ORIGINAL ARTICLE

Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia

Hua Zhong^{1,2,5}, Hong Shi¹, Xue-Bin Qi¹, Chun-Jie Xiao³, Li Jin⁴, Runlin Z Ma² and Bing Su¹

The regional distribution of an ancient Y-chromosome haplogroup C-M130 (Hg C) in Asia provides an ideal tool of dissecting prehistoric migration events. We identified 465 Hg C individuals out of 4284 males from 140 East and Southeast Asian populations. We genotyped these Hg C individuals using 12 Y-chromosome biallelic markers and 8 commonly used Y-short tandem repeats (Y-STRs), and performed phylogeographic analysis in combination with the published data. The results show that most of the Hg C subhaplogroups have distinct geographical distribution and have undergone long-time isolation, although Hg C individuals are distributed widely across Eurasia. Furthermore, a general south-to-north and east-to-west cline of Y-STR diversity is observed with the highest diversity in Southeast Asia. The phylogeographic distribution pattern of Hg C supports a single coastal 'Out-of-Africa' route by way of the Indian subcontinent, which eventually led to the early settlement of modern humans in mainland Southeast Asia. The northward expansion of Hg C in East Asia started ~40 thousand of years ago (KYA) along the coastline of mainland China and reached Siberia ~15 KYA and finally made its way to the Americas.

Journal of Human Genetics (2010) 55, 428–435; doi:10.1038/jhg.2010.40; published online 7 May 2010

Keywords: genetic divergence; Out of Africa; prehistoric migration; Y chromosome

INTRODUCTION

The Y-chromosome lineages in East Asian populations have been examined extensively. It has been shown that several dominant Y-chromosome haplogroups, such as O-M175, D-M174 and C-M130, and several relatively rare Y-chromosome haplogroups, such as F-M89, K-M9, P-M45 and N-M231, constitute the East Asian Y-chromosome gene pool.^{1–5} The ethnically diversified populations in East Asia have been suggested as the descendants of ancient modern humans of African origin, having a significant role in subsequent migrations into Siberia and the Americas.^{1,6} However, the migration routes of ancient modern humans into East Asia have long been debated, although two major routes have been proposed: the southern route and the northern route.^{1,6}

On the basis of the Y-chromosome lineage analysis, several research groups have attempted to elucidate the timing and the routes of the prehistoric migration of modern humans into East Asia. It is widely accepted that there is a genetic divergence between northern (NEAS) and southern (SEAS) East Asian populations.^{1–4} However, the relationship between NEAS and SEAS populations, and the cause of genetic divergence remain controversial.^{2–4} We have previously

suggested a southern origin for all East Asian populations based on the screening of 19 Y-chromosome single-nucleotide polymorphisms and a set of autosomal microsatellites in East Asian populations.^{3,7} Subsequently, an extended examination of Y-chromosome variation performed by Karafet *et al.*⁴ showed that NEAS populations have higher Y-chromosome diversity than do SEAS populations. Recently, Xue *et al.*² reported that the pooled Y-chromosome short tandem repeats (STRs) have a higher diversity in NEAS populations than in SEAS populations. Therefore, these two investigations of Y-chromosome diversity in East Asia suggested the potential existence of the northern route.

Through a detailed analysis of the expansion time and distribution pattern of one dominant Y-chromosome haplogroup in certain geographical regions, the timing and routes of the prehistoric migrations can be determined more objectively and the influence of recent population admixture can be avoided. This approach has been proven effective in inferring the prehistoric migrations of modern humans into Europe.^{8,9} Our previous study on Hg O3-M122 indicated a clear pattern of southern origin of this lineage and provided a solid evidence for the proposed southern route.¹⁰ Through detailed analysis

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology and Kunming Primate Research Centre, Chinese Academy of Sciences, Kunming, PR China; ²Center for Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, PR China; ³Human Genetics Centre, Yunnan University, Kunming, PR China; ⁴State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences, Fudan University, Shanghai, PR China and ⁵Graduate School, Chinese Academy of Sciences, Beijing, PR China

Correspondence: Dr RZ Ma, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 1 West Beichen Road, Chaoyang District, Beijing 100101, PR China. E-mail: rma@genetics.ac.cn or Dr B Su, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 East Jiaochang Road, Kunming 650223, PR China.

E-mail: sub@mail.kiz.ac.cn

Received 8 November 2009; revised and accepted 5 April 2010; published online 7 May 2010

of Hg N-M231, Rootsi *et al.*⁵ also detected the same migration route via Southeast Asia. The remaining question is whether the migrations of other haplogroups into East Asia followed the same route.

Hg C-M130 has a wide distribution across Asia^{1,2,4,11–27} and Oceania,^{12,15,23} less frequent in Europe^{11,13,16,28–31} and the Americas,^{26,32–34} and absent in Africa.^{11,18,35} As a non-African lineage, Hg C is highly informative in tracing the migration route of the African exodus in prehistory.⁶ However, when and where Hg C occurred, migrated and expanded is yet to be disclosed. At present, most of the archaeological and genetic evidence supports that the earliest African exodus went out of Africa via the Red Sea and then rapidly migrated to mainland Southeast Asia through the Indian coastline, and eventually reached Oceania.^{36–39} Recent Y-chromosome and mitochondrial DNA analysis in Australia and New Guinea has shown that Hg C is likely one of the earliest Out-of-Africa founder types,¹² which was also proposed in another study,⁶ and that mitochondrial DNA lineages consisting of the founder types (M and N) are dated to approximately 50–70 KYA.¹²

Given the early settlement in Oceania, it remains unknown whether modern humans migrated to mainland East Asia at the same time and when and by which route they expanded across East Asia. A previous study has suggested that Hg C migrated into East Asia via both the northern route and the southern route approximately 45–50 KYA.⁶ However, it was also suggested that Hg C in Central Asia had a Mongol origin.⁴⁰ On the other hand, the fossil records in East Asia indicate that the earliest record of modern humans was ~40 KYA.^{1,41} In addition, the inference based on the dental traits suggested that the earliest East Asians were the direct descendants of Southeast Asians and migrated into East Asia via the Sunda shelf.⁴² As an ancient haplogroup, Hg C could provide important clues to recover traces of the early colonization of Asia by anatomically modern humans.

MATERIALS AND METHODS

Samples

In this study, a total of 4284 unrelated males, including 4196 males from 134 East Asian populations and 88 males from 6 Southeast Asian populations (Figure 1 and Supplementary Figure 1), were recruited with informed consent. The protocol of this study was approved by the Institutional Review Board of the Kunming Institute of Zoology, Chinese Academy of Sciences. A total of 194 M130-derived Y chromosomes were extracted from the literature, and 10 M130-derived Australians typed in our previous project were included (Supplementary Table 1).

Y-chromosome genotyping

Using a hierarchical genotyping strategy,^{11,43} we first genotyped three Y-chromosome markers: M175, YAP and M130. The M130-derived individuals were then subjected to further typing of 12 biallelic markers, which define 13 subhaplogroups: C*-M130, C1-M8, C2-M38, C3*-M217, C3a-M93, C3b-P39, C3c-M48, C3d-M407, C3e-P53.1, C3f-P62, C4-M347, C5-M356 and C6-P55, the phylogenetic relationships of which are illustrated in Figure 1, according to the Y Chromosome Consortium (YCC 2002)⁴⁴ and Y Chromosomal Haplogroup Tree.⁴⁵ The genotyping primers were from the literature: M175, YAP, M130, M8, M38, M217, M93 and M48 from Underhill *et al.*;⁶ P39 from Zegura *et al.*;³² P53.1, P55 and P62 from Karafet *et al.*;⁴⁵ M407 and M356 from Sengupta *et al.*;¹⁴ and M347 from Hudjashov *et al.*¹² The biallelic markers were determined by sequencing PCR products, with the exceptions that the M130T allele was detected by PCR-restriction fragment length polymorphism (*Bsl* I digestion), M175 by running denatured PCR products on ABI 3730 and YAP by direct agarose electrophoresis of PCR products. To evaluate the phylogeographic structure of Hg C, we also typed eight commonly used Y-STR markers: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393 using fluorescence-labeled primers (obtained from Applied Biosystems, Foster

City, CA, USA) and then running denatured PCR products on ABI 3730. The Y-STR nomenclature follows the system proposed by Butler *et al.*⁴⁶

Data analyses

Together with the published data, the frequencies of M130 in worldwide populations are summarized in Figure 1 and applied to generate a contour map of frequency distribution (Figure 2) using the Surfer 7.0 software (Golden Software). The Y-STR data (Supplementary Table 1), including those from the literature,^{14,23–25,32–34,47–49} were used to construct the median-joining networks using the program NETWORK 4.5.0.7 (Fluxus Engineering),⁵⁰ and to calculate the average gene diversity and the R_{ST} genetic distances based on eight STR loci by Arlequin 3.01.⁵¹ Multidimensional scaling (MDS) analysis was performed based on the R_{ST} genetic distances using SPSS 15.0 (SPSS). The ages of STR variation and the divergence times of the Hg C subhaplogroups were estimated following Zhivotovskiy *et al.*, assuming an average Y-STR mutation rate of 0.00069 per locus per 25 years.^{8,14,52,53} The age of STR variation within a haplogroup reveals the time when variation occurred compared with a median haplotype in the given population; for the divergence time of a haplogroup, it represents the time when a subhaplogroup diverged from an ancestral haplogroup.

RESULTS AND DISCUSSION

Hg C is prevalent in various geographical areas (Figures 1 and 2), including Australia (65.74%), Polynesia (40.52%), Heilongjiang of northeastern China (Manchu, 44.00%), Inner Mongolia (Mongolian, 52.17%; Oroqen, 61.29%), Xinjiang of northwestern China (Hazak, 75.47%), Outer Mongolia (52.80%) and northeastern Siberia (37.41%). Hg C is also present in other regions, extending longitudinally from Sardinia¹³ in Southern Europe all the way to Northern Colombia,³² and latitudinally from Yakutia²⁴ of Northern Siberia and Alaska³² of Northern America to India, Indonesia and Polynesia, but absent in Africa.

As shown in Figure 1, most of the subhaplogroups of Hg C have a geographically pronounced distribution. Hg C6, which is defined by a recently identified marker,⁴⁵ was not detected in our samples. Hg C1 and C4 are completely restricted to Japan and Australia, respectively, and not detected in the other samples from East Asia and Southeast Asia. Hg C5 occurs in India and its neighboring regions Pakistan and Nepal.^{14,54} In mainland East Asia, four Hg C5 individuals were detected, including two in Xibe, one in Uyghur and one in Shanxi Han. Although the dispersal of Hg C2 is relatively wide, its distribution remains limited to Oceania and its neighboring regions, except Australia. In our samples, only three Hg C2 individuals were observed in Eastern Indonesia, which is consistent with previous reports.^{15,23} Hg C3 is the most widespread subhaplogroup, which was detected in Central Asia, South Asia, Southeast Asia, East Asia, Siberia and the Americas, but absent in Oceania. Different subhaplogroups of Hg C that do not overlap between the regions suggest that these individuals have undergone long-time isolation. As these subhaplogroups have a common origin by sharing the M130-derived allele, their geographical distributions enable us to infer the prehistoric migration routes of this lineage.

Hg C3 (defined by M217) can be further divided into six sub-branches: C3a-M93, C3b-P39, C3c-M48, C3d-M407, C3e-P53.1 and C3f-P62. As shown in the recent Y Chromosomal Haplogroup Tree (Figure 1), Hg C3* is the ancestral state at the M93, P39, M48, M407, P53.1 and P62 loci, and therefore presumably ancestral to the other Hg C3 sub-branches (Hg C3* may contain unidentified sub-branches, and therefore may not be a monophyletic group). Previous data have shown that Hg C3a and C3b were only detected in Japan¹¹ and North America,³² respectively. Hg C3c was detected in NEAS populations, Siberia and Central Asia.^{2,16,23–25,55} Hg C3d was detected only in a Yakut population.¹⁴ Hg C3* was detected in multiple regions,

including Southeast Asia, East Asia, Central Asia and Siberia. Unfortunately, in the published data, those Hg C3* individuals were not subtyped,^{2,4,6,14–16,23–25,54,55} and therefore it cannot be correctly assigned to the Hg C3 sub-branches.

In our samples, a total of 465 M130-derived Y chromosomes were identified (Figure 1 and Supplementary Table 1), and 430 of them were M217-derived (Hg C3) individuals, including 374 Hg C3* (all non-M93-, P39-, M48-, M407-, P53.1- and P62-derived individuals are assigned to the Hg C3* group in this study), 18 Hg C3c, 23 Hg C3d and 15 Hg C3e individuals. Hg C3a, C3b and C3f were not detected. As shown in Figure 1, the high frequencies of Hg C3* are observed in NEAS populations, including Inner/Outer Mongolians and Manchurian from Heilongjiang and Hazak (>30%). A total of 23 populations among the 31 NEAS tested have Hg C3* with frequencies >10%. Relatively low frequencies of Hg C3* are observed in SEAS populations. Only 9 populations out of the 47 SEAS have frequencies >10%, and Hg C3* is totally absent in 14 populations. As for Hg C3c and Hg C3e, they have similar distribution patterns and occur in Tibetan and Altaic populations with the exception of one Hg C3c individual and one Hg C3e individual detected in Heilongjiang Han and Gansu Han, respectively. Hg C3d is sparsely distributed in East Asian populations (Figure 1). In addition, there are 28 Hg C* individuals (Hg C* represents non-M8-, M38-, M217-, M347-, M356- and P55-derived individuals and is considered a potential ancestral haplogroup of the Hg C lineage in this study, although it may contain unidentified subclades), 7 in NEAS, 19 in SEAS and 2 in Southeast Asia (Figure 1). Combining the recently reported data,² Hg C* occurs from the southernmost to the northernmost in East Asia, but is more frequent in SEAS than in NEAS populations. Previous studies have shown that Hg C* might also exist in Central Asia.^{16,17} However, we believe that these Hg C* individuals should be Hg C3

because many sub-branch markers were not typed in the reported studies. This speculation is further supported by two lines of evidence. First, in Central Asia, all M130-derived individuals detected by Karafet *et al.*⁴ are M217-derived. Second, the assumed Hg C* individuals in Central Asia are shown to be the descendants of Mongols by subsequent Y-STR analysis.⁴⁰

The phylogeographic pattern of Hg C is consistent with the mitochondrial DNA evidence indicating rapid initial settlement, followed by prolonged isolation.³⁶ As shown in Figure 3a, most of the East Asian populations cluster together in the MDS plot, whereas other populations show separations from each other and have relatively large genetic distances, especially the Japanese-specific Hg C1 being clearly an outlier in the MDS plot. Interestingly, besides Hg C1, Japanese also have M217-derived individuals who have a close relationship with the Han Chinese (Figures 3a and b), rather than with the Altaic-speaking populations. Therefore, the two distinctive sets of Hg C lineages in Japan support the hypothesized two independent migration waves to Japan,²³ that is, the Paleolithic migration and the Neolithic migration likely due to the demic diffusion of the Han culture.⁵⁶ The Hg C5 sublineage in India is also distinctive in the MDS plot, but with relatively short genetic distances with the East Asian populations. As expected, Australians and Austronesians are clustered together and are relatively close to SEAS, including Hmong-Mien-, Daic- and Austro-Asiatic-speaking populations. Native Americans and Siberians are close in the MDS plot with short genetic distance with the Altaic-speaking populations, which can also be reflected when only analyzing the Hg C3 sublineage (Supplementary Figure 2).

To estimate STR gene diversity, we grouped the populations based on geographical regions and language families (Supplementary Table 2). A general east-to-west and south-to-north cline was observed. The Austronesian group has the highest diversity (0.582),

Figure 2 Frequency distribution of Hg C in worldwide populations and the inferred migration routes of the African exodus carrying the M130 mutation in prehistory.

Figure 1 The hierarchical phylogenetic relationships and distribution frequencies of Hg C and its subhaplogroups. In the Y-chromosomal haplogroup tree, Hg C2 is the combination of Hg C2* individuals and M208-derived individuals. Hg C4 includes Hg C4*, Hg C4a and Hg C4b. ^aThis study; ^bAustro-Asiatic-speaking populations; ^cAustronesian-speaking populations; ^dDaic-speaking populations; ^eHmong-Mien-speaking populations; ^fTibeto-Burman-speaking populations; ^gAltaic-speaking populations; ^hSouthern and Northern East Asia are geographically separated by Yangtze River; '—' indicates no available data.

Figure 3 The MDS plots. Populations in (a) are grouped according to geographic distributions and language families and include all M130-derived individuals. Their detailed information can be obtained from Supplementary Table 1. Populations in (b) include only Hg C3* individuals.

followed by Australian (0.545), Hainan aborigines (0.522) and Southern Han (0.508). In contrast, Siberian, Native American, Tibeto-Burman and Altaic groups show relatively low diversities (0.251, 0.359, 0.317 and 0.371, respectively). Hence, in combination with the above analysis of the MDS plot (Figure 3a), the STR diversity pattern (Supplementary Table 2) suggests that Southeast Asia might be the cradle land of the M130 lineage, and that the M130 lineage, derived from the M168 ancestral type (the shared marker in non-Africans),⁶ first migrated into mainland Southeast Asia by way of the Indian subcontinent, and then into Australia and mainland East Asia

separately. After its settlement in Southeast Asia in prehistory, the M130 lineage probably experienced a population expansion as reflected by the high STR diversity. It then began to migrate northward via the coastline, and gradually settled in southern and northern East Asia, then northeast Siberia, and finally into the Americas via Beringia.

The M217-derived (Hg C3) lineages are informative in revealing the eastward migration of modern humans into East Asia in prehistory because of its extensive distribution in East Asia, Central Asia and Siberia. It was suggested that the M217-derived individuals first reached South Asia and then started migrating eastward through two

routes: Central Asia and Southeast Asia.⁶ However, the Central Asian M217-derived individuals were shown having a recent Mongol origin (~1000 years ago).⁴⁰ The Han Chinese display a high STR diversity (Supplementary Table 2 and Figure 4), especially those in the eastern coastal region (0.467) as well as other eastern populations (Korean, 0.463; Japanese, 0.453), whereas populations in the north and west show low diversities (Altaic, 0.281; Tibetan, 0.366). Therefore, the distribution and gene diversities of the M217-derived lineages support a single eastward migration through the southern route and the subsequent northward migration of Hg C along the coastline of mainland East Asia in prehistory. The evidence from dental morphological traits pointed to the same direction.⁴²

The sub-branches (namely Hg C3c, C3d and C3e) of Hg C3 in East Asia can also tell the pattern of prehistoric migrations of regional populations. Hg C3c is restricted to Altai-speaking populations with only sporadic appearance in Northern Han Chinese (one individual), Tibetan (four individuals) and Japanese (three individuals) (Figure 1). Among the 82 Hg C3c individuals identified, 76 of them (92.7%) share a 9-repeat motif at the DYS391 locus. The median-joining network of Hg C3c (Figure 4) indicated a star-like/short-distanced network, implying that Hg C3c has a relatively recent origin. Hg C3d was detected in NEAS and SEAS populations (Figure 1), but it is more prevalent in NEAS. Moreover, the Y-STR diversity of Hg C3d is higher

in NEAS (0.313) than in SEAS (0.198). As shown in the median-joining network (Figure 4), the Hg C3d individuals in NEAS have more STR haplotypes than those in SEAS, suggesting that Hg C3d likely occurred in NEAS and then expanded to SEAS recently due to the demic diffusion of the Han culture.⁵⁶ Hg C3e was detected only in NEAS (Figure 1) with a low STR diversity (Figure 4), suggesting its recent origin in NEAS.

Table 1 The estimated ages of STR variation and the divergence times of the Hg C subhaplogroups

Subhaplogroup and sample size	Age ^a of STR variation	Divergence time ^a
C* (28)	5.5 ± 1.6	—
C1-M8 (20)	10.0 ± 3.5	41.9 ± 16.6
C3*-M217 (386)	18.9 ± 4.0	32.6 ± 14.1
C3b-P39 (41)	13.2 ± 4.1	14.5 ± 5.1
C3c-M48 (13)	10.8 ± 2.3	9.3 ± 3.3
C3d-M407 (23)	9.3 ± 2.7	12.0 ± 4.2
C3e-P53.1 (15)	4.8 ± 2.3	14.4 ± 5.1
C5-M356 (16)	14.2 ± 3.3	33.3 ± 19.1

Abbreviation: STR, short tandem repeat.
^aIn thousands of years ago (± s.e.).

Figure 4 The median-joining networks of Y-STR haplotypes within subhaplogroups of Hg C. The network of Hg C3* was constructed by the median-joining method after weighting STRs according to their repeat number variances and processing the data using the reduced median method. The sizes of the nodes are proportional to their frequencies. The lengths of the lines are proportional to the mutational steps.

On the basis of STR data, we estimated the ages of STR variation and the divergence times of Hg C subhaplogroups (Table 1). In general, the times estimated are highly consistent with the inferred migration events. The divergence times of C3*-M217 and C1-M8 were estimated as 32.6 ± 14.1 KYA and 41.9 ± 16.6 KYA, respectively, indicating that the proposed eastward migration of Hg C into East Asia started about 32–42 KYA. This is consistent with the mitochondrial DNA findings, in which the Japanese- and Korean-specific Haplogroup M7a was estimated as 37.0 ± 20.0 KYA.⁵⁷ In addition, the archaeological findings also provided strong evidence that an Upper Paleolithic wave of migration brought people into Japan more than 30.0 KYA.^{58,59} At that time, Pleistocene land bridges likely connected Japan to the mainland and there was a much shorter coastline between East Asia and Southeast Asia.³⁸ However, the STR-variation ages for Hg C3* and C1 were estimated as 18.9 ± 4.0 KYA and 10.0 ± 3.5 KYA, respectively, reflecting relatively recent population expansions, which is reasonable because this is the time that the Last Ice Age started to retreat and the climate became warmer.⁶⁰ Another ancient lineage is Hg C5 (33.3 ± 19.1 KYA), and its divergence time agrees well with the suggested midway station of the Indian subcontinent during the eastward migration of Hg C from Africa to East Asia. Similar to Hg C1 and Hg C3*, the STR-variation age of Hg C5 also reflects recent population expansion time (14.2 ± 3.3 KYA), which is a bit younger than the reported age by Sengupta *et al.*¹⁴ As expected, the subbranches of Hg C3 are young, which are consistent with the proposed later migration events associated with these sublineages. For example, M48-derived individuals have the highest Y-STR diversity (0.384) in NEAS but with a young age of ~ 10 KYA (Table 1). We believe that M48 originated in NEAS populations, which agrees well with the suggested recent migration (for example, the Mongol expansion) of M48-derived individuals into Central Asia and Siberia.^{24,40}

It should be noted that the estimated age is not necessarily always a reliable indicator of the founding date of a lineage/population. The STR-variation age of Hg C* is surprisingly young (5.5 ± 1.6 KYA), which seems to contradict the assumed ancestral status of Hg C*. As shown in Figure 4, the STR haplotypes of Hg C* form a star-like network and the mutational steps are short. There are two possible explanations. One is that there might be other unidentified young sublineages under Hg C*. The other would invoke an ancient bottleneck-related genetic drift or natural selection. In addition, the relatively small sample size of Hg C* may also cause the underestimation. However, we tend to believe that the Hg C* individuals detected in this study are the genetic footprints of the ancient lineage because they not only have a very wide distribution (although low frequency) but also have similar STR haplotypes (Figure 4 and Supplementary Table 1). Finally, Hg C3*, C1 and C5 discussed in this study possibly contain unidentified subclades; therefore, further studies are required for a well-resolved phylogeny and detailed phylogeographic inferences.

CONCLUSIONS

We demonstrated the phylogeographic distribution of one of the most ancient non-African Y-chromosome lineages, from which we inferred the prehistoric migration and expansion of the Hg C lineage. We propose that Hg C was derived from the African exodus and gradually colonized South Asia, Southeast Asia, Oceania and East Asia by a single Paleolithic migration from Africa to Asia and Oceania, which occurred more than 40 KYA. The prehistoric northward migration of Hg C in mainland East Asia likely followed the coastline and is consistent with the northward migration of other East Asian Y-chromosome haplogroups.

ACKNOWLEDGEMENTS

We are grateful to all the voluntary donors of DNA samples in this study. We thank Tatiana M Karafet, Li Hui, Sanghamitra Sengupta and Brigitte Pakendorf for providing their published STR data of Hg C. We thank Pingping Tan and Hui Zhang for their technical assistance. This study was supported by grants from the National 973 project of China (2006CB701506, 2007CB947701), the Chinese Academy of Sciences (KSCX1-YW-R-34), the National Natural Science Foundation of China (30413242, 30525028, 30700445, 30630013 and 30771181), and the Natural Science Foundation of Yunnan Province of China (2007C100M). We also thank the anonymous reviewers for their insightful comments and suggestions.

- Jin, L. & Su, B. Natives or immigrants: modern human origin in East Asia. *Nat. Rev. Genet.* **1**, 126–133 (2000).
- 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).
- Su, B., Xiao, J., Underhill, P., 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).
- Karafet, T., 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).
- Roots, S., Zhivotovsky, L. A., Baldovici, 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).
- Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, 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).
- Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T. *et al.* Genetic relationship of populations in China. *Proc. Natl Acad. Sci. USA* **95**, 11763–11768 (1998).
- Roots, 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).
- 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).
- 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).
- 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).
- Hudjashov, G., Kivisild, T., Underhill, P. A., Endicott, P., Sanchez, J. J., Lin, A. A. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl Acad. Sci. USA* **104**, 8726–8730 (2007).
- 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).
- Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E. *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).
- Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A. *et al.* Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol. Biol. Evol.* **23**, 2234–2244 (2006).
- 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).
- Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* **71**, 466–482 (2002).
- Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C. *et al.* The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* **74**, 532–544 (2004).
- Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L. *et al.* Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* **114**, 127–148 (2004).
- 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 the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* **28**, 458–472 (2003).
- 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).

- 22 Jin, H. J., Kwak, K. D., Hammer, M. F., Nakahori, Y., Shinka, T., Lee, J. W. *et al.* Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum. Genet.* **114**, 27–35 (2003).
- 23 Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. *et al.* Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* **51**, 47–58 (2006).
- 24 Pakendorf, B., Novgorodov, I. N., Osakovskij, V. L., Danilova, A. P., Protod'jakonov, A. P. & Stoneking, M. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum. Genet.* **120**, 334–353 (2006).
- 25 Pakendorf, B., Novgorodov, I. N., Osakovskij, V. L. & Stoneking, M. Mating patterns amongst Siberian reindeer herders: inferences from mtDNA and Y-chromosomal analyses. *Am. J. Phys. Anthropol.* **133**, 1013–1027 (2007).
- 26 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).
- 27 Regueiro, M., Cadenas, A. M., Gayden, T., Underhill, P. A. & Herrera, R. J. Iran: tricontinental nexus for Y-chromosome driven migration. *Hum. Hered.* **61**, 132–143 (2006).
- 28 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).
- 29 Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogvali, E. L. *et al.* The western and eastern roots of the Saami—the story of genetic 'outliers' told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* **74**, 661–682 (2004).
- 30 Passarino, G., Cavalleri, G. L., Lin, A. A., Cavalli-Sforza, L. L., Borresen-Dale, A. L. & Underhill, P. A. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *Eur. J. Hum. Genet.* **10**, 521–529 (2002).
- 31 Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A. V., Krumina, A., Kucinskas, V. *et al.* Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol. Biol. Evol.* **18**, 1077–1087 (2001).
- 32 Zegura, S. L., Karafet, T. M., Zhivotovsky, L. A. & Hammer, M. F. 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 (2004).
- 33 Bolnick, D. A., Bolnick, D. I. & Smith, D. G. Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol. Biol. Evol.* **23**, 2161–2174 (2006).
- 34 Malhi, R. S., Gonzalez-Oliver, A., Schroeder, K. B., Kemp, B. M., Greenberg, J. A., Dobrowski, S. Z. *et al.* Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. *Am. J. Phys. Anthropol.* **137**, 412–424 (2008).
- 35 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).
- 36 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
- 37 Forster, P. & Matsumura, S. Evolution: did early humans go north or south? *Science* **308**, 965–966 (2005).
- 38 Stringer, C. Palaeoanthropology. Coasting out of Africa. *Nature* **405**, 24–25, 27 (2000).
- 39 Walter, R. C., Buffler, R. T., Bruggemann, J. H., Guillaume, M. M., Berhe, S. M., Negassi, B. *et al.* Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* **405**, 65–69 (2000).
- 40 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).
- 41 Trinkaus, E. Early modern humans. *Annu. Rev. Anthropol.* **34**, 207–230 (2005).
- 42 Turner, C.G. II Major features of Sundadonty and Sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian aboriginals. *Am. J. Phys. Anthropol.* **82**, 295–317 (1990).
- 43 Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
- 44 Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
- 45 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).
- 46 Butler, J. M., Schoske, R., Vallone, P. M., Kline, M. C., Redd, A. J. & Hammer, M. F. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int.* **129**, 10–24 (2002).
- 47 Nonaka, I., Minaguchi, K. & Takezaki, N. Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann. Hum. Genet.* **71**, 480–495 (2007).
- 48 Li, D., Li, H., Ou, C., Lu, Y., Sun, Y., Yang, B. *et al.* Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS ONE* **3**, e2168 (2008).
- 49 Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y. *et al.* Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol. Biol.* **8**, 146 (2008).
- 50 Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
- 51 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
- 52 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).
- 53 Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, 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).
- 54 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 (2007).
- 55 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).
- 56 Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y. *et al.* Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
- 57 Kivisild, T., Tolk, H. V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H. J. *et al.* The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19**, 1737–1751 (2002).
- 58 Takamiya, H. & Obata, H. Peopling of Western Japan, focusing on Kyushu, Shikoku, and Ryukyu Archipelago. *Radiocarbon* **44**, 495–502 (2002).
- 59 Ono, A., Sato, H., Tsutsumi, T. & Kudo, Y. Radiocarbon dates and archaeology of the late Pleistocene in the Japanese Islands. *Radiocarbon* **44**, 477–494 (2002).
- 60 Forster, P. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Phil. Trans. R. Soc. Lond. B* **359**, 255–264 (2004).

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