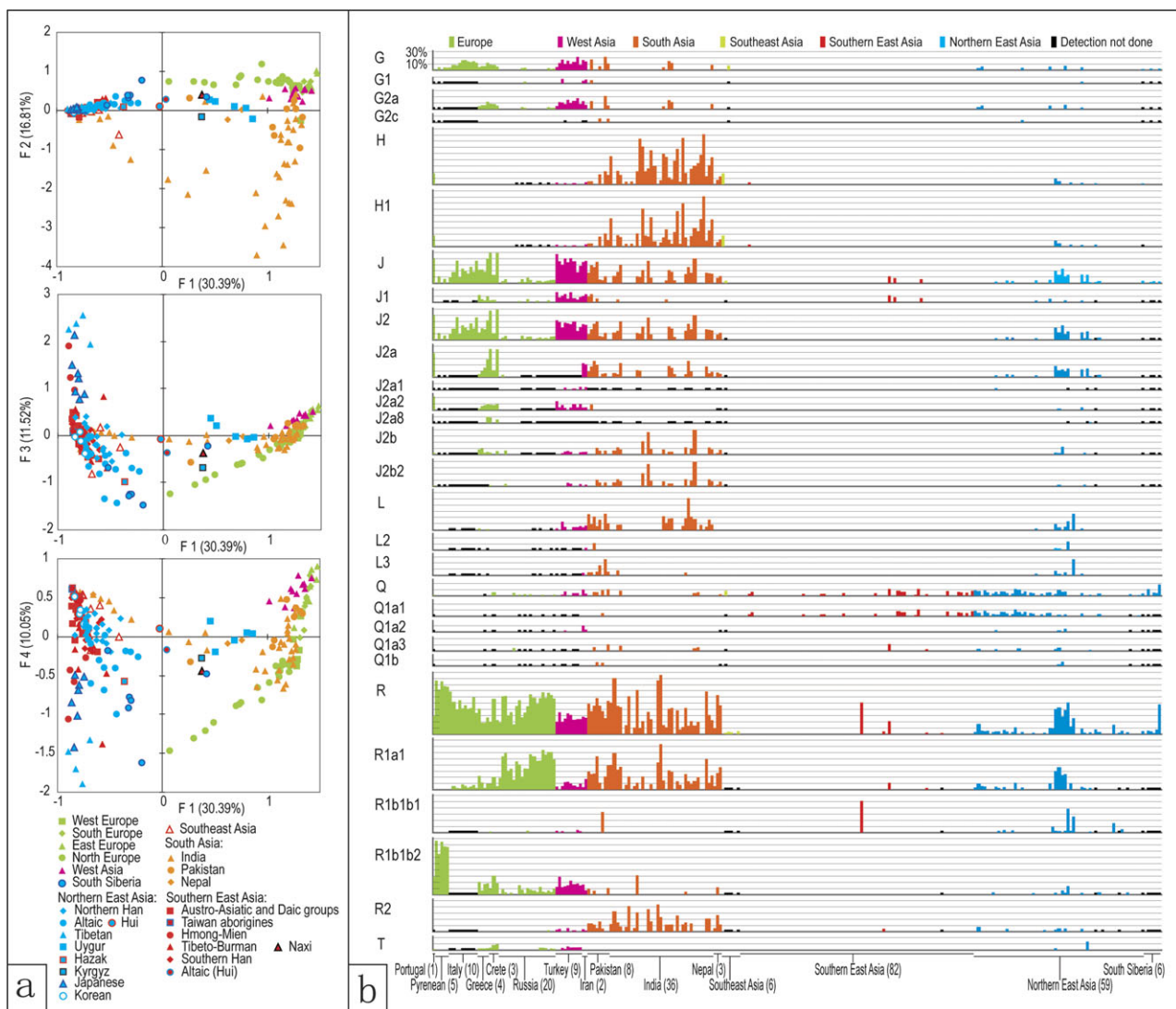


FIG. 3 Analysis of Y chromosome population substructure. (a) CA plots of 254 Eurasian populations. (b) Comparison of the frequencies of CSA- and WE-related haplogroups among 254 Eurasian populations. The horizontal axis denotes different Eurasian populations, which are successively marked. The number in the bracket denotes population size. The vertical axis denotes percentage frequency, which is labeled only in haplogroup G.

The rest of the Y chromosomes, accounting for 6.79% in total, were identified as haplogroups E-SRY4064, C5-M356, G-M201, H-M69, I-M170, J-P209, L-M20, Q-M242, R-M207, and T-M70, and we define these haplogroups as CSA- and WE-related haplogroups because they occur preferentially in CSA and/or WE populations (Semino et al. 2000; Underhill et al. 2000; Cinnioglu et al. 2004; Rootsi et al. 2004; Semino et al. 2004; Regueiro et al. 2006; Sengupta et al. 2006; Capelli et al. 2007; Martinez et al. 2007; Balanovsky et al. 2008; Gusmao et al. 2008; King et al. 2008; Lopez-Parra et al. 2009; Mirabal et al. 2009). Among them, E-SRY4064, C5-M356, G-M201, H-M69, I-M170, L-M20, and T-M70 have low frequencies (0.04–0.24%), whereas J-P209, Q-M242, and R-M207 have relatively high frequencies, 1.09%, 1.66%, and 3.06% (supplementary table S1, Supplementary Material online), respectively. Notably, these CSA- and WE-related haplogroups were detected mainly in NEAS populations. We conducted subtyping of the CAS- and WE-related haplogroups and the result is shown in figure 2.

Y Chromosome Population Substructure of East Asians

We performed CA based on the genotyping data of Y chromosome biallelic markers. In the CA plots (fig. 3a), we observed the following features: 1) the SEAS populations cluster tightly together except for the Naxi group from the literature (Sengupta et al. 2006) and Yunnan Hui (a Muslim population); 2) the SEAS and most of the NEAS populations cluster together with all Southeast Asian populations, but cluster with only 7 of 37 South Asian populations (3 Tibeto-Burman groups from northeastern India, 2 Austro-Asiatic groups from eastern India, 1 Dravidian group from southern India, and 1 Sino-Tibetan group from Nepal); 3) NEAS populations are loosely clustered because many of them are close to South Asian and West Eurasian populations; 4) in particular, Uyghur, Kyrgyz, and Hui populations in northwestern China show close relationship to West Eurasian populations. Although two SEAS populations (Hui and Naxi) are also close to West Eurasians, they

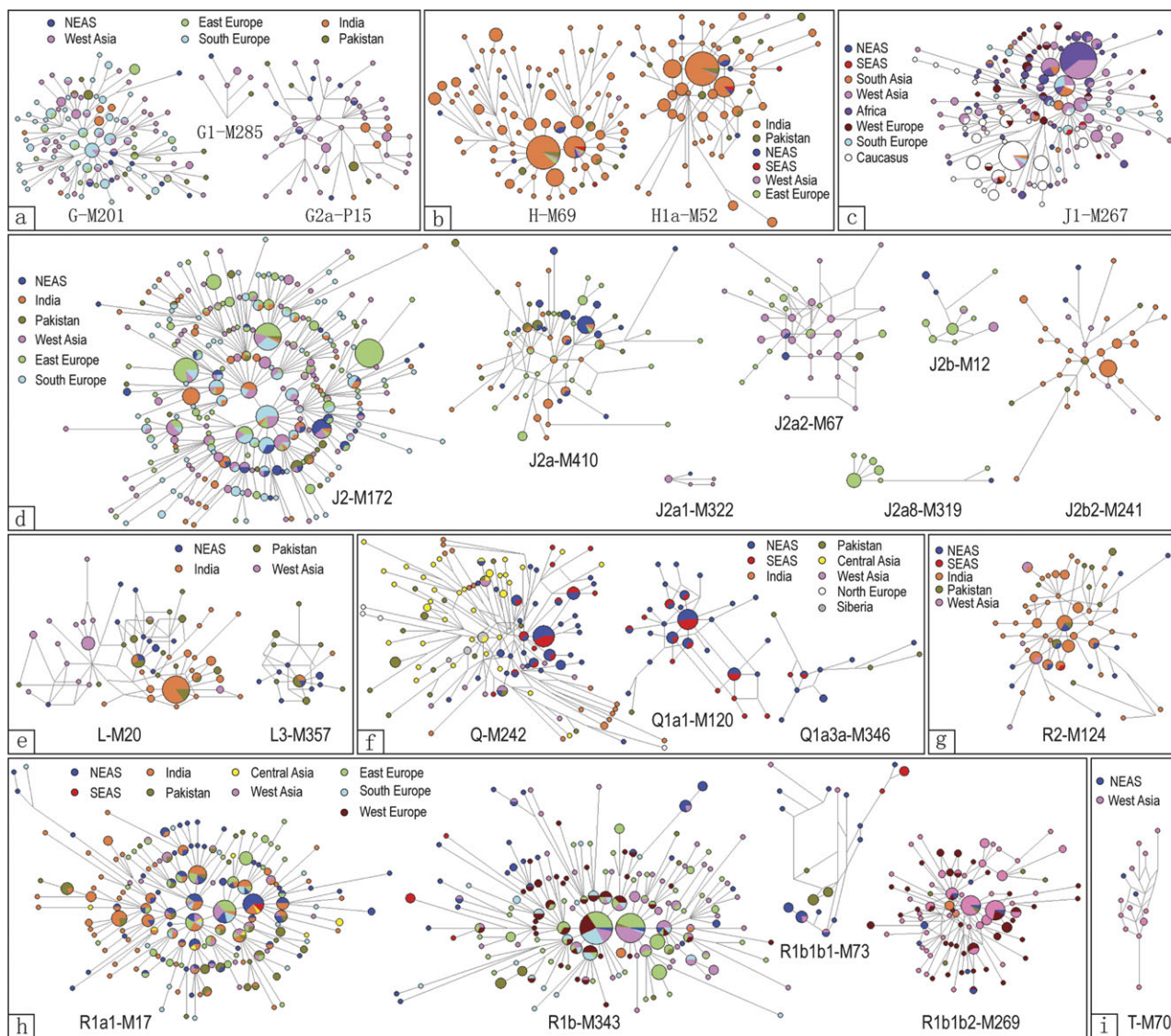


FIG. 4 The median-joining networks of Y-STR haplotypes within Y chromosome haplogroups. The median-joining networks were constructed on the basis of eight commonly used Y-STRs except that the network of J2b2-M241 was done by the eight Y-STRs and DYS472. Because many individuals within haplogroup G-M201, H-M69, J1-M267, J2-M172, L-M20, Q-M242, and R1b-M343 were not subtyped in the literature, all haplotypes within a haplogroup were included to construct a network at first and then subtyped individuals were used to construct networks of subhaplogroups. The networks of haplogroup G-M201, H-M69, J1-M267, J2-M172, R1a1-M17, and R1b-M343 were constructed by the median-joining method after processing the data with the reduced median method (reduction threshold = 1.0). The sizes of the nodes are proportional to their frequencies. The lengths of the lines are proportional to the mutational steps.

are the known immigrants from northwestern China (Du and Yip 1993). Collectively, these features indicate that CSA- and WE-related Y chromosome haplogroups occur almost exclusively in northwestern populations rather than in SEAS and Southeast Asian populations. As expected, the Xinjiang Uygur populations contain all those CSA- and WE-related Y chromosomes (fig. 2), which account for 71.79% of all Uygur males from Xinjiang, China.

Phylogeographic Analysis of CSA- and WE-Related Y Chromosome Haplogroups

Haplogroup E, C5, G, H, I, L, and T are prevalent in CSA and/or WE populations with more subhaplogroups, and they also appear in NEAS populations with low frequency,

but generally absent in SEAS populations (figs. 2 and 3b and supplementary tables S1 and S2, Supplementary Material online). The Y-STR haplotypes of CSA and WE populations are highly diverged with consecutive mutational steps, whereas those of NEAS populations display either partial cluster or sporadic distribution in the median-joining networks (fig. 4). Within haplogroup G and T, the NEAS Y-STR haplotypes mainly link to West Asians, whereas those of haplogroup H and L tend to link to Indians and Pakistanis, respectively.

Haplogroup Q occurs in Eurasians and its subhaplogroups show different distributions. Q1a1-M120 is widely distributed in both SEAS and NEAS populations, but absent outside East Asia except for one incidence

Table 1. Ages of Y-STR Variation within Y Chromosome Haplogroups

Haplogroup	Age ± Standard Error (n) in thousand years ago
F2-M427 ^a	2.42 ± 1.96 (15)
G1-M285 ^b	14.79 ± 4.77 (7)
G2a-P15 ^b	14.58 ± 2.90 (66)
H1-M52 ^b	9.68 ± 2.07 (167)
J1-M267 ^b	23.11 ± 11.34 (344)
J2a-M410 ^b	17.83 ± 4.63 (176)
J2a1-M322 ^b	5.18 ± 2.42 (7)
J2a2-M67 ^b	12.89 ± 2.39 (52)
J2a8-M319 ^b	8.05 ± 4.44 (18)
J2b-M12 ^b	12.14 ± 2.58 (69)
J2b2-M241 ^b	13.86 ± 3.15 (49)
L2-M317 ^b	12.45 ± 2.47 (6)
L3-M357 ^b	7.93 ± 2.43 (24)
Q1a1-M120 ^a	15.42 ± 7.25 (61)
Q1a2-M25 ^b	2.72 ± 1.24 (5)
Q1a3*-M346 ^b	17.77 ± 4.41 (13)
R1a1*-M17 ^a	15.37 ± 3.46 (84)
R1b-M343 ^b	17.35 ± 4.53 (338)
R2-M124 ^b	11.73 ± 1.67 (100)
T-M70 ^b	15.40 ± 1.35 (14)
C1-M8 ^a	9.96 ± 3.54 (20)
C3*-M217 ^a	18.78 ± 4.19 (374)
C3c-M48 ^a	9.31 ± 2.85 (18)
C3d-M407 ^a	9.47 ± 2.71 (22)
C3e-P53.1 ^a	4.83 ± 2.34 (15)
D*-M174 ^c	50.62 ± 16.60 (44)
D1-M15 ^a	32.67 ± 6.28 (233)
D2-M57 ^a	16.27 ± 3.33 (106)
D3-P47 ^a	12.97 ± 3.35 (282)
O1a*-M119 ^d	35.05 ± 4.58 (346)
O1a2-M110 ^d	39.93 ± 11.55 (21)
O2a*-M95 ^d	41.10 ± 7.61 (343)
O3*-M122 ^a	37.24 ± 4.56 (27)
O3a*-M324 ^a	39.29 ± 5.03 (355)
O3a3b-M7 ^a	41.74 ± 7.00 (88)
O3a3c1*-M133 ^a	29.23 ± 5.47 (242)
O3a3c-M134 ^a	35.32 ± 4.66 (139)

NOTE.—The numbers (n) in brackets are sample size.

^a Including East Asians only.

^b Including individuals from East Asia, CSA, and WE.

^c Including East Asians and Andamanese.

^d Including East and Southeast Asians.

observed in northern Pakistan (figs. 2 and 3b and supplementary table S2, Supplementary Material online), which is consistent with another published data of 49 Eurasian populations (Wells et al. 2001). Q1a1-M120 was not detected in Europe, West Asia, Caucasia, and Central Asia except for a Sino-Tibetan group in Kyrgyzstan. The other subhaplogroups (Q1a2-M25, Q1a3*-M346, and Q1b-M378) were detected in NEAS, CSA, and WE populations, but absent in SEAS population except for one Han group and Yunnan Hui (figs. 2 and 3b and supplementary table S2, Supplementary Material online) with known migrations from the north (Du and Yip 1993; Wen, Li, et al. 2004). In further network analysis (fig. 4f), the East Asian Y-STR haplotypes form a relatively separate cluster, whereas the CSA Y-STR haplotypes show a loose structure and long mutational steps; also, the NEAS Y-STR haplotypes display higher differentiation than the SEAS haplotypes (Q1a1-M120

and Q1a3*-M346). The high diversity and structured distribution of Q-M242 lineages in CSA, WE, and NEAS populations are indicative of their prehistoric migrations via the northern route into East Asia.

Haplogroup R occurs at high frequencies in Eurasians, especially in West Eurasians (fig. 3b and supplementary table S2, Supplementary Material online). In East Asia, it is frequent in NEAS but rare in SEAS (sparsely detected in 5 of 82 SEAS populations). The R subhaplogroups mainly occur in CSA and WE: R1a1-M17 in Europe, West Asia, and CSA; R1b1b2-M269 in Europe and West Asia; and R2-M124 in South Asia. However, R1b1b1-M73 was detected mainly in NEAS and was sporadically detected in South Asia and WE. As shown in the network of R1b-M343 (fig. 4h), most of the M73-derived individuals occur at the terminals of the network with multistep mutations, indicating a different origin of M73 in comparison with those West Eurasian individuals and a possible origin in NEAS because of its high Y-STR diversity. All M269-derived individuals from NEAS either link to or share haplotypes with West Eurasians. A recent study suggests a West Asian origin of R1b1b2-M269 and its Neolithic expansion (Balaesque et al. 2010). Therefore, the NEAS R1b1b2-M269 likely came from West Asia. Although R1a1-M17 occurs across Eurasia, its network (fig. 4h) does not exhibit region-specific cluster; also, the NEAS Y-STR haplotypes of R1a1-M17 show consecutive mutation steps and high differentiation. In addition, these M17-derived individuals were proven to be the ancestral R1a1* by subtyping seven downstream SNPs (fig. 2). Therefore, this lineage likely represents the early immigrants from CSA or WE. Haplogroup R2-M124 has a predominant distribution in South Asia (fig. 3b) and exhibits a star-like structure in the network (fig. 4g). Also, more diversified haplotypes occur in South Asia instead of NEAS. These features of R2-M124 reflect a relatively recent expansion from South Asia to NEAS via the northern route. The phylogeography of haplogroup R excludes the possibility that it migrated eastward via the southern route.

The WE origin haplogroup J (Semino et al. 2004) presents a similar distribution pattern with R in East Asia, primarily occurring in NEAS, but rare in SEAS (fig. 2). As shown in figure 4c and d, there is higher Y-STR diversity in CSA and WE than in East Asian populations. East Asian J1-M267 individuals share haplotypes with West Eurasians and tend to link to Caucasian haplotypes. J2b2-M241 individuals in NEAS contain an 8-repeat motif at the A7.2 STR locus (West Eurasian 8-repeat motif vs. South Asian 7-repeat motif; Cinnioglu et al. 2004; Sengupta et al. 2006) and link to West Asian haplotypes. In the network of J2-M172, 61% of the NEAS haplotypes share with West Eurasians and South Asians, again suggesting WE origin of these haplotypes.

The Age of STR Variation within Y Chromosome Haplogroups

To evaluate the early movements of different haplogroups detected in East Asia, the ages of STR variation within each

Y chromosome haplogroup were calculated based on the eight commonly used Y-STR loci except for O1a*-M119, O1a2-M110, and O2a*-M95 without including DYS389b (supplementary table S3, Supplementary Material online). As shown in table 1, the predominant haplogroups in East Asians all display ancient ages of STR variation, especially D1, O1, O2, and O3 (more than 30 Ka), whereas those mainly occurring in NEAS show relatively young ages, most of which occurred ~10–18 Ka. The estimated ages of the nonsouthern origin haplogroups using data only from East Asian populations suggest that there were Paleolithic migrations (>10 Ka) from CAS and/or WE via the northern route, although recent East-West admixture in NEAS (<3 Ka) also existed.

Discussion

Southern versus Northern Route

Our previous Y chromosome data, primarily from haplogroup O-M175, C-M130, and D-M174 (Su et al. 1999; Shi et al. 2005; Shi et al. 2008; Zhong et al. 2010), suggested that the southern route made a substantial contribution to the early peopling of anatomically modern humans in East Asia, which was also supported by the phylogeography of haplogroup N-M231 (Rootsi et al. 2007). In this study, the four predominant haplogroups account for 92.87% of the East Asian Y chromosomes and show continuous distribution in East Asians, again indicating the substantial contribution of the southern route.

However, the CSA- and WE-related Y chromosome haplogroups were also detected accounting for 6.79% of the East Asian males and exhibit structured distribution. They occur primarily in northwestern East Asia and gradually decrease from northwest to northeast. On the other hand, in contrast to the ages of STR variation within haplogroup C, D, and O (18.78–50.62 Ka), the CSA- and WE-related haplogroups and subhaplogroups show relatively young ages, ranging from 2.72 to 23.11 Ka (table 1). Among the 19 CSA- and WE-related haplogroups and subhaplogroups, the majority of them range from 9.68 to 17.83 Ka, suggesting ancient migrations from CSA and WE into NEAS via the northern route, which likely occurred following the last glacial maximum.

Furthermore, maternally inherited mtDNA shows similar distribution pattern with Y chromosomes in East Asia. The East Asian-specific mtDNA haplogroups are prevalent in both SEAS and NEAS populations (Kivisild et al. 2002; Derenko et al. 2007), whereas the West Eurasian-specific mtDNA haplogroups occur mainly in NEAS, especially in Uyghur and Hui (Yao et al. 2004; Derenko et al. 2007; Yang et al. 2008). Additionally, these West Eurasian mtDNA haplogroups also show postglacial dispersal in northern Asia (Derenko et al. 2007). Therefore, both Y chromosome and mtDNA data support that the Paleolithic migrations of modern humans via the southern route had major contribution to current East Asians, whereas the postglacial migrations via the northern route also contributed to current NEAS populations.

Early Postglacial and Neolithic Migrations to NEAS via the Northern Route

Although the CSA- and WE-related Y chromosome haplogroups account for only 6.79% of East Asian males, it is noteworthy that 69.55% of them belong to two haplogroups Q and R, which probably represent early migrations from CSA and WE into NEAS via the northern route.

Q1a1-M120 and Q1a3*-M346 are the two major sublineages of haplogroup Q, which have similar ages of STR variation, 15.42 and 17.77 Ka, respectively (table 1). Q1a1-M120 is an East Asian-specific subhaplogroup. It occurs in most of the NEAS populations. In SEAS populations, it occurs mainly in southern Han Chinese with relatively low Y-STR diversity (fig. 4f), implying that the spread of Q1a1-M120 was from north to south likely due to the demic diffusion of Han culture during Neolithic time (Wen, Li, et al. 2004). Q1a3*-M346 was not detected in the SEAS aborigines with only sporadic appearance in southern Han and Hui, again suggesting its southward migrations from the NEAS populations (Du and Yip 1993; Wen, Li, et al. 2004). Notably, Q1a3*-M346 is the ancestral haplogroup of Q1a3a-M3, which occurs only in Native American populations (Bortolini et al. 2003; Zegura et al. 2004; Karafet et al. 2008). The trace of Q1a3*-M346 in the NEAS populations provides important evidence in supporting the proposed prehistoric migration from Central Asia to the Americas. Collectively, the phylogeographic structure of haplogroup Q reveals early demographic expansions via northern Eurasia.

R1a1*-M17, recently renamed R1a1a* (Underhill et al. 2009), is the major sublineage of haplogroup R, and it has similar distribution pattern with Q1a1-M120 and Q1a3*-M346 in East Asia, prevalent in NEAS but rare in SEAS populations. The STR variation age of R1a1* in East Asia (15.37 Ka) is also similar with those of Q1a1-M120 and Q1a3*-M346, suggesting that R1a1* was one of the lineages entering East Asia via the northern route. It is well known that R1a1*-M17 is popular in West Asia and Europe. However, the currently known seven subbranches of M17 (Underhill et al. 2009) were not detected in our 84 M17-derived individuals, implying that these M17-derived individuals probably migrated into NEAS before the occurrence of the seven subbranches in West Asia and Europe. Interestingly, the STR variation age of East Asian R1a1*-M17 is similar with the age of West Indian R1a1*-M17 (15.8 Ka), both of which are older than the R1a1*-M17 in CSA and WE (Underhill et al. 2009). The ancient age of East Asian R1a1*-M17 can also be reflected in the network (fig. 4h) by exhibiting high differentiation and consecutive mutational steps. All these features indicate that R1a1*-M17 in East Asia were ancient immigrants probably from CSA and underwent long-period independent differentiation. However, more informative subbranch markers are required for a better phylogeography of M17.

Neolithic expansions in the near and middle east have been demonstrated contributing greatly to the peopling of Europe (Semino et al. 2004; Balareshque et al. 2010). Did the

Neolithic expansions contribute to NEAS populations? Several Neolithic expansion-signature Y chromosome haplogroups were detected in NEAS, such as R1b1b2-M269 (Balaesque et al. 2010), J1-M267 (Tofanelli et al. 2009), and J2a-M410 (Semino et al. 2004). Furthermore, these haplogroups and several other haplogroups (G-201, H-M69, L-M20, and R2-M124) detected in NEAS all show close relationship with CSA and WE haplotypes in the network (fig. 4), therefore suggesting the existence of Neolithic migrations to NEAS via the northern route. The Neolithic expansions is further supported by evidence from ancient human remains that the CSA- and WE-related genetic components were brought into NEAS and Siberia about 2,000–4,000 years ago (Francalacci 1995; Mair 1995; Lalueza-Fox et al. 2004; Keyser et al. 2009; Zhang et al. 2009; Li et al. 2010).

The early migrations into NEAS via the northern route can also be reflected by data from mtDNA. Several mtDNA haplogroups (H, V, and X), showing signatures of postglacial expansions (Torroni et al. 2001; Reidla et al. 2003; Derenko et al. 2007), were detected in NEAS populations (Yao et al. 2004; Yang et al. 2008). More detailed analysis of mtDNA in East Asia is needed to quantify the maternal contribution of the early migrations via the northern route.

More Recent Gene Flow from CSA and WE to NEAS

The trade via the ancient Silk Road was widely applied to explain the occurrence of CSA- and WE-related genetic components in NEAS (Yao et al. 2004; Yang et al. 2008). We show that the majority of the CAS- and WE-related genetic contribution to NEAS occurred much earlier than the time of ancient Silk Road. However, in the NEAS populations, the impact of the ancient Silk Road can also be reflected by the sporadic appearance of the minor CSA- and WE-related haplogroups, such as E-SRY4064, C5-M356, I-M170, J2a2*-M67, Jab2-M241, and T-M70, which had not have enough time to accumulate within-haplogroup genetic diversity. Among these rarely detected haplogroups, E-SRY4064, I-M170, and T-M70 are prevalent in WE populations. Interestingly, these features are consistent with a recent result that gene flows from WE populations to the extant ethnic groups in Northwest China were relatively weak (Shou et al. 2010), indicating that recent population admixture is rather limited and those CSA- and WE-related haplogroups mainly came from ancient demographic expansions via the northern route. Finally, it should be noted that the Muslim populations (Hui) in China were immigrants from Central Asia, the Persians and Arabs during Yuan Dynasty (~700 years ago) (Du and Yip 1993; Gladney 1996; Lipman 1997); indeed, the Hui populations were found carrying the majority of CSA- and WE-related Y chromosomes (fig. 2).

Conclusions

Our investigation of East Asian Y chromosomes, based on high-resolution genotyping and fine-scale sampling, supports the existence of demographic expansions toward East Asia via the northern route, which started ~15–18 Ka

(following the last glacial maximum). However, the demographic input via the northern route is mostly restricted to NEAS populations and just contributed limited Y chromosomes to current East Asian populations (6.79%). Among them, haplogroup Q-M242 and R-M207 likely represent the earliest settlers via the northern route. In addition, it was confirmed that the Paleolithic colonization of anatomically modern humans in East Asia via the southern route have made substantial contribution to the extant East Asian Y chromosome gene pool (92.87%). Paleolithic migrations via the southern route shaped the south-to-north clinal structure; the postglacial migrations via the northern route enlarged the genetic divergence between SEAS and NEAS populations.

Supplementary Material

Supplementary tables S1–S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We are grateful to all the voluntary donors of DNA samples in this study. We thank Hui Zhang and Xiangqun Li for their technical assistance. This study was supported by grants from the National 973 project of China (2007CB947701 and 2007CB815705), the Chinese Academy of Sciences (KSCX1-YW-R-34, Westlight Doctoral Program), the National Natural Science Foundation of China (30413242, 30771148, 30971566, 30700445, 30630013, and 30771181), and the Natural Science Foundation of Yunnan Province of China (2007C100M and 2009CD107).

References

- Abdulla MA, Ahmed I, Assawamakin A, et al. (92 co-authors). 2009. Mapping human genetic diversity in Asia. *Science* 326: 1541–1545.
- Balanovsky O, Rootsi S, Pshenichnov A, et al. (11 co-authors). 2008. Two sources of the Russian patrilineal heritage in their Eurasian context. *Am J Hum Genet.* 82:236–250.
- Balaesque P, Bowden GR, Adams SM, et al. (16 co-authors). 2010. A predominantly Neolithic origin for European paternal lineages. *PLoS Biol.* 8:e1000285.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.
- Bortolini MC, Salzano FM, Thomas MG, et al. (21 co-authors). 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet.* 73:524–539.
- Bosch E, Calafell F, Gonzalez-Neira A, et al. (13 co-authors). 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet.* 70:459–487.
- Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int.* 129:10–24.
- Capelli C, Brisighelli F, Scarnicci F, et al. (11 co-authors). 2007. Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Mol Phylogenet Evol.* 44:228–239.
- Cinnioglu C, King R, Kivisild T, et al. (15 co-authors). 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet.* 114:127–148.

- Derenko M, Malyarchuk B, Grzybowski T, et al. (12 co-authors). 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet.* 81:1025–1041.
- Du RF, Yip VF. 1993. Ethnic groups in China. Beijing, China: Science Press.
- Francalacci P. 1995. DNA analysis of ancient desiccated corpses from Xinjiang. *J Indoeur Stud.* 23:385–389.
- Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2007. The Himalayas as a directional barrier to gene flow. *Am J Hum Genet.* 80:884–894.
- Gladney DC. 1996. Muslim Chinese: ethnic nationalism in the People's Republic. 2nd ed., Harvard East Asian monographs. London: Harvard University Press.
- Gusmao A, Gusmao L, Gomes V, Alves C, Calafell F, Amorim A, Prata MJ. 2008. A perspective on the history of the Iberian gypsies provided by phylogeographic analysis of Y-chromosome lineages. *Ann Hum Genet.* 72:215–227.
- Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S. 2006. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet.* 51:47–58.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol.* 18:1189–1203.
- Hellenthal G, Auton A, Falush D. 2008. Inferring human colonization history using a copying model. *PLoS Genet.* 4:e1000078.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet.* 69:615–628.
- Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF. 2010. Major east-west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol.* 27:1833–1844.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.
- Keyser C, Bouakaze C, Crubezy E, Nikolaev VG, Montagnon D, Reis T, Ludes B. 2009. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet.* 126:395–410.
- King RJ, Ozcan SS, Carter T, et al. (12 co-authors). 2008. Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. *Ann Hum Genet.* 72:205–214.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol.* 19:1737–1751.
- Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L, Reddy BM. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol.* 7:47.
- Lalueza-Fox C, Sampietro ML, Gilbert MT, Castri L, Facchini F, Pettener D, Bertranpetit J. 2004. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci.* 271:941–947.
- Li C, Li H, Cui Y, et al. (13 co-authors). 2010. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* 8:15.
- Li H, Wen B, Chen SJ, et al. (19 co-authors). 2008. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol Biol.* 8:146.
- Lipman JN. 1997. Familiar strangers: a history of Muslims in Northwest China. Hong Kong, China: Hong Kong University Press.
- Lopez-Parra AM, Gusmao L, Tavares L, Baeza C, Amorim A, Mesa MS, Prata MJ, Arroyo-Pardo E. 2009. In search of the pre- and post-Neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations. *Ann Hum Genet.* 73:42–53.
- Mair VH. 1995. Prehistoric Caucasoid corpses of the Tarim Basin. *J Indoeur Stud.* 23:281–370.
- Martinez L, Underhill PA, Zhivotovsky LA, Gayden T, Moschonas NK, Chow CE, Conti S, Mamolini E, Cavalli-Sforza LL, Herrera RJ. 2007. Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet.* 15:485–493.
- Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, Herrera RJ. 2009. Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur J Hum Genet.* 17:1260–1273.
- Nonaka I, Minaguchi K, Takezaki N. 2007. Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann Hum Genet.* 71:480–495.
- Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol.* 24:710–722.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M. 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet.* 120:334–353.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Stoneking M. 2007. Mating patterns amongst Siberian reindeer herders: inferences from mtDNA and Y-chromosomal analyses. *Am J Phys Anthropol.* 133:1013–1027.
- Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. 2006. Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered.* 61:132–143.
- Reidla M, Kivisild T, Metspalu E, et al. (43 co-authors). 2003. Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet.* 73:1178–1190.
- Rootsi S, Magri C, Kivisild T, et al. (45 co-authors). 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet.* 75:128–137.
- Rootsi S, Zhivotovsky LA, Baldovic M, et al. (20 co-authors). 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet.* 15:204–211.
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet.* 73:700–705.
- Semino O, Magri C, Benuzzi G, et al. (16 co-authors). 2004. 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.
- Semino O, Passarino G, Oefner PJ, et al. (17 co-authors). 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159.
- Sengupta S, Zhivotovsky LA, King R, et al. (15 co-authors). 2006. 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.
- Sharma S, Rai E, Bhat AK, Bhanwer AS, Bamezai RN. 2007. A novel subgroup Q5 of human Y-chromosomal haplogroup Q in India. *BMC Evol Biol.* 7:232.

- Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet.* 77:408–419.
- Shi H, Zhong H, Peng Y, et al. (13 co-authors). 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6:45.
- Shou WH, Qiao EF, Wei CY, Dong YL, Tan SJ, Shi H, Tang WR, Xiao CJ. 2010. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. *J Hum Genet.* 55:314–322.
- Su B, Xiao C, Deka R, et al. (11 co-authors). 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet.* 107:582–590.
- Su B, Xiao J, Underhill P, et al. (21 co-authors). 1999. 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.
- Tofanelli S, Ferri G, Bulayeva K, et al. (23 co-authors). 2009. J1-M267 Y lineage marks climate-driven pre-historical human displacements. *Eur J Hum Genet.* 17:1520–1524.
- Torroni A, Bandelt HJ, Macaulay V, et al. (33 co-authors). 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet.* 69:844–852.
- Underhill PA, Myres NM, Rootsi S, et al. (34 co-authors). 2009. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet.* 18:479–484.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet.* 65:43–62.
- Underhill PA, Shen P, Lin AA, et al. (21 co-authors). 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet.* 26:358–361.
- Wells RS, Yuldasheva N, Ruzibakiev R, et al. (27 co-authors). 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A.* 98:10244–10249.
- Wen B, Li H, Lu D, et al. (18 co-authors). 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302–305.
- Wen B, Xie X, Gao S, et al. (13 co-authors). 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet.* 74:856–865.
- Xu S, Huang W, Qian J, Jin L. 2008. Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am J Hum Genet.* 82:883–894.
- Xue Y, Zerjal T, Bao W, et al. (12 co-authors). 2006. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 172:2431–2439.
- Yang L, Tan S, Yu H, Zheng B, Qiao E, Dong Y, Zan R, Xiao C. 2008. Gene admixture in ethnic populations in upper part of Silk Road revealed by mtDNA polymorphism. *Sci China C Life Sci.* 51:435–444.
- Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol Biol Evol.* 21:2265–2280.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol.* 21:164–175.
- Zhang F, Su B, Zhang YP, Jin L. 2007. Genetic studies of human diversity in East Asia. *Phil Trans R Soc Lond B.* 362:987–995.
- Zhang F, Xu Z, Tan J, et al. (11 co-authors). 2009. Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang. *Am J Phys Anthropol.* 142:314–320.
- Zhivotovsky LA. 2001. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol Biol Evol.* 18:700–709.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, et al. (17 co-authors). 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 74:50–61.
- Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, Su B. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet.* 55:428–435.