Y-chromosome haplotype distribution in Han Chinese populations and modern human origin in East Asians

KE Yuehai (柯越海)[1], SU Bing (宿 兵)[1−3], XIAO Junhua (肖君华)[1], CHEN Hua (陈 华)[1], HUANG Wei (黄 薇)[4], CHEN Zhu (陈 竺)[4], CHU Jiayou (褚嘉佑)[5], TAN Jiazhen (谈家桢)[1], JIN Li (金 力)[1,3] & LU Daru (卢大儒)[1]

1. Institute of Genetics and Morgan-Tan International Center for Life Science, Fudan University, Shanghai 200433, China;
2. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;
3. Human Genetics Center, University of Texas-Houston, Houston 77225, USA;
4. National Human Genome Center at Shanghai, Shanghai 201203, China;
5. Institute of Medical Biology, Chinese Academy of Medical Sciences, Kunming 650031, China

Correspondence should be addressed to Lu Daru (email: drlu@fudan.edu.cn)

Received May 15, 2000

Abstract We investigated the distribution of Y-chromosome haplotype using 19 Y-SNPs in Han Chinese populations from 22 provinces of China. Our data indicate distinctive patterns of Y chromosome between southern and northern Han Chinese populations. The southern populations are much more polymorphic than northern populations. The latter has only a subset of the southern haplotypes. This result confirms the genetic difference observed between southern and northern ethnic populations in East Asia. It supports the hypothesis that the first settlement of modern humans of African origin occurred in the southern part of East Asia during the last Ice Age, and a northward migration led to the peopling of northern China.

Keywords: Y-chromosome haplotype, SNPs, modern human origin, emigration, Chinese.

East Asia is regarded as the key region for studying modern human origin and migration since it is the region connecting directly or indirectly with other continents. A planthera of genetic evidence favors the hypothesis that modern human originated in Africa and dispersed into other parts of the world. While the expansions of modern human into Europe, the America and Oceania are now relatively well characterized, little is known of the earliest migratory routes by which modern humans spread from West Asia to East Asia[1−6].

Fossil records excavated in China and other regions of East Asia seem to show a continuous *in situ* evolution process from *Homo erectus* to *Homo sapiens sapiens*. This has been used as the strong evidence to support the hypothesis of independent origin of automatically modern human (*Homo sapiens sapiens*) in China although this claim itself is controversial among paleonanthropologists[7−12]. A recent study of Chinese populations using 30 microsatellites markers questioned the validity of the *in situ* origin hypothesis. Chu et al. (1998) suggested that modern humans in East Asia were originally from Africa, and southern and northern Chinese populations have a re-

cent common ancestor though genetic differences exist between them[13].

However, the study of Chu et al. fell short of providing unequivocal evidence in delineating migratory routes of modern humans during the historic peopling of China due to the high mutation rates associated with the microsatellite markers employed. The microsatellites, as the genetic markers for studying human evolution and population migration, have some limitations, especially for tracing deep evolution history. However, the availability of a large amount of Y chromosome SNP markers recently discovered triggered another wave of genetic study of human history. The biallelic markers are usually single base changes. The markers on the non-recombinant part of the Y-chromosome allow the reconstruction of intact haplotypes, which are not likely to be eroded by recombination and recurrent mutation. Therefore, Y-chromosome biallelic markers are more stable than microsatellite and highly informative for tracing ancient human migrations. Recognizing the fact that Y chromosomes have a smaller effective population size compared to autosomes, those Y chromosome polymorphic markers are probably the best genetic tool of studying early human migrations as bottleneck events which are often associated with such migrations become more pronounced[14—21]. In this study, 19 Y chromosome SNPs (single nucleotide polymorphism) were employed to study the genetic structure of Han Chinese population. The Y-chromosome haplotype distribution in extant Chinese population was used to reconstruct ancient migration patterns within China. In addition, the analysis of Y-chromosome microsatellites allows us to estimate the initial settlement of modern humans in East Asia.

1 Material and methods

1.1 Samples collection

A total of 362 Han Chinese samples were collected in 22 provincial areas whose geographical origins were assigned according to the birthplaces of their four grandparents. Those samples were further categorized into 16 geographic regions given small samples sizes in some provinces. Genome DNA was extracted using the standard Phenol-chloroform methods. DNA was stored at –20℃ after extraction.

1.2 Genotyping of Y-SNPs and microsatellites (STR)

A total of 19 Y chromosome biallelic loci were screened, seven of them were from the previous paper[17], and the other 12 SNPs were from our recent reports[22]. An allele specific PCR assay was employed to type Y chromosome biallelic loci. For each Y locus, the allelic-specific primers were designed to recognize two different alleles at this locus. The 10 μL PCR system includes: 20 ng/μL DNA template, 2 pmol of each primer, 2.5 mmol/L 4×dNTP, 1 μL of 10×PCR buffer, and 0.5 unit of Taq DNA polymerase. The primer sequence of 14 loci and the PCR conditions are listed in table 1.

Three Y chromosome microsatellites loci, DYS389, DYS390 and DYS391 were also typed using GeneScan (ABI). The PCR system includes: 0.5 μL of 10×PCR buffer, 3.0 mmol/L Mg^{2+},

0.6 μL of 4×0.2 mmol/L dNTPs, 0.2 unit Taq DNA polymerase, 0.05 μL of each primer and 1 μL of 20 ng/ μL DNA template. The ddH_2O was added up to the final volume of 5 μL. A touch down PCR assay was applied for amplification. After PCR, products were screened by the Genescan with an ABI373 DNA Sequencer (Perkin-Elmer). The Genotyper^{Tm} software was used for allele calling[18].

Table 1 The sequence of primers for 16 Y chromosome SNPs and amplification condition

Locus		Primer sequence	Mg^{2+} concentration	Anneal
M122	forward primer	(1) AAT TGA GAT ACT AAT TCA C	6 mmol/L	50℃
	forward primer	(2) AAT TGA GAT ACT AAT TCA T		
	reverse primer	AAA ACT TTA TCA TAT TGA G		
M120	forward primer	GAG CTT GGA CTT TAG GAC GG	4 mmol/L	60℃
	reverse primer	(1) TTT CCC TTA AAA ACA GCA TGG		
	reverse primer	(2) TTT CCC TTA AAA ACA GCA TGA		
M119	forward primer	ACT CAC CCT AAG GAA GTC ACG A	4 mmol/L	60℃
	reverse primer	(1) CCA ATT CAG CAT ACA GGC G		
	reverse primer	(2) CCA ATT CAG CAT ACA GGC T		
M7	forward primer	TGG ATT AGT CAC TTT AGA CCT AC	4 mmol/L	60℃
	reverse primer	(1) GTA GTT GAG TTA CTG TTC TTC		
	reverse primer	(2) GTA GTT GAG TTA CTG TTC TTG		
M50	forward primer	(1) GCC TAC CCA AAC CAC ACC T	4 mmol/L	60℃
	forward primer	(2) GCC TAC CCA AAC CAC ACC C		
	reverse primer	GGA AGC CTG TGT CTA CTC TGC		
M110	forward primer	AAA AGG CTG TGG TGC TGA TC	4 mmol/L	60℃
	reverse primer	(1) AAT GCA ACA GTT TAC AAG AAC ATG		
	reverse primer	(2) AAT GCA ACA GTT TAC AAG AAC ATA		
M111	forward primer	(1) AGA ACA AGT TCT GTT TTT CAC ATT G	4 mmol/L	54℃
	forward primer	(2) AGA ACA AGT TCT GTT TTT CAC AG		
	reverse primer	AAG AGA TGA AGA TAC CTT ATA TGC CC		
M89	forward primer	AGA AGC AGA TTG ATG TCC C	4 mmol/L	60℃
	reverse primer	(1) TCA GGC AAA GTG AGA GAT G		
	reverse primer	(2) TCA GGC AAA GTG AGA GAT A		
M88	forward primer	(1) CTT ATT CCT GCT TCT TCT GCG	4 mmol/L	58℃
	forward primer	(2) CTT ATT CCT GCT TCT TCT GCA		
	reverse primer	AGG TGT GAC CAC AGA GAC TCA G		
M103	forward primer	CTC CAA GGA CAC AGA ACA GG	4 mmol/L	58℃
	reverse primer	(1) CAA GAC CAG AGA AGG TGG G		
	reverse primer	(2) CAA GAC CAG AGA AGG TGG A		
M45	forward primer	(1) GGC AGT GAA AAA TTA TAG ATA G	4 mmol/L	58℃
	forward primer	(2) GGC AGT GAA AAA TTA TAG ATA A		
	reverse primer	ACC TTC CAC AGA CCA CAG		
M9	forward primer	(1) GCC TAA GAT GGT TGA ATC	4 mmol/L	54℃
	forward primer	(2) GCC TAA GAT GGT TGA ATG		
	reverse primer	CTC AAG CGT AAA TGT ACT GT		
M95	forward primer	(1) GAT AAG GAA AGA CTA CCA TAT TAG TGC	4 mmol/L	62℃
	forward primer	(2) GAT AAG GAA AGA CTA CCA TAT TAG TGT		
	reverse primer	GGG TGG GTG TGT TTG AAG G		
M17	forward primer	(1) TTG CTG GTT GTT ACG GGG	4 mmol/L	52℃
	forward primer	(2) GTTG CTG GTT GTT ACG GGT		
	reverse primer	GCT ATT CTT GTT TCT CCA GGC		

Among them, the M15 Taq DNA polymerase is ordinary Taq; all the others are Biolase Platinum DNA polymerase (Bioline). After PCR, 2 μL products were visualized through 4% agarose gel electrophoresis.

1.3 Statistical method

For estimating the age of M112C haplotypes in Han Chinese, the equation $t=-N_e\ln(1-V/N_e\mu)$ was used. It was based on the single step mutation model for a haploid population assuming constant population size, where N_e is the effective population size, V is the variance of repeat numbers of Y chromosome microsatellite in the population, and μ is the mutation rate of the microsatellite loci.

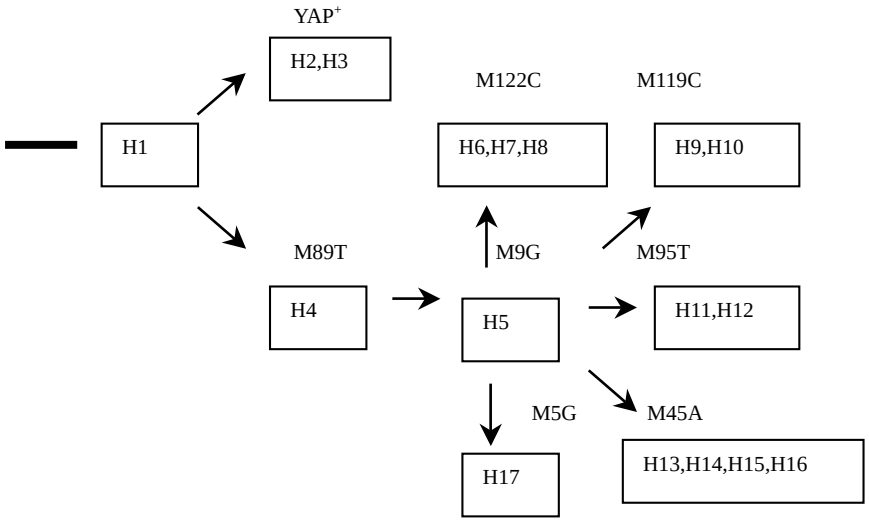
2 Result

In all the individuals studied for the 19 Y-chromosome biallelic markers, 17 haplotypes were obtained (table 2). The frequency distribution of Y haplotypes in all the populations in Han Chinese populations and other areas is listed in table 3 (Su et al. 1999), where African includes *Lissongo*, *Biaka Pygmy* and *Mbuti Pygmy*, American Indian includes *Kartiana*, *Guavm* and *Trapa*, European includes *Italian* and northern Europeans, and Oceanian includes *Nasioi Melanesian* and *New Guinea Highlander*.

The H1 and H2 haplotypes are relatively ancient, appearing in both African and non-African populations, implying that the occurrence of the mutations defining those haplotypes preceded the initial migration of modern humans out of Africa. All of Africans carry the C allele at locus M9 that has a C to G mutation, but the frequency of the M9G allele holds 84.3% in non-African populations. This suggests that most of non-African population have relatively close relationship and recent common ancestral Y chromosome. H5 appears to be the common ancestor of all the other haplotypes (H6—H17) which are regionally distributed and probably arose after the out-of-Africa migration[17]. H15 and H17 are American Indian and Oceanian specific haplotypes, respectively, H14 holds high frequency in European population, and H6—H13 only exist in East Asia population. Therefore, those regionally specific haplotypes provide an excellent tool of tracing population migrations and detecting gene flows among continental populations. Of the eight aforementioned Asia-specific haplotypes, H6, H7 and H8 share a T to C mutation at locus M122. They are the predominant haplotypes in most of the East Asian populations studied, particularly in Han Chinese (54.1% on the average) and are absent in non-Africans. When we compared the frequency distributions of Y haplotypes between southern and northern Han Chinese populations with the boundary of Yangtze River, it showed that southern populations are more polymorphic than northern populations. The difference shows as follows: (ⅰ) The number of East Asian specific haplotypes in southern populations is much more than those in northern populations. The haplotypes found in the northern populations consist of only a subset of those found in the southern populations. For example, H7, H10, H11 and H12 can only be found in the southern Han populations but absent in the northern Han populations. This comparison becomes more pronounced by comparing non-Han ethnic populations. This contrast suggests that the south-north gene flows are more frequent in Han populations than that in non-Han ethnic populations. (ⅱ)

Southern populations are much more polymorphic than northern populations. The gene diversity values are 62.4% for northern non-Han populations, 71.5% for northern Han populations, 76.1% for southern Han populations, and 81.3% for southern non-Han populations respectively. This pattern supports the hypothesis that modern human entered into East Asia from the south[22]. Interestingly, data from mtDNA are also consistent with this hypothesis[23].

Table 2 17 Y chromosome haplotype systematic evolution relationship



H1 is ancestral haplotype, other haplotypes derive from H1, and every haplotype specific SNPs type is shown on the pane. Each haplotype with 19 biallelic is listed in our previous report[20].

To estimate the time of the entry of modern human into China, we typed three Y chromosome microsatellites loci (DYS391, DYS390 and DYS389) for individuals carrying the M112C allele shared by the East Asian-specific haplotypes H6, H7 and H8. To minimize the possible influence of population substructure on the estimation, only 160 Han Chinese samples with M122C were screened. As mentioned in the method, the equation $t=-N_e\ln(1-V/N_e\mu)$ was used. t can show that this formula is still approximately valid even if the population went throughout a strong bottleneck followed by a rapid population expansion. The widely accepted estimation of the effective population size of modern humans is 5000—10000. Given a relatively small genetic diversity in Asian populations compared with that seen in Africans, a value of 750—2000 should be a reasonable assumption. The mutation rate of 0.18% was used according to the published data[24,25]. The number of generations estimated is 919—3032 for DYS390, the oldest among all the three estimations. Therefore, the age of M122C is approximately 18000—60000 years assuming a 20-year generation time[22]. The estimated age of the initial entry of modern humans into East Asia is consistent with the morphological and archeological studies[26—29]. The knowledge on archeology shows that the so-called Sinodont dentition in northern Asian people, including the whole

Table 3 The frequency distribution of Y chromosome haplotypes in Han Chinese populations and other populations in the world (Shadows cover the East Asia specific haplotype)

Region	Population	Size	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17
Northeast Asia	Buryat	4	75.0			25.0													
	Ewenki	8	50.0				12.5	12.5		25.0									
	Manchurian	18	16.7			11.1	22.2	27.8		16.7	5.6								
	Mongolian	24	58.3	4.2		8.3	12.5	4.2		4.2	4.2							4.2	
	Tibetan	8		12.5	25.0	12.5				50.0									
Northern Han	Liaoning	6	16.7				50.0				16.7				16.7				
	Hebei	6					16.7	50.0		33.3									
	Shangdong	32	9.4			3.1	18.8	28.1		28.1	9.4				3.1				
	Henan	28	7.1			3.6	25.0	32.1		14.3	10.7				7.1				
	others	10	10.0				10.0	30.0		40.0	10.0								
Southern Han	Anhui	22	13.6				18.2	27.3		18.2	18.2				4.6				
	Zhejiang	50	12.0				6.0	24.0		26.0	26.0		6.0						
	Jiangsu	55	7.3	1.8		1.8	18.2	23.6	3.6	21.8	16.4		3.6		1.8				
	Shanghai	30	6.7				16.7	26.7	3.3	16.7	26.7				3.3				
	Hubei	18	5.6				11.1	27.8	5.6	33.3	16.7								
	Sichuan	14	7.1				28.6	7.1	7.1	35.7	7.1		7.1						
	Jiangxi	21	4.8			9.5	19.1	23.8		23.8	14.3		4.8						
	Hunan	15					13.3	33.3		26.7	13.3		13.3						
	Fujian	13	7.7				7.7	38.5		38.5				7.7					
	Yunnan	27	11.1			3.7	3.7	18.5		55.6	3.7		3.7						
	Guangdong	15						40.0		26.7	20.0			6.7	6.7				
Southeast Asia	Zhuang	28	3.6		3.6	7.1	3.6	3.6		25.0	17.9		25.0	10.7					
	Cambodian	26	3.8		3.8	11.5	11.5	3.8		15.4	3.8	3.8	23.1	11.5		3.8		3.8	
	Thailand	20				5.0	5.0	5.0	5.0		5.0	5.0	45.0	20.0		5.0			
	Malayan	13				7.7	7.7	30.8		15.4	7.7	23.1	7.7						
	Batak	18	5.6			5.6	11.1	11.1	16.7		22.2		27.8						
	Javan	11	9.1			9.1	27.3	9.1			18.2	9.1	18.2						
Other continents	African	24	20.8	79.2															
	American Indian	26														3.8	96.2		
	European	39	10.3			12.8	25.6									51.3			
	Oceanian	93	15.1			2.2	41.9												40.9

northern Chinese population, occurred some 18000—25000 years ago. The similar dentition pattern predominates among all of the Southeast Asian population and was thought to be ancestral to the Sinodont pattern. Consequently, this evidence of dental evolution tends to rule out an 18000-years colonization dating, the lower bound of our age estimation. In addition, archeological evidence from the Altai Mountain and Lake Baikal regions of southeastern Siberia is beginning to show the presence of modern human lithic cultures dating from 25000—45000 years ago. Hence, when we admit the fact that the Southeast Asia mainland is the homeland for all East Asian population, including Siberian and Oceanian, the first entry of East Asia should predate the emergence of the lithic culture in northern Asia. Therefore, the upper bound of 60000 years seems to be in a more acceptable value of the entry of modern human.

In conclusion, followed by a northward migration coinciding with glaciers receding in that area, the first entry of modern humans into the southern China occurred about 60000 years ago, and then a subpopulation made their arduous journey to the north, which results in peopling of northern China and then Siberia.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 39993420).

References

1. Cann, R. L., Stoneking, M., Wilson, A. C., Mitochondria DNA and human evolution, Nature, 1987, 325: 31—36.
2. Vigilant, L., Stoneking, M., Harpending, H. et al., African populations and the evolution of human mitochondrial DNA, Science, 1997, 253: 1503—1507.
3. Cavalli-Sforza, L. L., Piazza, M. P., The History and Geography of Human Genes, Princeton: Princeton University Press, 1994.
4. Brooks, A. S., Wood, B., Paleoanthropology, The Chinese side of the story, Nature, 1990, 344: 288—289.
5. Li, T., Etler, D. A., New middle Pleistocene hominid crania from Yunxian in China, Nature, 1992, 357: 404—407.
6. Wu, X. Z., Poirier, F. E., Human Evolution in China, Oxford: Oxford University Press, 1995.
7. Etler, D. A., The fossil evidence for human evolution in Asia, Annu. Rev. Anthropol., 1996, 25: 275—301.
8. Wolpoff, M. H., Interpretations of multiregional evolution, Science, 1996, 274: 704—707.
9. Stringer, C. B., Andrew, P., Genetic and fossil evidence for the origin of modern humans, Science ,1988, 239: 1263—1268.
10. Wilson, A. C.,Cann, R. L., The recent African genesis of humans, Scientific American, 1992, (4): 68—75.
11. Weng, Z., Yuan, Y., Du, R., Analysis of the genetic structure of human populations in China, Acta Anthropol. Sin. (in Chinese)1989, 8: 261—268.
12. Zhao, T., Zhang, G., Zhu, Y. et al., The distribution of immunoglobulin Gm allotypes in forty Chinese populations, Acta Anthropol. Sin. (in Chinese), 1986, 6: 1—8.
13. Chu, J. Y., Huang, W., Kuang, S. Q. et al., Genetic relationship of populations in China, Proc. Natl. Acad. Sci., 1998, 95: 11763—11768.
14. Jobling, M. A., Tyler-Smith, C., Fathers and sons: the Y chromosome and human evolution, Trends in Genetics,1995, 11: 449—455.
15. Oefner, P. J., Underhill, P. A., Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC), Am. J. Hum. Genet., 1995, 57: A266.
16. Oefner, P. J., Underhill, P. A., DNA mutation detection using denaturing high performance liquid chromatography (DHPLC), Current Protocols in Human Genetics, Supplement, 1998, 19(7): 10.1—7.10.12.

17. Underhill, P. A., Jin, L., Zemans, R. et al., A pre-Columbian human Y chromosome-specific transition and its implications for human evolution, Proc. Natl. Acad. Sci., 1996, USA 93: 196—200.
18. Underhill, P. A., Jin, L., Lin, A. A. et al., Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography, Genome Research, 1997, 7: 996—1005.
19. Vollrath, D. S., Vollrath, D., Foote, S. et al.,The human Y chromosome: A 43-interval map based on naturally occurring deletions, Science,1992, 258: 52—59.
20. Underhill, P. A., Jin, L., Lin, A. A. et al., An African origin of human Y-chromosomes, Program and Abstract, 1997 Annual Meeting of The American Society of Human Genetics, Am. J. Hum. Genet, 1997, 61(supplement): A18.
21. Kayser, M., Caglia, A., Corach, D. et al., Evaluation of Y-chromosomal STRs: a multicenter study, Int. J. Legal Med., 1997, 110: 125—133.
22. Su, B., Xiao, J. H., Underhill, P. et al., Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age, Am. J. Hum. Genet., 1999, 6: 1718—1724.
23. Ballinger, S. W., Schurr, T. G., Torroni, A. et al., Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations, Genetics, 1992, 130: 139—152.
24. Heyer, E., Puymirat, J., Dieltjes, P. et al., Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees, Human Molecular Genetics,1997, 6: 799—803.
25. Bianchi, N. O., Catanesi, C. I., Bailliet, G. et al., Characterization of ancestral and derived Y-chromosomal haplotypes of New World native populations, Am. J. Hum. Genet., 1998, 63: 1862—1871.
26. Turner, II C. G., Shifting continuity: modern human origin, in The Origin and Past of Modern Humans as Viewed From DNA (eds. Brenner, S., Hanihara, K.), New York: World Sciencific,1993, 216—243.
27. Vasil'ev, S. A., The upper Paleolithic of northern Asia, Current Anthropology,1993, 34: 82—92.
28. Brown, P., Recent human evolution in East Asia and Australia, Philos. Trans. R. Soc., 1992, 337(1280): 235—242.
29. Swisher, C. C., 3rd, Rink, W. J., Anton, S. C. et al., Latest *Homo erectus* of Java: potential contemporaneity with *Homo sapiens* in Southeast Asia, Science ,1996, 274: 1870—1874.