Paternal Population History of East Asia: Sources, Patterns, and Microevolutionary Processes

Tatiana Karafet,[1,3] Liping Xu,[4] Ruofu Du,[4] William Wang,[5] Shi Feng,[5] R. S. Wells,[6] Alan J. Redd,[1] Stephen L. Zegura,[2] and Michael F. Hammer[1,2]

[1]Laboratory of Molecular Systematics and Evolution and [2]Department of Anthropology, University of Arizona, Tucson; [3]Laboratory of Human Molecular and Evolutionary Genetics, Institute of Cytology and Genetics, Novosibirsk, Russia; [4]Institute of Genetics, Chinese Academy of Sciences, Beijing; [5]Department of Electronic Engineering, City University of Hong Kong, Kowloon; [6]Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, United Kingdom

Asia has served as a focal point for human migration during much of the Late Pleistocene and Holocene. Clarification of East Asia's role as a source and/or transit point for human dispersals requires that this region's own settlement history be understood. To this end, we examined variation at 52 polymorphic sites on the nonrecombining portion of the Y chromosome (NRY) in 1,383 unrelated males, representing 25 populations from southern East Asia (SEAS), northern East Asia (NEAS), and central Asia (CAS). The polymorphisms defined 45 global haplogroups, 28 of which were present in these three regions. Although heterozygosity levels were similar in all three regions, the average pairwise difference among haplogroups was noticeably smaller in SEAS. Multidimensional scaling analysis indicated a general separation of SEAS versus NEAS and CAS populations, and analysis of molecular variance produced very different values of Φ_{ST} in NEAS and SEAS populations. In spatial autocorrelation analyses, the overall correlogram exhibited a clinal pattern; however, the NEAS populations showed evidence of both isolation by distance and ancient clines, whereas there was no evidence of structure in SEAS populations. Nested cladistic analysis demonstrated that population history events and ongoing demographic processes both contributed to the contrasting patterns of NRY variation in NEAS and SEAS. We conclude that the peopling of East Asia was more complex than earlier models had proposed—that is, a multilayered, multidirectional, and multidisciplinary framework is necessary. For instance, in addition to the previously recognized genetic and dental dispersal signals from SEAS to NEAS populations, CAS has made a significant contribution to the contemporary gene pool of NEAS, and the Sino-Tibetan expansion has left traces of a genetic trail from northern to southern China.

Introduction

The evolutionary history of human populations has been characterized by range expansions, colonizations, and recurrent gene flow restricted by isolation by distance (IBD) (Gamble 1993; Cavalli-Sforza et al. 1994; Lahr and Foley 1998; Templeton 1998; Fix 1999; Hammer et al. 2001). The relative importance of these migration processes and other evolutionary forces (such as selection and genetic drift) in shaping the patterns of variation in the human genome is not well understood (Przeworksi et al. 2000). Therefore, one of the main challenges facing human evolutionary geneticists is to disentangle the effects of past population history events from ongoing demographic processes. The study of population history in East Asia is particularly relevant to

this endeavor. The reconstruction of patterns of migration involving East Asia is of interest because (1) the route(s) of the earliest dispersals of anatomically modern humans out of Africa are not well understood from the archaeological and paleontological records, and (2) this region serves as a point of origin for subsequent migrations to Japan, Siberia, and the Americas.

Two major routes of migration into East Asia have been proposed: one through central Asia (CAS) and one through southern East Asia (SEAS) (Nei and Roychoudhury 1993, Cavalli-Sforza et al. 1994). The hypothesis that both routes played a significant role in the peopling of Asia has been referred to by Ding et al. (2000) as the "pincer" model. Analyses of classical genetic markers, as well as of dermatoglyphic and somatometric data, are consistent in showing a major division between northern East Asian (NEAS) and SEAS populations, with populations from northern and southern China falling in separate major clusters (Chu et al. 1998). In addition, Cavalli-Sforza et al. (1994, p. 232) inferred that the initial genetic differences between populations arriving in East Asia via different migration routes were maintained despite a long subsequent period of gene

Table 1

Haplogroup Frequencies and Population Diversities for 25 Asian Populations, Grouped by Geography and Language

Group, Population Number, and Population (Code)	Language Family	N	Frequency of Haplogroup[a]												
			11	12	14	16	17	20	21	22	23	24	25	26	27
NEAS:															
1. Northern Han (Nha)	Sino-Tibetan	44					2				1	1	5		
2. Hui (Hui)	Sino-Tibetan	54	3	2		3	6	2			1	4		1	1
3. Tibetans (Tib)	Sino-Tibetan	75	25	12			2	1				1	3		
4. Manchu (Man)	Altaic	52					14				1		3		
5. Chinese Evenks (Cev)	Altaic	41					18						2		
6. Oroqen (Oro)	Altaic	23					21							1	
7. Uygurs (Uyg)	Altaic	68	3			1	4			3	3	4	4		3
8. Mongolians (Mon)	Altaic	147	2	2			77	1		1	1	3	9	4	
9. Siberian Evenks (Sev)	Altaic	95					65	1	5			1	2	16	
10. Buryats (Bur)	Altaic	81					49					1	3	23	
11. Koreans (Kor)	Isolate/Altaic	74	3				8						2		
SEAS:															
12. Yizu (Yiz)	Sino-Tibetan	43		7			1	2					14		
13. Tujians (Tuj)	Sino-Tibetan	49		1		3	9						2		
14. Southern Han (Sha)	Sino-Tibetan	40					2						7		
15. Taiwanese Han (Tha)	Sino-Tibetan	82		1			6						3	2	
16. Zhuang (Zhu)	Austro-Asiatic	19													
17. She (She)	Austro-Asiatic	51													
18. Miao (Mia)	Austro-Asiatic	57		4		1	2								
19. Yao (Yao)	Austro-Asiatic	60		1		12									
20. Vietnamese (Vie)	Austro-Asiatic	70		2			3	1			1	1	2		
21. Malaysian (Mal)	Austronesian	32		1		1	1			1			2		
CAS:															
22. Kazakhs (Kaz)	Altaic	30					12	4		1	3	1		3	
23. Altai (Alt)	Altaic	29	1				7	1					2		
24. Uzbeks (Uzb)	Altaic	54			2		11	3	1	1	2	4	3	1	2
25. Kirghiz (Kir)	Altaic	13					1								
Regions:															
NEAS		754	36	16		4	266	5	5	5	6	15	33	45	4
SEAS		503		17		17	23	4		1	1	1	30	2	
CAS		126	1		2		31	8	1	2	5	5	5	4	2
Total		1,383	37	33	2	21	320	17	6	8	12	21	68	51	6

[a] Haplogroup designations are as given in figure 2. SE = standard error.

flow between regions. A competing model based primarily on dental data derives NEAS populations from the south, solely via dispersals originating in Sundaland (Turner 1990).

Three recent high-resolution studies at the DNA level (Chu et al. 1998; Su et al. 1999; Ding et al. 2000) have reached mutually contradictory conclusions concerning the origin(s), structure, and microevolutionary processes reflected in the contemporary East Asian gene pool. Chu et al. (1998) used a set of autosomal microsatellites to reconstruct phylogenies of East Asian populations. On the basis of the presence of a paraphyletic northern group and a nearly monophyletic southern group, they inferred a distinction between northern and southern populations, as well as a southern origin for northern populations. Su et al. (1999) reached similar conclusions on the basis of a survey of 19 biallelic polymorphisms on the nonrecombining portion of the Y chromosome (NRY) in 25 Asian populations. They in-

ferred support for northern and southern regional clusters through a principal-components analysis. Because northern populations were less polymorphic and had a subset of the NRY haplogroups found in southern populations, they suggested a southern origin for all East Asian populations. Ding et al. (2000) reached very different conclusions on the basis of their study of mtDNA sequences and five autosomal microsatellites, as well as a reanalysis of the polyomavirus JC sequence data of Sugimoto et al. (1997) and the NRY data of Su et al. (1999). They found no support for either a major north-south division or a southern origin of northern populations on the basis of principal-component maps for each marker system and concluded that patterns of variation throughout East Asia are best explained by a simple model of IBD.

To summarize the aforementioned disagreements, Chu et al. (1998) and Su et al. (1999) concluded that a southern origin is most likely, whereas Ding et al.

						Frequency of Haplogroup[a]										
28	29	30	31	32	33	34	35	36	37	40	42	43	44	45	$h \pm SE$	$p \pm SE$
	7	18	4		2	1				2				1	.80 ± .01	2.63 ± .22
	4	10	1	1	2					1	4		2	6	.93 ± .00	4.84 ± .33
1	1	24												5	.76 ± .00	3.85 ± .23
3	7	8	5	2	3		2				2	1		1	.88 ± .00	4.22 ± .30
		6	4	1	2	1	1			4				2	.78 ± .01	4.68 ± .37
					1										.17 ± .02	1.09 ± .16
	1	4	2	2						3	4		12	15	.90 ± .00	4.34 ± .26
1	4	23	6	1	2					4	2		1	3	.70 ± .00	4.16 ± .17
										4				1	.50 ± .01	2.71 ± .15
			2								1		1	1	.56 ± .01	3.24 ± .19
1	8	13	7	2	2		24	3						1	.84 ± .00	3.63 ± .22
	1	12	1		1	4									.80 ± .01	3.17 ± .26
	6	11	9	4	2									2	.87 ± .00	4.12 ± .30
	1	9	3	6		12									.82 ± .01	2.92 ± .25
	7	20	22	9	6	6									.84 ± .00	2.92 ± .17
1		3		2		13									.52 ± .03	1.67 ± .24
		3	29	1		18									.56 ± .01	2.06 ± .16
	12	6	22	4		6									.78 ± .01	2.81 ± .20
		3	29	1	12	2									.69 ± .01	3.33 ± .22
	7	11	11	4	1	19	1	1		4				1	.87 ± .00	3.41 ± .21
	4	3	3	3		11			1					1	.86 ± .01	3.32 ± .31
		3								1			2		.81 ± .01	4.11 ± .38
										2				16	.65 ± .02	4.09 ± .39
		1			1	1				4	2	1	5	9	.91 ± .00	4.69 ± .32
2				2										8	.62 ± .04	3.56 ± .54
6	32	106	31	9	14	2	27	3		18	13	1	16	36	.84 ± .00	4.72 ± .21
1	38	81	129	34	22	91	1	1	1	4				4	.86 ± .00	3.32 ± .08
2	—	4	—	2	1	1	—	—	—	7	2	1	7	33	.86 ± .00	4.71 ± .08
9	70	191	160	45	37	94	28	4	1	29	15	2	23	73	.90 ± .00	4.71 ± .00

(2000) proposed that recent IBD processes have erased earlier genetic signatures. Thus, rather than a southern origin for NEAS populations, Ding et al. (2000) stressed the potential importance of more recent gene flow from CAS populations, via the Silk Road. Chu et al. (1998) and Su et al. (1999) support a north-south genetic division in East Asia, which Ding et al. (2000) contend is more properly restricted to post-Pleistocene cultural phenomena.

A major shortcoming of these three DNA studies is the lack of sampling in CAS to assess the role of these populations in contributing to the genetic composition of East Asia. Another potential bias can result from oversampling of SEAS populations (Chu et al. 1998; Su et al. 1999). For example, there were more than two times as many individuals and populations from SEAS in the NRY study by Su et al. (1999), compared with populations from NEAS. Moreover, the choice of the 19 NRY markers was influenced by previous knowledge that they were polymorphic in East Asian populations

(Su et al. 1999). This could lead to an ascertainment bias when assessing the relative diversity of NEAS versus SEAS populations. The present study was designed to include additional NRY polymorphisms ascertained in a global sample, as well as more populations from both NEAS and CAS. We performed phylogenetic, nonmetric multidimensional scaling (MDS), spatial autocorrelation, and nested cladistic analyses (NCA), as well as analysis of molecular variance (AMOVA), to test the hypothesis of a north-south East Asian division and to assess the relative roles of population history and population structure in the shaping of patterns of NRY variation in East Asia.

Subjects and Methods

Populations

We analyzed a total of 1,383 unrelated males, representing 25 Asian populations (table 1). To avoid the

possible effects of strong genetic drift, we restricted our sample composition to continental Asia (except in the case of Taiwan, where recent immigrants from mainland China were sampled). The 25 populations were divided into three regional groups, primarily on the basis of geography: 11 formed the NEAS group, 10 formed the SEAS group, and 4 formed the CAS group (fig. 1). Table 1 also presents linguistic affiliations (Ruhlen 1992) and the three-letter and numerical population codes used throughout the remainder of the article.

Genomic DNAs from 10 populations were collected in the laboratory of R. Du in China. These populations (and the province of origin) were: Yao (Guangxi), Uygurs (Xinjiang), Hui (Ningxia), Manchu (Liaoning), Tujians (Hunan), Tibetans (Xizang), Zhuang (Guangxi), southern Han (Guangdong), northern Han (Shaanxi), and Yizu (Sichuan). The families of the sampled males had lived at their present locations for at least three generations. Several populations analyzed here (Siberian Evenks, Chinese Evenks, Oroqen, Mongolians, Buryats, Koreans, Miao, Vietnamese, Kazakhs, Altai, Uzbeks, Taiwanese Han, and Malaysians) have been described in our previous studies (e.g., Hammer et al. 2001). Buccal samples from the She population were collected by W. Wang and were extracted in Tucson. DNA samples from the final population, the Kirghiz, were provided by R. S. Wells. All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

Genetic Markers

Many of the >200 biallelic polymorphisms known on the NRY are highly geographically localized (Hammer and Zegura 1996; Underhill et al. 2000; Hammer et al. 2001). Ideally, a large number of randomly selected markers should be used for studies of geographically dispersed populations, and all markers should be genotyped on all samples. In practice, this is difficult because of limitations in time and materials (Underhill et al. 2000). Thus, the choice of markers is an important consideration for minimizing the ascertainment bias introduced in studies of NRY variation. We chose to survey a set of 42 polymorphisms from the literature and from our own mutation-detection experiments, which were performed on a geographically diverse sample of Y chromosomes (Hammer et al. 2001). In addition, we genotyped another 10 biallelic polymorphisms that were known to be polymorphic in Asia, some of which were used in the study by Su et al. (1999). These latter sites included M20, M119, M134, M172, M173, M174, M175, M217, UTY1-1330+18, and UTY1-3678+537 (Shen et al. 2000; Underhill et al. 2000). We screened all samples with 15 biallelic markers that define the branches on a haplogroup tree (Hammer et al. 2001).
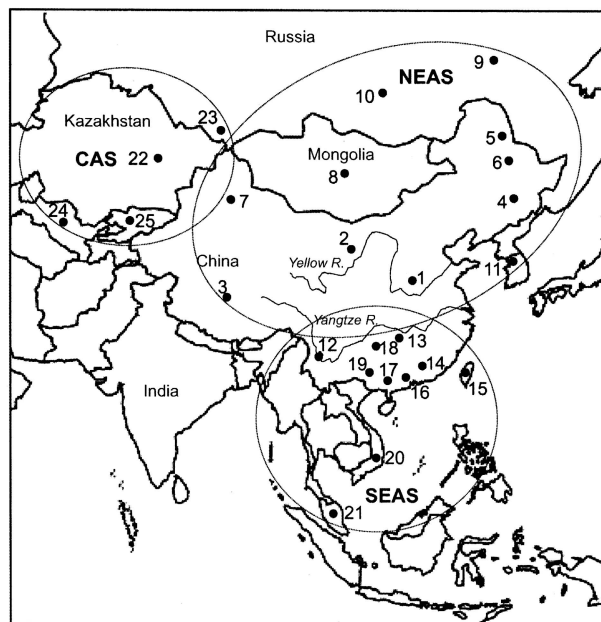

Figure 1 Map of 25 sampling localities, divided into three regional groupings (NEAS, SEAS, and CAS). Numbers indicate population numerical codes, given in table 1.

Subsequent analysis of a sample was restricted to markers on the appropriate branch of the haplogroup tree.

Statistical Analysis

We used Arlequin 2.0 software (Schneider et al. 1998) to perform AMOVA, to estimate Φ_{ST} distances, to test the correlation between genetic and geographic distances (Mantel 1967), and to measure haplogroup diversity. Two measures of haplogroup diversity were employed, including Nei's h (Nei 1987) and the mean number of pairwise differences among haplogroups (p). We performed nonmetric MDS (Kruskal 1964) on the Φ_{ST} distances, using the software package NTSYS (Rohlf 1998). MDS is an ordination technique for representing the dissimilarity among objects (e.g., populations) in an n-dimensional graph, such that the interpoint distances in the graph space correspond as well as possible to the observed genetic differences between populations. The goodness of fit between the distances in the graphic configuration and the monotonic function of the original distances is measured by a statistic called "stress," wherein a value of 0 is a perfect fit and a value of 1 is a total mismatch. Spatial autocorrelation analysis was performed using the autocorrelation index for DNA analysis (AIDA) (Bertorelle and Barbujani 1995). AIDA is a form of spatial autocorrelation analysis that summarizes genetic variation among individuals as a function of their distance in space. Measures of molecular similarity are estimated within arbitrary intervals, and

their departure from null expectations is tested by randomization. Much like correlation coefficients, autocorrelation statistics are positive when individuals are genetically similar at a certain distance, are negative when they are dissimilar, and are expected to be zero under the hypothesis of spatial randomness. A haplotype-based analog to Moran's I, denoted as "II," was employed as the measure of spatial effects on haplogroup frequencies (Fix 1999). To test genetic differentiation under an IBD model, we analyzed the regression of genetic distance estimates for pairs of populations on geographic distances. In a two-dimensional stepping-stone model (Kimura and Weiss 1964) with IBD, an approximately linear relationship between genetic distances, $\Phi_{ST}/(1 - \Phi_{ST})$, and the logarithm of geographic distances is expected (Slatkin and Maddison 1990; Rousset 1997; Dupanloup de Ceuninck et al. 2000). Analyses of genetic variance that are based on Wright's island model and AIDA cannot provide insights into the relative roles of historical events and ongoing processes in generating patterns of genetic variation (Hammer et al. 2001). For this reason, we also used the approach of Templeton et al. (1995) to disentangle the effects of population history (e.g., contiguous range expansion, long-distance colonization, and fragmentation) from population structure (e.g, recurrent gene flow restricted by IBD and long-distance dispersal). Population structure processes operate over short time intervals and tend to establish migration-drift equilibria, whereas population history events are considered to be nonrecurrent phenomena that disrupt equilibria (Hammer et al. 2001). We used GeoDis version 2.0 (GeoDis Home Page) to conduct NCA. This method attempts to clarify, in terms of inferred population history and/or population structure considerations, the causal factors of any statistically significant associations between haplogroups and geography (Templeton et al. 1995).

Results

Geographic Distribution of NRY Haplogroups in Asia

Of the 52 mutational sites surveyed here, 30 were found to be polymorphic in the NEAS, SEAS, and CAS populations. Figure 2 shows the evolutionary relationships among the 45 haplogroups (h1–h45) defined by variation at these 52 polymorphic sites. Of the resulting 45 global NRY haplogroups, 28 were present in NEAS, SEAS, and CAS populations. In contrast to the survey by Su et al. (1999), NRY haplogroups of NEAS were not found to be a subset of those in SEAS. Fourteen haplogroups were shared among the three geographic regions. Seven haplogroups were shared between CAS and NEAS, and five haplogroups were shared between
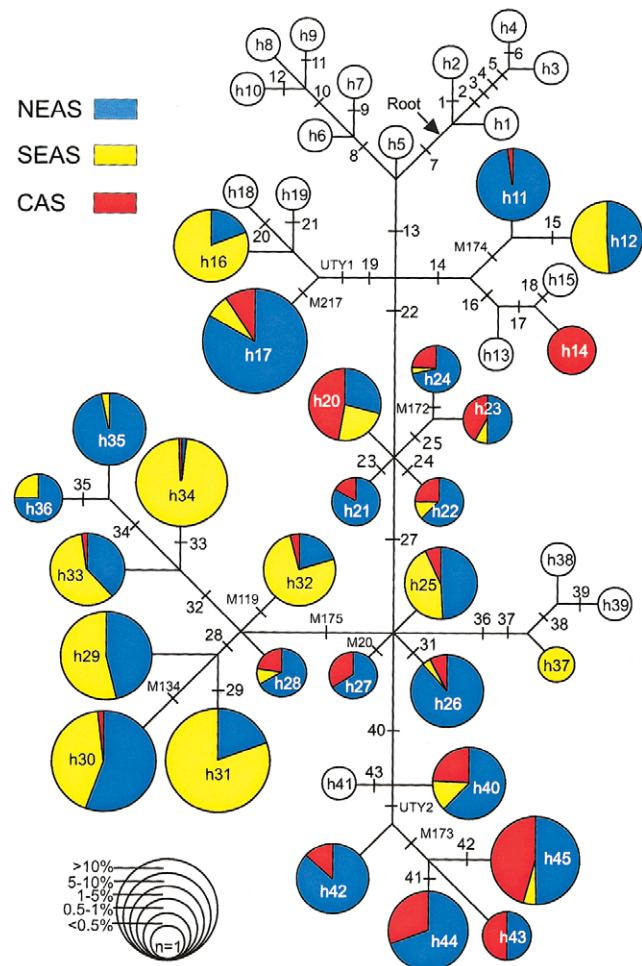

Figure 2   Evolutionary tree for 45 NRY haplogroups (h1–h45). The root of the haplogroup tree is indicated by an arrow. The 52 mutational events, of which the first 42 appear in table 1 in the report by Hammer et al. (2001), are shown by cross-hatches. The 10 additional mutational sites are from reports by Underhill et al. (2000) and Shen et al. (2000), where UTY1 = UTY1-3678+537 and UTY2 = UTY1-1330+18. Haplogroups are color-coded according to geographic region (see figure for key); circles without colors denote absence in NEAS, SEAS, and CAS. The pie charts represent the frequency of occurrence of the haplogroups within each of the three Asian geographic regions shown in figure 1. The overall size of each pie chart corresponds to one of six frequency classes (see figure for frequency-class key) and represents the frequency of that haplogroup in the total sample of 1,383 chromosomes.

SEAS and NEAS. Two haplogroups (h14 in CAS and h37 in SEAS) were specific to a single region.

NRY Haplogroup Diversity

NRY haplogroup diversity for each individual population and for the three major regional groups is also shown in table 1. Nei's (1987) diversity statistic, h, which is based on the frequency and number of haplogroups, ranges from .170 in the Oroqen to .928 in the

Hui, both from northern China. The mean regional h value was the highest in CAS, followed by SEAS and NEAS. As in the case of h, the mean number of pairwise differences among haplogroups, p, ranged from a low value in the Oroqens (1.09) and to a high value in the Hui (4.84). At the regional level, however, p values showed a different pattern: the highest p value was found in NEAS, whereas SEAS had the lowest p value. The relatively low mean number of pairwise differences among haplogroups in SEAS occurred because 85% of SEAS Y chromosomes belong to seven closely related haplogroups (e.g., the haplogroups marked by M175 only differed by an average of 2.75 mutations) (fig. 2). The high values of p in populations from NEAS and CAS reflect a more marked divergence among haplogroups in these regions.

MDS Plot

Figure 3 portrays the results of MDS based on Φ_{ST} genetic distances. All SEAS populations cluster together on the left half of plot, whereas nearly all NEAS populations occupy the right half of the plot. The only exceptions are the northern Han Chinese, Korean, and Manchu populations, which have closer genetic affinities with southern populations than with northern populations. All but one of the CAS populations (the Kazakhs), as well as the Uygurs from western China, occupy the lower right part of the plot. The stress value (.10) of the

MDS plot indicates a good fit between the two-dimensional graph and the original distance matrix.

AMOVA

Table 2 presents variance components and three Φ statistics at three different grouping levels, which summarize the geographic partitioning of NRY diversity. When all 25 populations were combined into three regions, the overall Φ_{ST} was .31 (i.e., ~69% of the variance occurred within populations). This value was surprisingly close to our global Φ_{ST} of .36, which is based on 2,858 males from 50 worldwide populations (Hammer et al. 2001). The among–populations within groups variance component and the among-group variance component showed similar values of ~15% each. NRY differentiation was significantly higher among NEAS populations (Φ_{ST} = .23) than among SEAS populations (Φ_{ST} = .09). CAS populations showed an intermediate value (Φ_{ST} = .12), which may partially reflect the fact that only four populations were sampled from this region. To identify the extent of among-group variation, we performed AMOVA for pairwise groupings. The highest resulting between-group value was found between SEAS and CAS (Φ_{CT} = .28), followed by the value between SEAS and NEAS (Φ_{CT} = .16). Consistent with the MDS plot (fig. 3), the among-group variance component between CAS and NEAS was not statistically significant (Φ_{CT} = .04, P = .156).
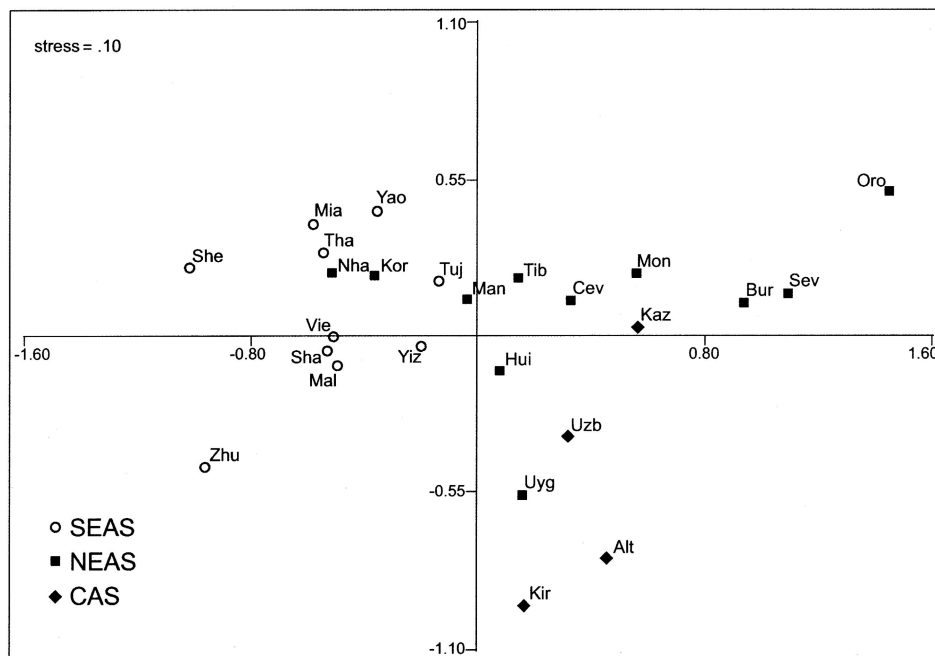

Figure 3 MDS plot of 25 Asian populations, based on Φ_{ST} genetic distances. For three-letter population codes, see table 1.

Spatial Autocorrelation Analysis

AIDA analysis was performed three times, initially including all 25 populations and then separately analyzing the SEAS and NEAS regional groupings. In the analysis of all populations (fig. 4*a*), the II values are positive and significant for distances <2,400 km and are negative and significant for distances >3,000 km. Although the overall correlogram does not exhibit a monotonic decrease, the pattern is very clearly clinal. The peak at 4,800 km shows that some geographically distant populations are not correspondingly genetically distant. The autocorrelation indices show very different patterns of variation in the NEAS and SEAS groupings (fig. 4*b*). In NEAS populations, the values of II generally decrease from positive and significant to negative and significant as geographic distance increases. The NEAS populations show some evidence of IBD, because populations within 600 km of one another are very similar. However, there is no strong geographic structuring at greater distances. Significantly negative dips at ~1,200–1,800 km and ~3,000–3,600 km were followed by upward fluctuations. Such a pattern is referred to as "long distance differentiation" in classical spatial autocorrelation studies (Sokal et al. 1989; Barbujani et al. 1994). These patterns are regarded as ancient clines on which the effects of gene flow, genetic drift, and/or adaptation to local environmental factors have been superimposed. There is no evident structure in SEAS populations, aside from the expected increased similarity among populations within the zero distance class. This lack of a structural signal may be due to insufficient sampling in the SEAS region.

IBD Model and Mantel Tests

Plotting the $\Phi_{ST}/(1 - \Phi_{ST})$ distances versus the log of geographic distances for 25 Asian populations (fig. 5*a*) indicated a significantly positive relationship between genetic and geographic distances ($r = .357$; Mantel test:
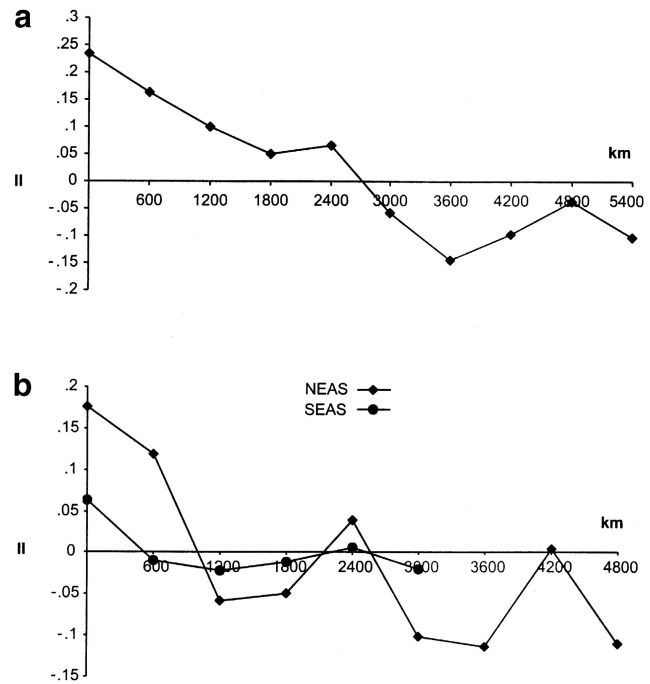

Figure 4 Spatial autocorrelation plots. *a*, 25 Asian populations. *b*, NEAS (*diamonds*) versus SEAS (*circles*) populations.

$P = .000$). This result should be interpreted with caution, because Slatkin (1993) showed that a "signature" of IBD can result from regional differences in patterns of gene flow. Because the AMOVA and AIDA results revealed different patterns of variation in NEAS and SEAS, we plotted genetic versus geographic distances separately for NEAS and SEAS populations (figure 5*b* and 5*c*, respectively). Whereas northern groups exhibited a statistically significant positive correlation ($r = .349$, $P = .022$), southern groups showed a statistically significant negative relationship between genetic and geographic distances ($r = -.381$, $P = .033$).

Table 2

Results of AMOVA

Group	N	No. of Populations	Within Populations Variance (%)	Φ_{st}	Among Populations Within Groups Variance (%)	Φ_{sc}	Among Groups Variance (%)	Φ_{ct}
Asia	1,383	25	69.2	.31	15.4	.18	15.4	.15
East Asia	1,254	21	68.1	.32	15.7	.19	16.2	.16
NEAS	754	11	76.6	.23				
SEAS	503	10	91.2	.09				
CAS	126	4	88.0	.12				
NEAS/SEAS	1,254	21	68.1	.32	15.7	.19	16.2	.16
NEAS/CAS	880	15	75.0	.25	21.2	.22	3.8	.04
SEAS/CAS	629	14	64.9	.35	6.8	.09	28.3	.28

NOTE.—All Φ statistics P values are <.001, except for the NEAS/CAS Φ_{CT}, for which $P = .156$.
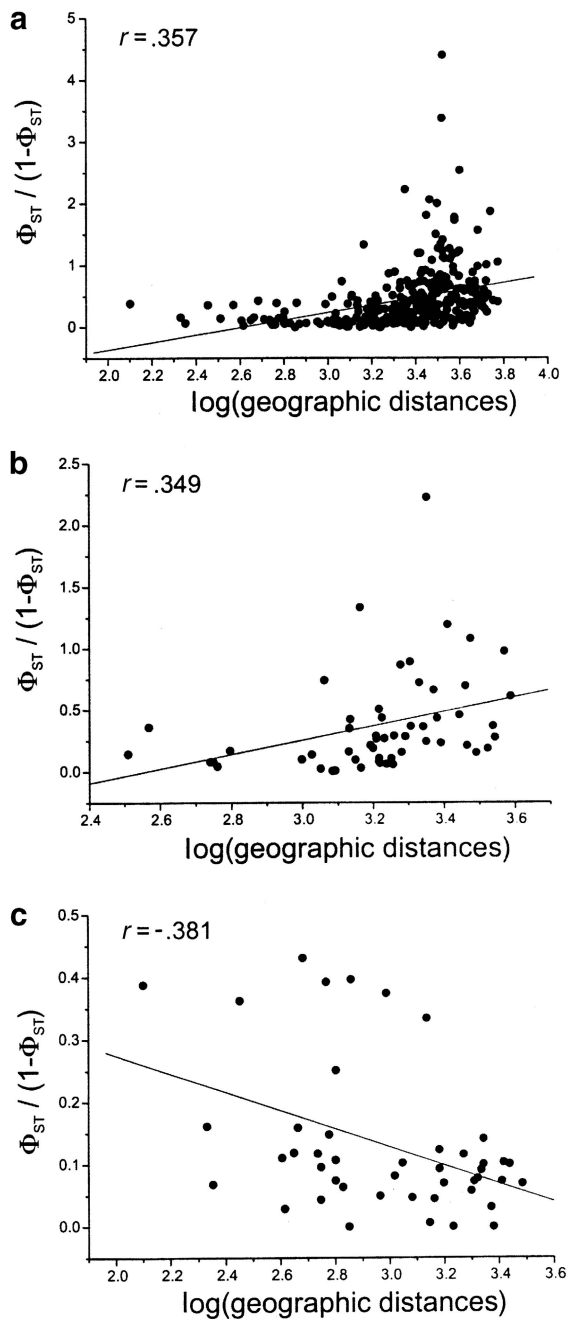
Figure 5 Regression of genetic-distance estimates for pairs of populations against the log of geographic distances. *a,* 25 Asian populations. *b,* NEAS populations. *c,* SEAS populations.

NCA

The nesting methodology of Templeton and Sing (1993) produced 14 one-step clades, 6 two-step clades, 3 three-step clades, 1 four-step clade, and a single five-step clade that nested the entire cladogram (fig. 6). A random permutation procedure indicated highly statistically significant associations between clades and geographic locations

for the entire cladogram. Of a total of 25 nested clades, 17 exhibited statistically significant associations with geography. With the aid of a key published on the GeoDis 2.0 Web site, we were able to infer the probable causes of these 17 patterns (table 3). Consistent with our previous analyses (Hammer et al. 1998, 2001; Karafet et al. 1999), the present study indicated that the distribution of NRY haplogroups has been influenced by both population history and population structure factors. In contrast with previous results, a larger proportion of the signals (11 of 17) were the result of population structure processes such as recurrent gene flow restricted by IBD ($n = 10$) and long-distance dispersal ($n = 1$). Only six signals involved unique historical events such as contiguous range expansions ($n = 4$) and long-distance colonizations ($n = 2$).

Discussion

Geographic Patterns of Diversity in East Asia: Is There a North-South Division of NRY Variation?

Important differences between NEAS and SEAS populations have been noted in many studies based on both genetic and morphological characters (Cavalli-Sforza et al. 1994). These findings have inspired several hypotheses concerning the origin of the Chinese people (Matsumoto 1988; Zhao and Lee 1989; Chu et al. 1998; Qian et al. 2000), the origin of linguistic families in Asia (Chu et al. 1998; Su et al. 2000), and prehistoric migrations to and within Asia (Nei and Roychoudhury 1993; Cavalli-Sforza et al. 1994; Su et al. 1999, 2000).

The present survey increased both the number of populations examined for markers on the NRY and the number of NRY markers studied in East Asian populations. Populations were sampled from NEAS and SEAS, as well as from CAS. The patterns of NRY haplogroup variation presented here have several features in common with other data sets, as well as several features that are novel. The MDS plot in figure 3 shows that the majority of SEAS populations are clearly differentiated from NEAS populations. Despite the extremely high internal diversity in the SEAS region (within-population variance = ~91%, table 2), the results of both the MDS plot and AMOVA are consistent in demonstrating a closer genetic relationship between CAS and NEAS populations than between either of the former groups and SEAS populations. This fact alone makes it clear that the genetic history of CAS populations must be considered in any analysis of the origins of East Asian populations (Ding et al. 2000).

In contrast to the results reported by Su et al. (1999), our results do not indicate that southern populations are more polymorphic than northern populations in East Asia, or that haplogroups found in the north represent a subset of those in the south (table 1). On the contrary,
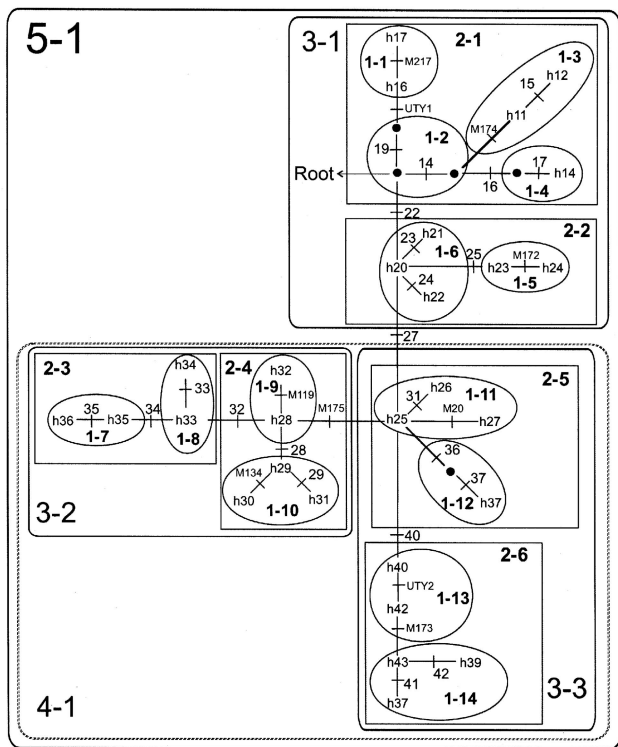
Figure 6 Nested cladistic design for 45 NRY haplogroups. The 52 mutational events (*cross-hatches*), 45 haplogroups (h1–h45), and cladogram root (*arrow*) are as described in figure 2. Black circles represent haplogroups that were missing in this sample of Y chromosomes. Ovals contain one-step clades (designated as "1-1" through "1-14"); rectangles contain two-step clades (designated as "2-1" through "2-6"); and rounded rectangles contain three-step clades (designated as "3-1" through "3-3"). Double-dotted lines contain one four-step clade (designated as "4-1"). A single five-step clade (5-1) encompasses the entire cladogram.

26 haplogroups were found in the north, whereas SEAS populations had only 21 haplogroups. Although caution should be exercised when comparing levels of diversity among different studies (Hammer et al. 2001), our results do not support the general conclusion of an exclusively southern origin for NEAS populations.

Two possible reasons for the difference between our results and those of previous studies of NRY variation in Asia include thinner sampling of NEAS populations in the study by Su et al. (1999) and/or different choices of NRY markers. To assess the contribution of these two sources of bias, we compared the ratio of northern to southern diversity in our data set (i.e., *p* in NEAS versus *p* in SEAS) with (1) a similar ratio based on the data presented by Su et al. (1999) and (2) a similar ratio based on our data but restricted to the haplogroups analyzed by Su et al. (1999) (using our markers, we were able to construct 13 of the 17 haplogroups analyzed by Su et al. [1999]). In the study by Su et al. (1999), there was

a relatively low ratio of NEAS to SEAS diversity (*p* ratio = 0.88). When we restricted the analysis of our data to the haplogroups analyzed by Su et al. (1999), we found a slightly higher ratio of NEAS to SEAS diversity (*p* ratio = 1.2); however, this was not as high as the diversity ratio in our study (*p* ratio = 1.42). These rough comparisons suggest that both the markers used and the populations sampled contributed to differences between the results of Su et al. (1999) and those presented here.

There were also contrasting patterns of NRY haplogroup diversity within the three regions surveyed here. Nei's (1987) diversity statistic (*h*) was slightly lower in NEAS than in SEAS. A larger effect of genetic drift may be part of the explanation for this observation because there are lower population densities and smaller population sizes in the north. The AMOVA results are consistent with this explanation. While significant genetic structure was observed in both regions, a nearly three-fold-higher Φ_{ST} value was found in NEAS compared with SEAS. If the extent of molecular differences between haplogroups is *not* taken into account, and genetic structure is estimated only from haplogroup frequencies, we still observe a higher Φ_{ST} value in the north than in the south (.157 vs. .101, respectively; data not shown), although this is a less marked numerical difference than for the corresponding Φ_{ST} values. This result suggests that both genetic drift and haplogroup composition have played important roles in the differentiation of populations in these two regions. Because of the large pairwise differences found in NEAS populations, it appears that highly divergent haplogroups are very important in distinguishing the populations within NEAS. This may reflect migrations from different source populations.

How does the among-group variance component between NEAS and SEAS populations compare with other regions of the world? A direct comparison can be made with the Φ_{CT} value observed between CAS and SEAS populations, which was ~75% higher (table 2). With the caveat that we are comparing across studies using different but overlapping sets of NRY markers, the level of differentiation between NEAS and SEAS populations observed here (i.e., Φ_{CT} = .16) is intermediate with respect to levels of differentiation between Middle Eastern and European populations (Φ_{CT} = .10) and between northern African and European populations (Φ_{CT} = .27) (Hammer et al. 2000).

Factors Shaping Genetic Variation in East Asia: Testing the IBD Model

The MDS, AMOVA, and NCA results provided evidence of geographic structuring in East Asia. We wanted to discern what factors shaped patterns of NRY diversity in East Asia, and, in particular, we wanted to test the

Table 3

Inferences from NCA

Clade[a]	Inference Chain[b]	Inference
1-1	1→2→11→12→13→Yes	Long-distance colonization
1-7	1→2→11→RE→12→Yes	Contiguous range expansion
1-8	1→2→3→4→No	Recurrent gene flow restricted by IBD
1-9	1→2→3→4→No	Recurrent gene flow restricted by IBD
1-10	1→2→3→4→No	Recurrent gene flow restricted by IBD
1-11	1→2→11→RE→12→13→Yes	Long-distance colonization
1-13	1→2→3→4→No	Recurrent gene flow restricted by IBD
1-14	1→2→3→4→No	Recurrent gene flow restricted by IBD
2-2	1→2→3→4→No	Recurrent gene flow restricted by IBD
2-3	1→2→3→5→6→7→Yes	Recurrent gene flow with some long-distance dispersal
2-4	1→2→3→4→No	Recurrent gene flow restricted by IBD
2-5	1→2→11→RE→12→Yes	Contiguous range expansion
2-6	1→2→3→4→No	Recurrent gene flow restricted by IBD
3-1	1→2→11→RE→12→Yes	Contiguous range expansion
3-3	1→2→3→4→No	Recurrent gene flow restricted by IBD
4-1	1→2→3→4→No	Recurrent gene flow restricted by IBD
Total	1→2→11→RE→12→Yes	Contiguous range expansion

[a] Only clades resulting in the rejection of the null hypothesis are included.

[b] See the GeoDis Home Page for inference key. RE = range expansion.

hypothesis of Ding et al. (2000) that NRY marker patterns result from simple IBD. The term "isolation by distance" was introduced by Wright (1943) to describe the accumulation of local genetic differences under geographically restricted dispersal. Under the IBD model, genetic differentiation at neutral loci increases with geographic distances (Wright 1943; Malécot 1968; Morton 1973). Although this model is more suitable for short-range migrations between neighboring populations, covariation of geographic and genetic distances has been observed at a large geographic scale in several human population systems (Jorde 1980; Excoffier et al. 1991; Cavalli-Sforza et al. 1992, 1994; Barbujani and Pilastro 1993; Poloni et al. 1997; Hammer et al. 2000).

To distinguish the *pattern* of increasing genetic distance with geographic distance from the *process(es)* generating this pattern, we refer to the pattern as "spatial correspondence." For example, when we considered all 25 populations from the CAS, NEAS, and SEAS regions, there was a statistically significant positive correlation between genetic and geographic distances (fig. 5a). This pattern of spatial correspondence is consistent with IBD (i.e., equilibrium under geographically restricted gene flow); however, this inference must be treated with caution, since the geographic region under investigation is vast, and other processes may be responsible for the observed positive correlation (Barbujani and Sokal 1991). When the NEAS and SEAS populations were analyzed separately, we found very different patterns (fig. 5b and 5c). This leads to the inference that NRY structuring in the SEAS region is not maintained by recurrent gene flow among local populations. The apparent lack of an IBD signal may be due to several nonmutually exclusive factors, in-

cluding recent population movements, language and/or cultural boundaries, sampling at an inappropriate scale, and/or unrealistic assumptions of the IBD model (Slatkin 1993; Zegura et al. 1995; Rousset 1997; Dupanloup de Ceuninck et al. 2000).

Spatial autocorrelation (Sokal and Oden 1978) is another method for analyzing geographic patterns of genetic diversity. This method compares data within each of several distance classes, and inferences are based on the degree of genetic similarity at various geographic distances. The particular method of spatial autocorrelation employed here, AIDA (Bertorelle and Barbujani 1995), also takes sequence differences among haplogroups into account. In other words, two localities are considered more alike if the same haplogroups occur at similar frequencies and if the various haplogroups differ by fewer mutations. The distribution of II across distance classes (i.e., the correlograms) can be compared with patterns predicted by different evolutionary processes. The resulting patterns are interpreted as clines, depressions (i.e., clinal variation encompassing only a part of the study area), IBD, random genetic variation, or as reflecting various selective regimes (Sokal 1979; Fix 1999; Barbujani 2000).

Spatial patterns of genetic similarity did not show a pure pattern of IBD in our data set (fig. 4). Under IBD, variations in population size affect only the impact of genetic drift (i.e., the larger the population, the smaller the allele-frequency fluctuations). When only drift and short-range dispersal affect populations, genetic similarity is expected to decrease from positive to insignificant in spatial autocorrelation analysis (Barbujani et al. 1994). NRY variation over the whole range of samples

is more consistent with a clinal pattern than with an IBD pattern. Clines are usually associated with distinct population movements. Demic diffusion (a combination of demographic growth, range expansion, and limited admixture) is an example of a form of directional population expansion causing allele-frequency clines (Ammerman and Cavalli-Sforza 1984). Clines resulting from demic diffusion can be stable for long periods of time (Cavalli-Sforza et al. 1994). Clines may also be generated by loss of genetic variation through repeated founder effects occurring in a phase of population expansion not accompanied by admixture (Barbujani et al. 1995; Fix 1999), by admixture between two genetically distinct groups initially separated by a nonpopulated area (Endler 1977), or by a selection gradient (Fix 1999).

In our NEAS sample, 7 of 11 populations are speakers of languages within the Altaic linguistic family. Our samples from SEAS include four Sino-Tibetan–speaking populations, five Austro-Asiatic–speaking populations, and one Austronesian-speaking population. Austro-Asiatic and Austronesean languages are considered to be branches of the Austric linguistic superfamily (Ruhlen 1992). In a study based on classical markers, Altaic-speaking populations exhibited a pattern of clinal variation (Barbujani et al. 1994). Their data and our NRY results (i.e., evidence for ancient clines in the NEAS correlogram plot, fig. 4*b*) are both consistent with Renfrew's (1991) prediction that Altaic languages should be included among those that were propagated by demic diffusion. This demic-diffusion hypothesis is also concordant with the inference that population history has played a key role in generating patterns of NRY variation in NEAS populations. On the other hand, NRY variation appears generally random even at small distances in SEAS. Likewise, Barbujani et al.'s (1994) classical marker study indicated that populations of Sino-Tibetan speakers and Austric speakers exhibited random genetic variation even at small distances. Under an IBD model, positive autocorrelation at the first distance classes must be observed. The insignificant and even negative autocorrelation detected at relatively short distances for SEAS populations suggests the rather puzzling conclusion that short-range gene flow was not important in establishing the observed levels of diversity among these populations.

The difference that we see between patterns of variation in NEAS and SEAS populations could be random, because an insufficient number of populations was included in our analyses. Another, and perhaps more cogent, explanation points to the different linguistic and cultural affiliations of populations inhabiting NEAS and SEAS. Interestingly, when we performed AMOVA on populations grouped according to a linguistic-family criterion (i.e., Sino-Tibetan, Altaic, Austro-Asiatic, and Austronesian), among-group distances were slightly larger ($\Phi_{CT} = .18$, data not

shown) than when populations were grouped into the three major geographic regions ($\Phi_{CT} = .16$, table 2). To summarize, a high level of population subdivision (caused or enhanced by linguistic barriers) may have led to a random distribution of NRY diversity in SEAS. On the contrary, in NEAS there was a significant cline, consistent with the effects of large-scale directional expansions of Altaic-speaking populations. Recurrent gene flow within short geographic distances (suggested by the NCA results) probably affected genetic diversity at a local scale and was reflected in the fluctuations of autocorrelation indices at intermediate distance classes. Nevertheless, the difference between the NEAS and SEAS regional gene pools was large enough to produce a broad gradient when all populations were jointly analyzed by spatial autocorrelation.

Conclusions

In sum, it appears that different kinds of evolutionary forces shaped patterns of genetic variation in East Asia. Although it is difficult to quantify the genetic impact of major population movements in the past, our results seem to be incompatible with random genetic variation and/or pure IBD as the only explanations for the patterns of NRY variation in East Asia. Our current results also suggest that some recent conclusions concerning the origin(s) of East Asian populations may have been premature. We do not find the simple ancestor-descendant relationship between SEAS and NEAS populations suggested by Su et al. (1999). Chu et al. (1998) posited that Altaic-speaking populations occupying the most northern parts of Asia originated from an East Asian population that was originally derived from SEAS. Their reasons were based, in part, on the claim that because the last glaciers started to recede only 15,000 years ago, an early migration route from CAS to Siberia was unlikely. However, the great boreal forest or taiga of Siberia was established in essentially its present character early in the Pleistocene and was never wholly displaced. Central and southern Siberia were not under glaciers and constituted a major refugium in northern Eurasia during the periods of maximum cold (Kuzmin and Orlova 1998).

The very different patterns of variation in NEAS and SEAS may better fit a two-prong (i.e., pincer) model for the origins of East Asian populations, similar to that championed by Cavalli-Sforza et al. (1994). Certainly, the closer genetic and linguistic relationships of CAS and NEAS populations lend support to the hypothesis of a separate migratory route coming from CAS (Cavalli-Sforza et al. 1994). Although we found a rather clear dichotomy of NEAS and SEAS populations, the sharing of haplogroups between southern and northern populations may be explained by subsequent short-range and long-range migration processes, perhaps as-

sociated with the advent of agriculture and animal domestication. Although this postulated south-to-north gene flow has begun to obscure the initial genetic differences between NEAS and SEAS, it has certainly not yet erased them (contra Ding et al. 2000). Su et al.'s (2000) hypothesis that the Sino-Tibetan language family originated near the Yellow River (in the NEAS region) and subsequently dispersed to the SEAS region adds an additional complication, this being the strong possibility of bidirectional gene flow across the Yangtze River during the Holocene (fig. 1). This hypothesis of relatively recent bidirectional migration is clearly underscored by visual inspection of figure 2. For instance, most of the haplogroups derived from (i.e., to the left of) the mutational site M175 are present at relatively high frequencies in SEAS populations (*yellow*) and are shared primarily with NEAS populations (*blue*) rather than with CAS populations (*red*). These haplogroups may well have originated in SEAS and spread to the north (Su et al. 1999, 2000). On the other hand, many of the haplogroups derived from the backbone of the gene tree are shared between NEAS and CAS populations and account for the majority of NEAS haplogroups. This phylogeographic pattern may be a signal of dispersals from CAS towards NEAS; moreover, the minor sharing of these haplogroups with SEAS populations could, indeed, be the signal of north-to-south dispersals such as the aforementioned Sino-Tibetan expansion. These observations imply that future, more realistic models for the underlying processes leading to the modern population structure of East Asia will have to accommodate more complex multidirectional biological and—*especially*—cultural influences than earlier explanatory paradigms.

Acknowledgments

Electronic-Database Information

The URL for data in this article is as follows:

GeoDis Home Page, http://bioag.byu.edu/zoology/crandall_lab/geodis.htm (for GeoDis version 2.0 software)

References

Ammerman AJ, Cavalli-Sforza LL (1984) Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Barbujani G (2000) Geographic patterns: how to identify them and why. Hum Biol 72:133–153

Barbujani G, Pilastro A (1993) Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. Proc Natl Acad Sci USA 90:4670–4673

Barbujani G, Pilastro A, De Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs. Paleolithic colonisation. Am J Phys Anthropol 95:137–154

Barbujani G, Sokal RR (1991) Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. Am J Hum Genet 48:398–411

Barbujani G, Sokal RR, Oden NL (1995) Indo-European origins: a computer-simulation test of five hypotheses. Am J Phys Anthropol 96:109–132

Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811–819

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Cavalli-Sforza LL, Minch E, Mountain JL (1992) Coevolution of genes and languages revisited. Proc Natl Acad Sci USA 89:5620–5624

Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. Proc Natl Acad Sci USA 95:11763–11768

Ding YC, Wooding S, Harpending HC, Chi HC, Li HP, Fu YX, Pang JF, Yao YG, Yu JG, Moyzis R, Zhang Y (2000) Population structure and history in East Asia. Proc Natl Acad Sci USA 97:14003–14006

Dupanloup de Ceuninck I, Schneider S, Langaney A, Excoffier L (2000) Inferring the impact of linguistic boundaries on population differentiation: application to the Afro-Asiatic-Indo-European case. Eur J Hum Genet 8:750–756

Endler JA (1977) Geographic variation, speciation, and clines. Princeton University Press, Princeton

Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A (1991) Spatial differentiation of *RH* and *GM* haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities. Hum Biol 63:273–307

Fix AG (1999) Migration and colonization in human microevolution. Cambridge University Press, Cambridge

Gamble C (1993) Timewalkers: the prehistory of global colonization. Harvard University Press, Cambridge

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427–441

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

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonné-Tamir B (2000) Jewish and middle eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc Natl Acad Sci USA 97:6769–6774

Hammer MF, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. Evol Anthropol 5:116–134

Jorde LB (1980) The genetic structure of subdivided human populations: a review. In: Mielke JH, Crawford MH (eds) Current developments in anthropological genetics. Plenum, New York, pp 135–208

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Pyschometrika 29:1–27

Kuzmin Y, Orlova LA (1998) Radiocarbon chronology of the Siberian Paleolithic. J World Prehist 12:1–53

Lahr MM, Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol Suppl 27:137–176

Malécot G (1968) The mathematics of heredity. Freeman, San Francisco

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Matsumoto H (1988) Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. Hum Genet 80:207–218

Morton NE (1973) Genetic structure of populations. University of Hawaii Press, Honolulu

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10:927–943

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015–1035

Przeworski M, Hudson RR, Di Rienzo A (2000) Adjusting the focus on human variation. Trends Genet 16:296–302

Qian Y, Qian B, Su B, Yu J, Ke Y, Chu Z, Shi L, Lu D, Chu J, Jin L (2000) Multiple origins of Tibetan Y chromosomes. Hum Genet 106:453–454

Renfrew C (1991) Before Babel: speculations on the origins of linguistic diversity. Cambridge Archaeol J 1:3–23

Ruhlen M (1992) An overview of genetic classification. In: Hawkins JA, Gell-Mann M (eds) The evolution of human languages. Addison-Wesley, Redwood City, pp 159–189

Rohlf FJ (1998) NTSYS-pc: numerical taxonomy and multivariate analysis system. Exeter Software, Setauket

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Schneider S, Kueffer J-M, Roessli D, Excoffier L (1998) Arlequin: a software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354–7359

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Slatkin M, Maddison WP (1990) Detecting isolation by distance using phylogenies of genes. Genetics 126:249–260

Sokal RR (1979) Ecological parameters inferred from spatial correlograms. In: Patil GP, Rozenzweig M (eds) Contemporary quantitative ecology and related econometrics. International Cooperative Publishing House, Fairland, pp 167–196

Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthropol 80:267–294

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. Biol J Linn Soc 10:229–249

Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza LL, Chakraborty R, Jin L (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–590

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza LL, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (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

Sugimoto C, Kitamura T, Guo J, Al-Ahdal MN, Shchelkunov SN, Otova B, Ondrejka P, Chollet YJ, El-Safi S, Ettayebi M, Gresenguet G, Kocagoz T, Chaiyarasamee S, Thant KZ, Thein S, Moe K, Kobayashi N, Taguchi F, Yogo Y (1997) Typing of urinary JC virus DNA offers a novel means of tracing human migrations. Proc Natl Acad Sci USA 94:9191–9196

Templeton AR (1998) Human races: a genetic and evolutionary perspective. Am Anthropol 100:632–650

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genetics 140:767–782

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics 134:659–669

Turner C (1990) 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

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH,

Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd K, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner P (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Wright S (1943) Isolation by distance. Genetics 28:114–138

Zegura SL, Simic D, Rudan P (1995) Malécot's isolation by distance model: empirical behavior and theoretical considerations. J Quant Anthropol 5:171–189

Zhao TM, Lee TD (1989) *Gm* and *Km* allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum Genet 83:101–110