STUDIES ON THE GEOGRAPHIC DISTRIBUTION OF HUMAN Y CHROMOSOME DNA VARIATION:

INFERENCES OF THE HUMAN DISPERSIONS FROM THE PHYLOGEOGRAPHIC POINT OF VIEW

Atsushi Tajima

DOCTOR OF SCIENCE

Department of Biosystems Science
School of Advanced Sciences
The Graduate University for Advanced Studies (Sokendai)

2001 (School Year)

ACKNOWLEDGMENTS

This dissertation could have been made impossible without any support. Foremost, I wish to express my sincere gratitude to my supervisor, Satoshi Horai, who has provided me every possible opportunity to mature my ideas under his continuous guidance.

I am deeply obliged to Naoyuki Takahata, Tomoya Akimichi, Yoko Satta, and Satoshi Horai for willingly being on my doctoral committee. Their useful comments on my researches have enabled me to finish this dissertation.

I would like to thank all the men who donated DNA. Without their consent to the researches, I could not have completed this dissertation. Also, I am grateful to I-Hung Pan, Goonnapa Fucharoen, Supan Fucharoen, Masafumi Matsuo, Katsushi Tokunaga, Takeo Juji, Masanori Hayami, and Keiichi Omoto for providing precious DNA and blood samples.

I would like to thank all the members of Department of Biosystems Science at the Graduate University for Advanced Studies (Sokendai). Their enthusiasms for advanced sciences have motivated me. In addition, I am grateful to Naoko Takezaki and Tatsuya Ota for helpful advice and comments on data analyses. I would also like to thank Sang-Hee Lee for improving my writing in English.

Several have provided generous help with this dissertation. In particular, I would like to thank Peter Underhill. Our discussion on my researches in China greatly helped to clarify the subjects. I am also grateful to Isao Yuasa, Jiayou Chu, and Yaping Qian, who have given me valuable comments. Moreover, I am much indebted to Yuttana Pengcham for his support

and hospitality during my stay at Khon Kaen University in Thailand.

Finally, I dedicate this dissertation to my wife, Chiko as a token of my appreciation. Her devoted supports let me concentrate only on my studies throughout the course of Ph.D. I deeply thank her.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

CHAPTER 1

1 Introduction

1.1 Background

Many studies of genetic variation in human populations have revealed a recent common African origin of modern humans (Cann et al. 1987; Nei and Roychoudhury 1993; Horai et al. 1995; Nei and Takezaki 1996; Hammer et al. 1998). According to this evolutionary history, modern human populations have dispersed into various parts of the world outside Africa for the past 150,000 years or so. Moreover, there have been many lines of molecular genetic evidence showing geographic variation of the human genome (e.g., Hammer et al. 1997). These genetic differences must be a result of multiple evolutionary forces (such as mutation, genetic drift and selection) and evolutionary events (such as range expansion, migration and isolation by distance). However, there is incomplete knowledge of how geographic patterns of variation have been shaped (Przeworski et al. 2000).

In general, time and space are jointly considered to deduce a population history from genetic variation in contemporary populations. In other words, we have to examine (1) spatial distributions of genetic variants (such as alleles and haplotypes) and (2) phylogenetic (historical) relationships among the variants. Recently, this field of study has been referred to as *phylogeography* (Avise et al. 1987).

Phylogeography is a sub-discipline of biogeography, and mainly deals with principles and processes governing geographic distributions of genetic variants whose

phylogenetic relationships can be inferred (Avise 2000). Most phylogeographic analyses of human genetic variation have been conducted with data for mitochondrial DNA (mtDNA) polymorphism, because mtDNA has phylogeographically favorable properties: the maternal transmission, extensive intraspecific variation, and absence of intermolecular recombination (Avise 2000). However, mtDNA is a single locus, and can be used to trace only maternal lineage in the human evolutionary history. For verifying the maternal view of the past, much attention has turned to the human Y chromosome, which is treated as a counterpart of mtDNA in the inherited manner.

1.2 Potential Contributions from Y-Chromosome Studies

The human Y chromosome is composed of a non-recombining region (NRY), flanked by pseudoautosomal regions. The NRY accounts for 95% of its length (about 60 Mb), and is the largest non-recombining region in the human genome (Jobling and Tyler-Smith 1995). This means that the NRY has been paternally inherited as a haploid genome during the human evolutionary history. Thus, the NRY and mtDNA have similarities in their mode of inheritance and lack of recombination (Table 1-1).

The phylogeographic analysis with haplotypes (which consist of several linked polymorphisms) generally has an advantage of obtaining more robust phylogeny for the genetic variants, rather than simply analyzing individual polymorphisms at random (Avise 2000; Stoneking 2001). Because of the lack of recombination, any combinations of polymorphisms on the NRY can be regarded as *Y-chromosome haplotypes*. As summarized in Table 1-1, the mutation rate *per haplotype* for the NRY is expected to be higher than that for mtDNA, while the opposite is evidently observed for the base substitution rate *per site*. These

imply that Y-chromosome haplotype carries more detailed information on the human evolutionary history than mtDNA haplotype, if DNA polymorphisms are examined in long stretch of the NRY which is more than 30 times as large as that of mtDNA. In addition, the lower mutation rate *per site* for the NRY usually allows to deduce an ancestral state more easily at every polymorphic site on the NRY than mtDNA, leading to a more reliable phylogenetic inference among the haplotypes.

Table 1-1. Comparisons between the NRY and mtDNA in human

	NRY[a]	mtDNA
Mode of inheritance	Paternal	Maternal
Meiotic recombination	No	No
Nucleotide length of the non-recombining region	60Mb (30Mb; on excluding the heterochromatin region)	16.5Kb
Substitution rate (/site/year)[b]	1.24×10^{-9}	3.89×10^{-8}
Expected mutation rate[c] (/haplotype/year)	7.4×10^{-2}	6.4×10^{-4}

[a] non-recombining region of the human Y chromosome

[b] for NRY, Thomson et al. (2000); for mtDNA, Horai et al. (1995)

[c] Calculated by multiplying the substitution rate and the length of non-recombining region

Recent studies have revealed the effectiveness of Y-haplotype analyses on the human evolution (Hammer et al. 1998, 2001; Underhill et al. 2000), and contributed to increase the number of available polymorphic markers on the NRY (The International SNP Map Working Group 2001; Underhill et al. 2001). However, the findings do not answer all of the questions about the human evolutionary history. In particular, questions regarding human migrations to specific regions are unsolved (see next section). To clarify a specific migration of interest, two

types of studies are required: one is to investigate geographic variation of Y-haplotypes minutely in target populations (or regions), and the other is to find population-specific (or region-specific) polymorphisms for an extensive phylogeographic analysis with Y-haplotypes.

1.3 The Question: Human Migrations to East and Southeast Asia

East and Southeast Asia are considered to be relatively important regions for human dispersal during the late Pleistocene and the Holocene, because the regions must have played principal roles in subsequent migrations to Australia, Oceania, and North Asia (e.g., Karafet et al. 2001). However, our understanding of the prehistoric settlements of modern humans in East and Southeast Asia is unsatisfactory (Cavalli-Sforza et al. 1994; Lahr and Foley 1994; Cann 2001).

As a result of many studies on various kinds of biological diversities among Asian populations, there are several hypotheses for the peopling of East and Southeast Asia. These hypotheses can be divided broadly into two categories: a single migration model, and a model with several waves of migration. From the dental morphological study of Turner (1990), the first model for a single migration event postulates that Southeast Asia is the homeland for all modern Asian populations, followed by their expansions toward the north and Oceania. Recent genetic studies of Asian populations with microsatellite DNA markers (Chu et al. 1998) and Y-chromosome haplotypes (Su et al. 1999; Jin and Su 2000) have supported this model. In contrast, the second model for multiple waves of migration has been proposed following the analyses of various polymorphic loci, such as classic genetic markers (Nei and Ota 1991; Omoto 1995), mtDNA haplotypes (Horai and Hayasaka 1990; Horai et al. 1996), human leukocyte antigen (HLA) haplotypes (Tokunaga et al. 1996) and Y-chromosome

haplotypes (Underhill et al. 2001). Each hypothesis explains part of the existing evidence in relation to the diversities among Asian populations, while none does the whole in a general way. Thus, the current situation demands to set up a verifiable hypothesis, but not to test the available hypotheses.

A hypothesis regarding the human migrations into East and Southeast Asia should explain at least the demographic history of the following populations: the Ainu in Japan, and aboriginal Australians. There is some agreement that each of these two modern populations is mostly composed of a direct descendant of people who had colonized the respective area (the Japanese archipelago or Australia) in the late Pleistocene (Turner 1990; Hanihara 1991). This may allow to consider that there has been a certain amount of genetic continuity in each population up to the present. In fact, based on mtDNA sequences from ancient Japanese bones (3,000-6,000 years BP) and Ainu bones (200-300 years BP), Horai et al. (1991) have reported that the two temporal groups of ancient Japanese show a close phylogenetic affiliation, suggestive of temporal genetic continuity in the Ainu people.

1.4 The Question: Peopling of the Japanese Archipelago

The evolutionary history of Japanese populations is another controversial subject in relation to the human dispersal in East and Southeast Asia (reviewed by Hanihara 1991). Currently, there are two major rival hypotheses for the formation of modern Japanese populations, viz., the transformation theory and the hybridization theory. Both theories accept that human migration from the Asian Continent to the Japanese archipelago occurred at least twice: once each in the late Pleistocene and in the Holocene. The transformation theory claims that the Japanese populations originated as a single ancient population from either Southeast

Asia (including southern China; Suzuki 1969; Turner 1990) or Northeast Asia (Nei 1995) in the late Pleistocene. Namely, this theory posits that the immigrants (the so-called Jomon people) have gradually transformed their genetic and morphological characteristics to form modern Japanese populations, while the subsequent immigrants (the so-called Yayoi people) are considered not to contribute genetically to the contemporary populations (Mizoguchi 1986; Nei 1995).

In the hybridization theory, however, genetic admixture of the two immigrants with different ancestries and in different periods of time is posited to form the modern Japanese (Hanihara 1991; Horai et al. 1996; Omoto and Saitou 1997). Several regional differences in genetic and physical characteristics of Japanese populations are interpreted as different extent of intermixture from region to region. In this context, the Ainu and Okinawans (living in Okinawa of the southwestern islands in Japan; the so-called Ryukyuans) are considered to be pure descendants of the Jomon people, although several oppositions were recently raised from genetic (Hatta et al. 1999) and morphological studies (Dodo et al. 1998).

In sum, it makes the difference between the two theories whether Yayoi people have contributed to the gene pool of modern Japanese populations. Furthermore, there are questions as to where Jomon people came from, and whether the Ainu and Okinawan populations have a common genetic background. To address these questions, it is important to understand the formation process of modern Japanese populations in a series of the human dispersal in East and Southeast Asia.

1.5 Objectives of the Study

This thesis consists of five chapters including this introduction. Chapter 2 discusses

early migration events of modern humans into East and Southeast Asia. For a phylogeographic analysis with Y-chromosome haplotypes, seven known polymorphisms on the NRY are analyzed in 610 males from 14 global populations in North, East, and Southeast Asia, and other regions of the world. On the basis of eight Y-haplotype frequencies, the paternal relationships of the 14 populations are examined. From the phylogenetic relationships of the Asian populations and the geographic distributions of Y-haplotypes in Asia, this chapter proposes a hypothesis on the peopling of East and Southeast Asia.

Chapter 3 deals with a specific subject in order to find population-specific (or region-specific) polymorphisms for an extensive phylogeographic analysis. To discover new single-nucleotide polymorphisms (SNPs), a 12.6-kb region on the NRY is sequenced in eight males from East Asia, Southeast Asia, Europe, and Africa. For verifying and characterizing novel SNPs, other population samples are genotyped with newly developed PCR-RFLP (restriction fragment length polymorphism) methods.

In Chapter 4, an extensive Y-haplotype analysis with a total of 15 polymorphic markers (including the novel SNPs) in the 14 populations is performed to consider the aforementioned problems regarding the formation of modern Japanese populations. In particular, the 'Ainu-Okinawan problem' is centrally discussed as viewed from the Y-haplotypes. Chapter 5 summarizes the thesis.

CHAPTER 2

2 Three major lineages of modern Asian populations : Implications for the peopling of East and Southeast Asia

As summarized in Chapter 1, formation process of modern human populations in East and Southeast Asia is one of the controversial matters. From a phylogeographic point of view, this is probably because the evolutionary relationships among geographic variants have been uncertain in many genetic and morphological studies (e.g., Nei and Ota 1991; Omoto 1995). Moreover, Australo-Melanesian populations (such as aboriginal Australians and Papua New Guineans) have been mostly excluded from the analyses (e.g., Horai et al. 1996; Chu et al. 1998; Su et al. 1999), although they are necessary to discuss early migration events of modern humans into East and Southeast Asia. Especially, aboriginal Australians are crucial when one hypothesizes and verifies the peopling of East and Southeast Asia as described in Chapter 1.

For a phylogeographic analysis with Y-chromosome haplotypes, the primary objectives of this chapter are (1) to clarify spatial distributions of the Y-chromosome DNA variation in 14 modern human populations (including 12 Asian populations), and (2) to obtain robust phylogenetic relationships among the haplotypes. In combination with the previous findings, a phylogenetic tree for a total of 20 Asian and Australo-Melanesian populations (including aboriginal Australians) is constructed to examine the paternal relationships on the basis of the Y-haplotype frequencies. Furthermore, the coalescence of the observed haplotypes is analyzed to estimate the ages of each of mutations, leading to a further

understanding of human dispersal. Finally, this chapter deduces patterns of human migrations to prehistoric East and Southeast Asia.

2.1 Materials and Methods

2.1.1 Subjects and DNA Samples

A total of 610 unrelated males from 14 global populations (including 12 Asian populations) were analyzed. The DNA samples were as follows: 61 Buryats in Siberia (Tokunaga et al. 1995); 21 Nivkhi in Sakhalin (Yamashita et al. 1997); 16 Japanese-Ainu (JP-Ainu) in Hokkaido, the northern island of Japan (Harihara et al. 1988); 82 Japanese-Honshu (JP- Honshu; 55 from Shizuoka and 27 from Aomori), 45 Japanese-Okinawa (JP-Okinawa) and 21 Taiwan Han Chinese (Horai 1991; Hammer and Horai 1995); 49 Han Chinese in northern China (Tokunaga et al. 1998); 34 Thais (19 from Khon Kaen and 15 from Chiang Mai; Fucharoen et al. 2001); 46 Jewish living in the Uzbekistan (39 from Bukhara and 7 from Samalkand; Kato et al. 1997); and four Africans (one each from Zaire, Nigeria, Zambia, and Zimbabwe; Horai and Hayasaka 1990). In addition, blood samples were collected from the following populations: 104 Japanese living in the Kyushu islands in southwestern Japan (JP-Kyushu); 65 Thai-Khmers in Thailand; 12 Malaysians in Malaysia; and 50 Philippines in the Philippines. All the individuals gave their informed consent prior to their inclusion in the study. Genomic DNA was isolated from the buffy coat in blood by treatment with sodium dodecyl sulfate and proteinase K, and subsequent phenol/chloroform extraction.

2.1.2 Genotyping of Biallelic Polymorphisms

Seven biallelic polymorphic sites studied were as follows (Fig. 2-1): the $DYS257_{108}$

G→A transition (Hammer et al. 1998); the *DYS287* Y *Alu* polymorphic (YAP) element insertion (Hammer 1994); the SRY_{4064} G→A transition and SRY_{10831} A↔G recurrent transition (Whitfield et al. 1995; Hammer et al. 1998); the M9 C→G transversion and M15 9-bp insertion (Underhill et al. 1997); and the $RPS4Y_{711}$ C→T transition (Bergen et al. 1999). Genomic DNA was used as template to amplify a DNA fragment encompassing the polymorphic site by the polymerase chain reaction (PCR) method. Each 25-µl PCR mixture contained 20-50 ng genomic DNA, 0.12-0.2 µM each primer, 0.2 mM each dNTPs, 1.5-3.0 mM $MgCl_2$, 1 × GeneAmp PCR Buffer II (10 mM Tris-HCl, pH 8.3; 50 mM KCl), and 0.5 U Ampli*Taq* DNA polymerase (Applied Biosystems, USA).

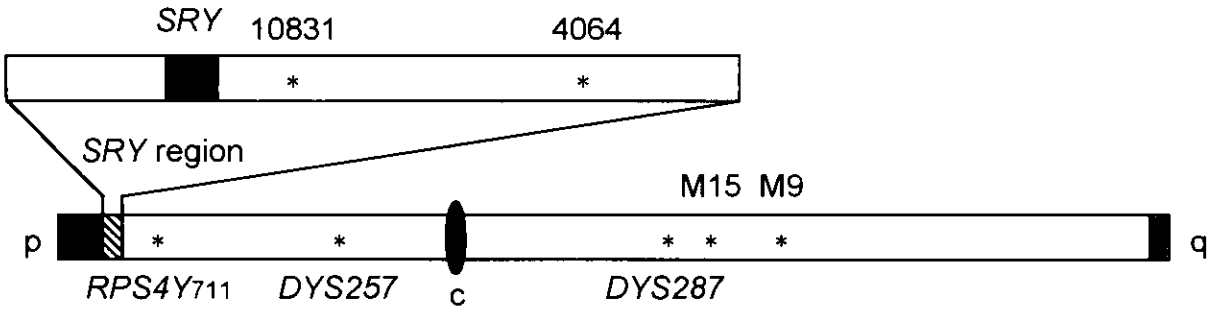


Figure 2-1. Schematic representation of human Y chromosome and chromosomal locations of the polymorphic sites examined. The short arm (p), the centromere (shaded oval; c), and the long arm (q) of the Y chromosome are indicated in the bottom illustration. The shaded boxes at the termini of the chromosome represent the pseudoautosomal regions. The top shows scaled enlargement of the *SRY* region. The seven asterisks represent the polymorphic sites examined.

$DYS257_{108}$ and *DYS287* (YAP) were genotyped by the method of Hammer et al. (1998) and Hammer and Horai (1995), respectively. DNA fragments for SRY_{4064}, SRY_{10831} and M9 were amplified with the primer sets reported by Qamar et al. (1999), Hammer et al.

(1998) and Thomas et al. (1999), respectively. PCR product containing $RPS4Y_{711}$ was obtained by amplification with the following primer set: 5'-GATTTTGTTGGGTGGTGGTC-3' and 5'-TGCTGCTACTGCAATTTAGCC-3'. The cycling condition for $RPS4Y_{711}$ was 94°C for 2 min, and then 30 cycles of 94°C for 30 s, 57°C for 30 s and 72°C for 30 s in GeneAmp PCR System 9600 (Applied Biosystems, USA). For typing the allele, the PCR products for SRY_{4064}, SRY_{10831}, M9 and $RPS4Y_{711}$ were digested with the appropriate restriction endonuclease (*Bsr*BI, *Mae*III, *Hin*fI and *Bsl*I, respectively). The digested DNA fragments were separated by electrophoresis on 2% agarose gel (Bio-Rad Laboratories, USA) in 0.5 × TBE buffer, and detected by staining with ethidium bromide for genotyping. The DNA fragment for M15 was amplified with the following primer set: 5'-CCTCATGCGCATATACAATCA-3' and 5'-CCACTGCACCTAGGGAGACA-3'. The cycling condition for M15 was 94°C for 2 min, and then 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s. The size of PCR product (86 bp or 95 bp) for M15 was determined by electrophoresis on 4% NuSieve 3:1 agarose gel (BME, USA) in 1 × TBE buffer, followed by staining with ethidium bromide (more detailed experimental conditions for the seven polymorphims are summarized in Appendix A. Electrophoretic detection of the allele variation for each of the seven polymorphic sites is typically shown in Appendix B).

2.1.3 Phylogenetic Analysis of Y-Chromosome Haplotypes

Y-chromosome haplotypes were inferred from the analyses of all the seven polymorphic sites. A phylogenetic tree for all Y-haplotypes was constructed by the maximum parsimony method (Fitch 1977). To infer the root in the tree, the orthologous nucleotide sequences in several non-human primates (Hammer 1995; Whitfield et al. 1995; Underhill et

al. 1997; Hammer et al. 1998; Bergen et al. 1999) were used as outgroups.

2.1.4 Characterization of Genetic Polymorphisms for Y-Haplotypes

To estimate the demographic parameters for the populations, the extent of genetic differentiation of the populations was estimated by the G_{ST} statistic (Nei 1987) on the basis of the Y-haplotypes. The distribution of pairwise sequence differences in 610 male samples was also calculated.

2.1.5 Gene Genealogy of Y-Haplotypes

Coalescence analysis of the Y-haplotype tree was performed to estimate the time back to the most recent common ancestor (TMRCA), and the age of each mutation with the GENETREE program, conditional on the topology of the haplotype tree (Griffiths and Tavaré 1994; for GENETREE, http://ftp.monash.edu.au/pub/). Under the infinitely-many-sites mutation model, mutations were supposed to occur by a Poisson process with rate $M/2$, where $M = 2N_e\mu$ (N_e is the effective population size of males, and μ is the total mutation rate per sequence per generation). For application of this program to entire data set, two insertions at *DYS287* and M15 were regarded as single-nucleotide substitutions. Maximum likelihood estimate of M (M_{ML}) was measured under a panmictic (random mating) population model with a constant N_e (for verification of the assumptions, see Section 2.2.3). By using M_{ML} as a parameter, the mean and standard deviation (SD) of the TMRCA, and the ages of mutations in units of N_e generations were estimated. All the estimates were computed by Markov chain simulation techniques with one million replicate runs. To convert the expected time in units of N_e generations into time in years, a long-term N_e of 5,000 and a 20-year generation time were

assumed (Hammer 1995).

2.1.6 Population Analysis from Y-Haplotypes

On the basis of the Y-haplotype frequencies, D_A genetic distance (Nei et al. 1983) between populations was calculated with the POPTREE program (provided by N. Takezaki; for POPTREE, http://www.bio.psu.edu/People/Faculty/Nei/Lab/Programs.html). The D_A distance is defined as

$$D_A = 1 - \sum_{i=1}^{q} \sqrt{x_i y_i}$$

where q is the number of haplotypes, and x_i and y_i are the frequencies of the i-th haplotype in population X and Y, respectively. A phylogenetic tree for the 14 populations was constructed from the matrix of pairwise D_A distances by the neighbor-joining (NJ) method (Saitou and Nei 1987). The reliability of the NJ tree was not examined by a bootstrap test (Felsenstein 1985) because the present Y-haplotypes could be considered as a single-locus system with multiple alleles.

2.2 Results and Discussion

2.2.1 Evolutionary Relationships of Y-Chromosome Haplotypes

By analyzing the seven biallelic polymorphic markers on the NRY, eight haplotypes were observed and named ht1-ht8. Figure 2-2 shows the maximum parsimony tree with a root for the eight haplotypes. This haplotype tree had no ambiguity, although the recurrent transitional mutations at SRY_{10831} have been suggested (Hammer et al. 1998; Ramana et al. 2001). The phylogenetic tree also coincided with that reported by Hammer et al. (2001) (for

correspondences regarding the nomenclature of haplotypes, see legend to Fig. 2-2).

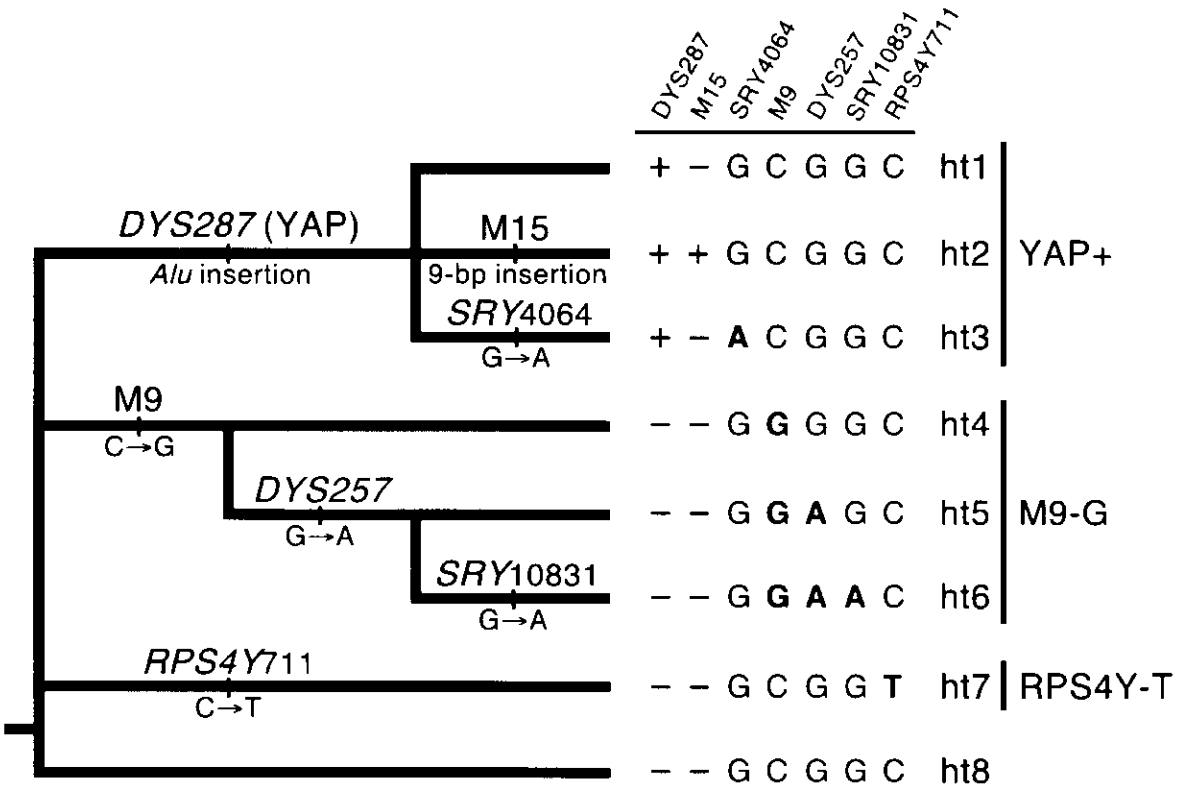


Figure 2-2. Eight Y-chromosome haplotypes on the basis of seven biallelic markers, and their most parsimonious relationships. +, – Alleles with insertion and no insertion, respectively, at *DYS287* and M15; *A, T, G, C* nucleotide sequences of five polymorphic sites (SRY_{4064}, M9, $DYS257_{1085}$, SRY_{10831} and $RPS4Y_{711}$). *Vertical bars* on the branches of the Y-haplotype tree indicate the expected mutational events at the seven polymorphic sites. Eight haplotypes (ht1-ht8) are defined by the allelic associations among the seven polymorphic sites. Three major lineages are indicated *right* with the lineage names: YAP$^+$, M9-G, and RPS4Y-T. The correspondences between the haplotype designation and that reported by Hammer et al. (2001) are as follows: *ht1* haplotype 11; *ht2* haplotype 12; *ht3* haplotypes 13-15; *ht4* haplotypes 24-35; *ht5* haplotypes 36, 37, and 39; *ht6* haplotype 38; *ht7* haplotypes 16-18; *ht8* haplotypes 1-10 and 19-23.

In general, almost all SNPs on the NRY are biallelic, but some are known to be reversion and triallelic polymorphism, each of which is caused by multiple mutation at the same nucleotide site (Hammer et al. 1998; Underhill et al. 2000). However, all the multiple markers can define different haplotypes in combination with other markers, and make no ambiguous relationship among the haplotypes (Underhill et al. 2000). Therefore, the Y-haplotype tree in this population survey may be considered as a robust phylogeny.

The eight haplotypes could be classified into four distinct lineages defined by three key mutations: an insertion of the YAP element at *DYS287*, a C-to-G transversion at M9, and a C-to-T transition at $RPS4Y_{711}$. These lineages were designated as YAP^-, M9-G, and RPS4Y-T, respectively (Fig. 2-2). Of the four lineages, three major lineages (YAP^-, M9-G, and RPS4Y-T) accounted for 98.6% of the Asian samples as shown in Table 2-1. An important point is that the evolutionary relationships of the three major lineages are not hierarchical (Underhill et al. 2000), indicative of independent occurrence of the respective lineages during the human evolutionary history. Furthermore, all the three key markers that distinguish the Y-lineages here also occupy important positions in the Y-chromosome DNA phylogeny from worldwide human populations (Hammer et al. 2001; Underhill et al. 2000, 2001). Thus, this classification of the Y-haplotypes is adequate for characterizing the paternal composition of Asian populations, indicating that modern Asian populations mainly consist of the three paternal lineages (Ke et al. 2001).

2.2.2 Geographic Distributions of Y-Chromosome Haplotypes

The eight Y-haplotypes were unevenly distributed among the 14 populations. On the basis of their localities, these populations were divided into four groups: North Asian, East

Asian, Southeast Asian, and other (non-Asian) populations (see Table 2-1). The frequencies of the Y-haplotypes for the populations are shown in Fig. 2-3.

Table 2-1. Comparison of Y-chromosome haplotype frequencies in 14 populations

Population	*N*	Haplotype frequency (%)[b]							
		ht1[a]	ht2	ht3	ht4	ht5	ht6	ht7	ht8
North Asian									
Buryat	61	0	0	0	9.8	0	0	83.6	6.6
Nivkhi	21	0	0	0	28.6	19.0	9.5	38.1	4.8
East Asian									
JP-Ainu[c]	16	87.5	0	0	0	0	0	12.5	0
JP-Honshu[c]	82	36.6	0	0	57.3	0	0	6.1	0
JP-Kyushu[c]	104	27.9	0	0	59.6	0	0	11.5	1.0
JP-Okinawa[c]	45	55.6	0	0	37.8	2.2	0	4.4	0
Northern Han	49	0	0	0	83.7	8.2	0	8.2	0
Taiwan Han	21	0	0	0	81.0	0	4.8	14.3	0
Southeast Asian									
Thai	34	0	2.9	0	88.2	2.9	0	2.9	2.9
Thai-Khmer	65	0	1.5	0	84.6	1.5	10.8	1.5	0
Malaysian	12	0	0	0	83.3	0	8.3	0	8.3
Philippine	50	0	0	0	96.0	2.0	2.0	0	0
Total (Asia)	560	17.5	0.4	0.0	60.5	2.1	2.1	15.9	1.4
Other									
Jewish-Uzbekistan	46	0	0	28.3	6.5	2.2	2.2	0	60.9
African	4	0	0	100	0	0	0	0	0
Overall	610	16.1	0.3	2.8	56.1	2.1	2.1	14.6	5.9

[a] The haplotypes were defined by the following polymorphic sites and the respective derived alleles: *ht1* YAP$^+$; *ht2* YAP$^+$, M15$^+$; *ht3* YAP$^+$, SRY_{4064}-A; *ht4* M9-G; *ht5* M9-G, $DYS257_{108}$-A; *ht6* M9-G, $DYS257_{108}$-A, SRY_{10831}-A; *ht7* $RPS4Y_{711}$-T; *ht8* none.

[b] Some rows do not sum to 100% because the frequencies (%) were rounded to one decimal place.

[c] *JP* Japanese

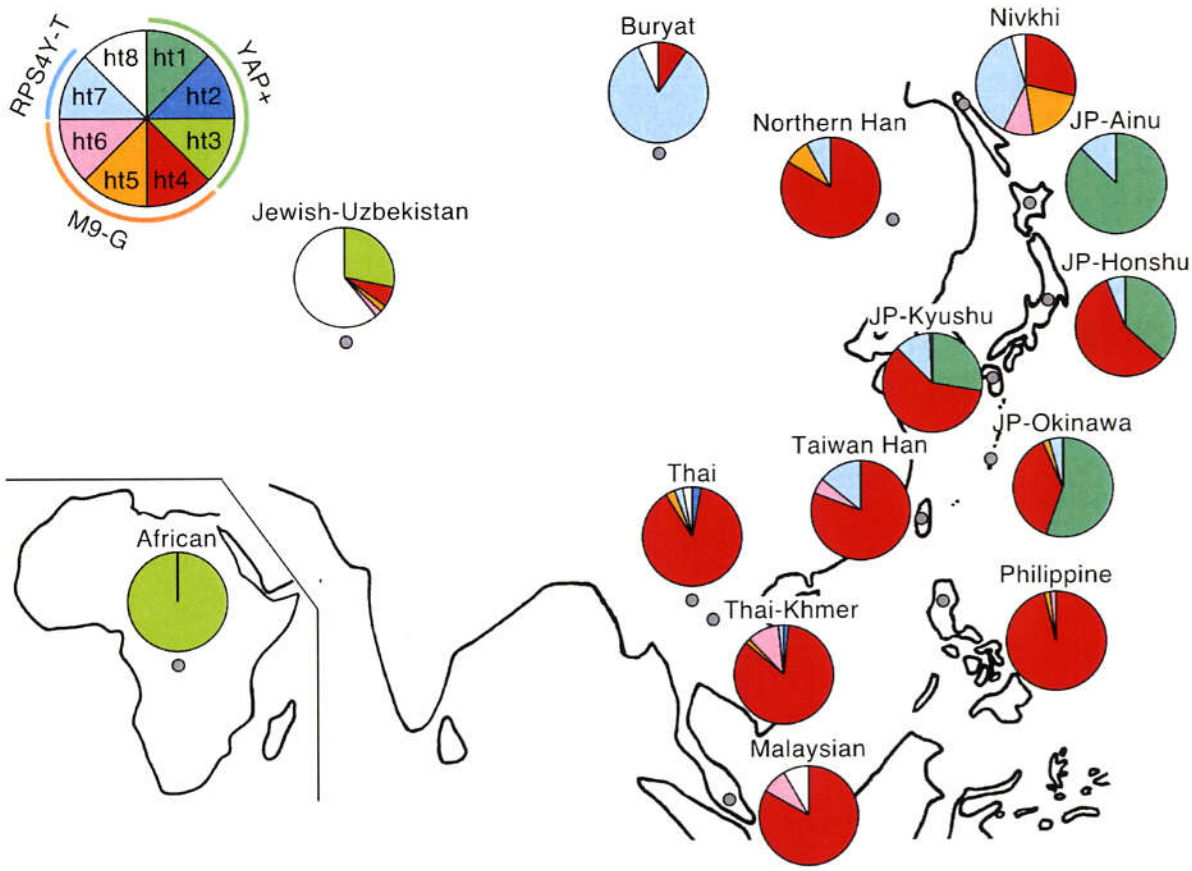


Figure 2-3. Frequency distributions of the eight Y-chromosome haplotypes for the 14 global populations, with their approximate geographic locations. The frequencies of the eight haplotypes are shown as *colored pie charts* (for color codes, see *upper left insert*). *JP* Japanese.

Only four Japanese populations exhibited ht1 (defined only by YAP^+) at various frequencies (also see Table 2-1). The highest frequency (87.5%) was found in JP-Ainu, followed by JP-Okinawa (55.6%) living in the southwestern islands of Japan, JP-Honshu (36.6%) and JP-Kyushu (27.9%). The ht2 haplotype (defined by $YAP^+/M15^+$) was found in only two males, one each from Thais and Thai-Khmers. The ht3 haplotype (defined by YAP^+/SRY_{4064}-A) was completely absent in the Asian populations examined, whereas Jewish in the Uzbekistan and African populations had this haplotype with a frequency of 28.3% and

100%, respectively (Hammer et al. 2000). Thus, the YAP$^+$ lineage was found in restricted populations among Asian populations, consistent with the previous reports (Hammer and Horai 1995; Hammer et al. 1997; Shinka et al. 1999).

The ht4 haplotype (defined only by M9-G) was widely distributed among North, East, and Southeast Asian populations, except for the Ainu. This haplotype was frequent (60.5%) in overall Asian populations (Table 2-1). Among them, the Han Chinese and Southeast Asian populations were characterized by high frequencies ranging from 81.0% to 96.0%. In contrast to ht4, ht5 (defined by M9-G/$DYS257_{108}$-A) and ht6 (defined by M9-G/$DYS257_{108}$-A /SRY_{10831}-A) were small contributors to Asian populations. The highest frequency of ht5 was observed in Nivkhi (19.0%), and that of ht6 in Thai-Khmers (10.8%). The ht5 haplotype is widely distributed among European, Asian, and Native American populations, and is proposed to be one of the candidates for the founder haplotypes in the Americas (Karafet et al. 1999). Furthermore, high frequencies of ht6 are observed in North Europe, Central Asia, and India (Karafet et al. 1999). Thus, the presence of ht5 in Nivkhi may account for the founder effect of peopling of the Americas.

The ht7 haplotype (defined by $RPS4Y_{711}$-T) was also widely distributed throughout Asia excluding Malaysia and the Philippines, whereas this was absent in two non-Asian populations. The highest frequency of ht7 was found in Buryats (83.6%), followed by Nivkhi (38.1%). Thus, the geographic distribution of ht7 in Asia appears to contrast with that of ht4.

Only the eight individuals (1.4%) in Asia belonged to ht8, which was the major haplotype in the Jewish population (Table 2-1). The ht8 haplotype may not be useful for inferring the population relatedness among Asian populations because it is defined by no lineage-specific mutations. Additional Y-polymorphic markers such as M89 and M168

(Underhill et al. 2000; Ke et al. 2001) will be needed to investigate details of the formation of modern Asian populations.

2.2.3 Coalescence Analysis in the Y-Haplotype Tree

The genealogical inference with Y-chromosome DNA variation is affected by demographic characteristics such as structure and size of human populations (Pritchard et al. 1999), but there is no general agreement on methodology (Bertranpetit 2000; Stumpf and Goldstein 2001). In this section, coalescence analysis of the observed Y-haplotypes was performed to reconstruct the population history of modern humans.

To measure the effect of population subdivision on the coalescent estimation, the degree of genetic differentiation of populations was measured by the G_{ST} statistic (Nei 1987). The African population was excluded from this measurement because there were only four Africans in this study, possibly leading to an unreliable conclusion. The estimated G_{ST} in a total of 13 populations was 0.42 (on including the Africans, it was estimated to be 0.49). This value did not differ greatly from the published estimates (0.230-0.645) for Y-chromosome DNA polymorphisms (Hammer et al. 1997, 2001; Karafet et al. 1997; Poloni et al. 1997; Seielstad et al. 1998; Kittles et al. 1999), although most other studies have sampled from worldwide populations. The present G_{ST} value is also consistent with those for mtDNA (0.31, Stoneking et al. 1990; 0.46, Merriwether et al. 1991). Furthermore, it is comparable to those for diploid nuclear loci (about 0.1, Takahata 1993), because it is appropriate to consider that the effective number of Y chromosomes is one-fourth as large as that of autosomes. According to a theoretical study on human genetic variation (Takahata 1993), the estimates for the nuclear loci allow to regard the entire human population as a single panmictic

population under Wright's island model of population structure. Therefore, this may be applied to the present Y-haplotype data because of the above equivalence between the G_{ST} values for the diploid genes and Y-chromosome DNA.

In addition, a bimodal distribution of sequence differences including insertions in all pairwise comparisons of 610 individuals was found in the overall populations (Fig. 2-4). A similar pairwise distribution was also observed in the 13 populations excluding Africans (data not shown). These suggest that there was no rapid expansion in population size, because a distribution with a single peak for nucleotide differences would be expected under a population-growth model (Sherry et al. 1994). The bimodal distribution also implies absence of recombination because intragenic recombination affects pairwise distribution, which is expected to be Poisson-like unimodal as the rate of recombination increases (Takahata and Satta 2002). This implication is in good agreement with the expectation from no meiotic recombination in the NRY. In any case, the model of constant population size fits these data better than that of population growth, consistent with the finding from European Y-chromosome DNA variation (Pereira et al. 2001). Therefore, a total of 610 individuals are assumed to be randomly sampled from a panmictic population with a constant N_e.

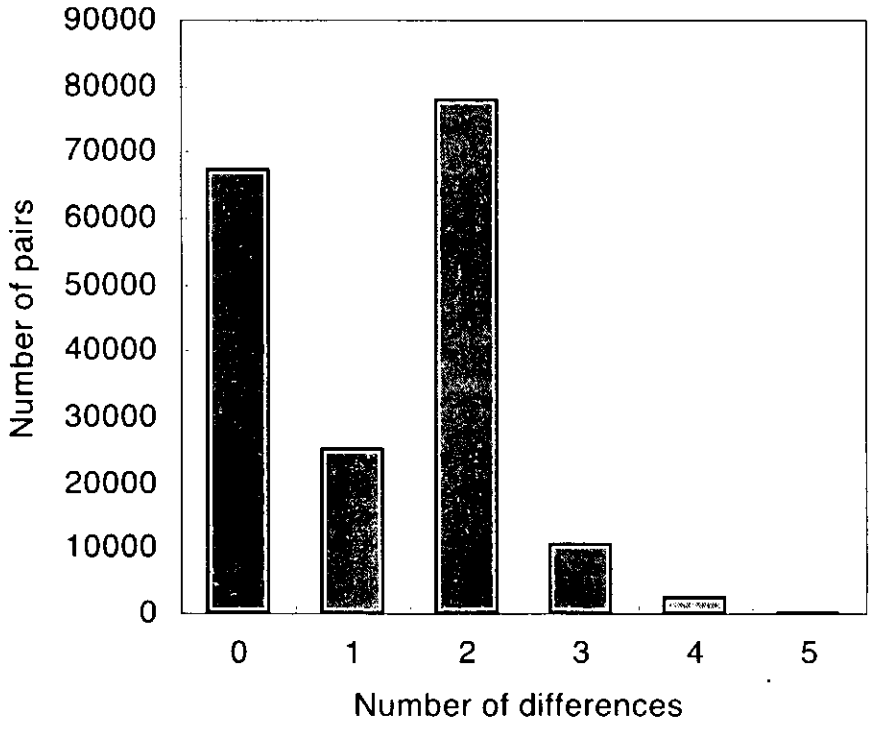


Figure 2-4. Frequency distribution of sequence differences for all possible pairs of the haplotypes in the 14 populations. The pairwise differences were calculated in 610 male individuals for the eight Y-haplotypes with seven segregating sites.

The maximum likelihood estimate of the parameter M (M_{ML}) was calculated at 1.0 with the GENETREE program, conditional on varying values of M ranging from 0.5 to 1.5 by 0.1. By using this M_{ML}, TMRCA of the Y-haplotypes, and the ages of mutations were estimated in units of N_e generations in the Y-haplotype tree. Table 2-2 shows their converted ages in years with stochastic errors, on the assumption that the long-term N_e is 5,000 with a generation time in humans of 20 years (Hammer 1995). This estimate for TMRCA was compatible with others based on the Y-chromosome DNA sequences (Hammer 1995; Underhill et al. 1997) and mtDNA sequences (Horai et al. 1995).

Table 2-2. TMRCA, and the estimated ages of mutations

	Coalescence Time[a] (in years; mean ± SD)
TMRCA	183,000 ± 70,000
Age of mutation	
DYS287 YAP insertion	69,000 ± 49,000
M15 9-bp insertion	2,300 ± 5,200
SRY_{4064} G→A	9,700 ± 11,000
M9 C→G	95,000 ± 51,000
$DYS257_{108}$ G→A	19,000 ± 16,000
SRY_{10831} G→A	5,300 ± 5,100
$RPS4Y_{711}$ C→T	53,000 ± 41,000

[a] Maximum likelihood estimate of M ($M_{\text{MLE}} = 1.0$) was used to compute the ages in the Y-haplotype tree under a constant population size model with the GENETREE program. To convert the estimated time into years, a long-term effective population size of 5,000 males with a 20-year generation time were assumed. The simulation results are based on one million replicate runs, and represented as the mean estimates ± SD.

All the lineages characterized by the three key mutations (YAP^-, M9-G, and RPS4Y-T) are geographically widespread in the world: YAP^+ in Africa, southern Europe, Middle East and East Asia; M9-G in Eurasia, Oceania and the Americas; RPS4Y-T in eastern Eurasia, Oceania and the Americas (Hammer et al. 1998; Karafet et al. 1999; Semino et al. 2000; Underhill et al. 2000; Wells et al. 2001). This means that the three mutations predate global population subdivisions in modern humans. The mean estimates for the three markers were 69,000 years ago for *DYS287* (YAP), 95,000 years ago for M9, and 53,000 years ago for $RPS4Y_{711}$, respectively (Table 2-2). These estimates are consistent with the above inferences of dating, although each has a large stochastic error. Therefore, the findings suggest that the three lineages (YAP^+, M9-G, and RPS4Y-T) were separated from one another during early stages of human evolutionary history.

2.2.4 Relationships of Populations on the Basis of Y-Haplotype Frequencies

To investigate relationships of the 14 populations, D_A distances (Nei et al. 1983) between populations were calculated on the basis of the frequencies of the eight haplotypes. The D_A distances between African and all the Asian populations showed the maximum value of 1.0, although the sample size of the African population was small (Table 2-3). The maximum D_A values (1.0) were also observed between the Ainu and three other populations (Malaysians, Philippines, and Jewish in the Uzbekistan), indicating that the Ainu has a unique paternal genetic background among Asian populations.

Table 2-3. D_A distances for 14 populations based on the frequencies of eight Y-chromosome haplotypes

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) BUR[a]													
(2) NIV	0.212												
(3) AIN	0.677	0.782											
(4) JPH	0.537	0.443	0.347										
(5) JPK	0.423	0.356	0.386	0.012									
(6) OKI	0.616	0.477	0.228	0.032	0.060								
(7) NTH	0.452	0.209	0.899	0.237	0.197	0.335							
(8) TWH	0.372	0.218	0.866	0.225	0.177	0.367	0.068						
(9) THA	0.507	0.281	0.940	0.247	0.200	0.362	0.043	0.090					
(10) THK	0.600	0.278	0.957	0.274	0.248	0.391	0.088	0.054	0.074				
(11) MAL	0.640	0.360	1.000	0.309	0.267	0.439	0.165	0.115	0.094	0.066			
(12) PHI	0.693	0.371	1.000	0.258	0.244	0.377	0.063	0.087	0.056	0.035	0.065		
(13) JWU	0.720	0.582	1.000	0.807	0.725	0.821	0.724	0.738	0.602	0.699	0.500	0.708	
(14) AFR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.468

[a] *BUR* Buryat, *NIV* Nivkhi, *AIN* Japanese-Ainu, *JPH* Japanese-Honshu, *JPK* Japanese-Kyushu, *OKI* Japanese-Okinawa, *NTH* Northern Han, *TWH* Taiwan Han, *THA* Thai, *THK* Thai-Khmer, *MAL* Malaysian, *PHI* Philippine, *JWU* Jewish-Uzbekistan, *AFR* African

A phylogenetic tree for the 14 populations was constructed from the matrix of pairwise D_A distances by NJ method (Fig. 2-5). The tree revealed that there were four monophyletic clusters, which consisted of Non-Asian, North Asian, Japanese, and the Han Chinese/Southeast Asian populations, respectively. Two Han Chinese populations showed closer relationships with Southeast Asian populations than with Japanese populations. The long branch leading to JP-Ainu may be due to long-term isolation with small effective population size.

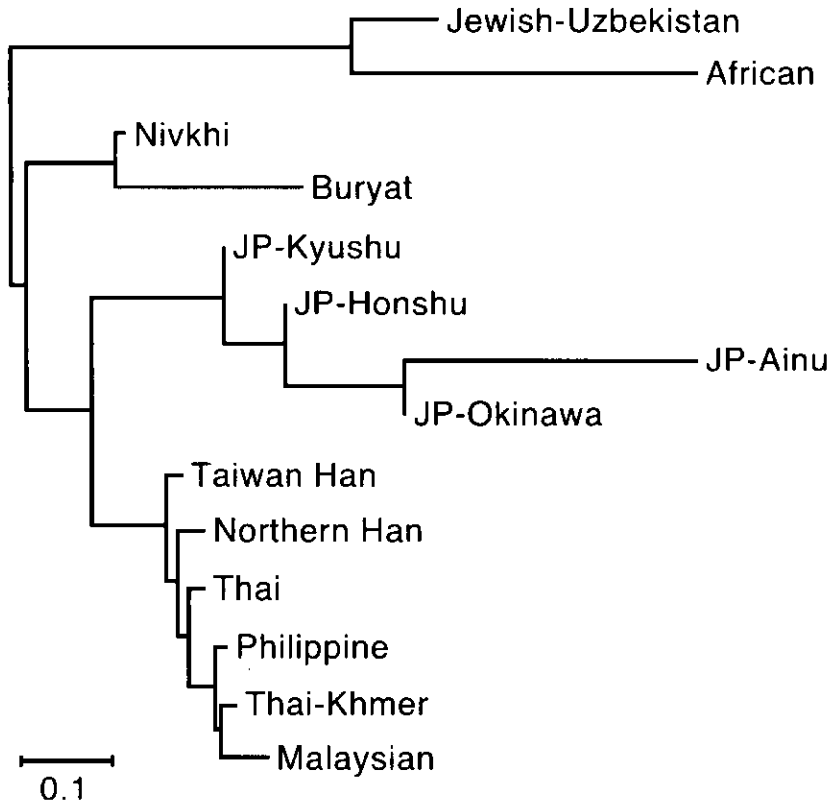


Figure 2-5. Neighbor-joining (NJ) tree, showing the relationships of the 14 populations on the basis of D_A distances. The scale for the distance is shown *bottom left*.

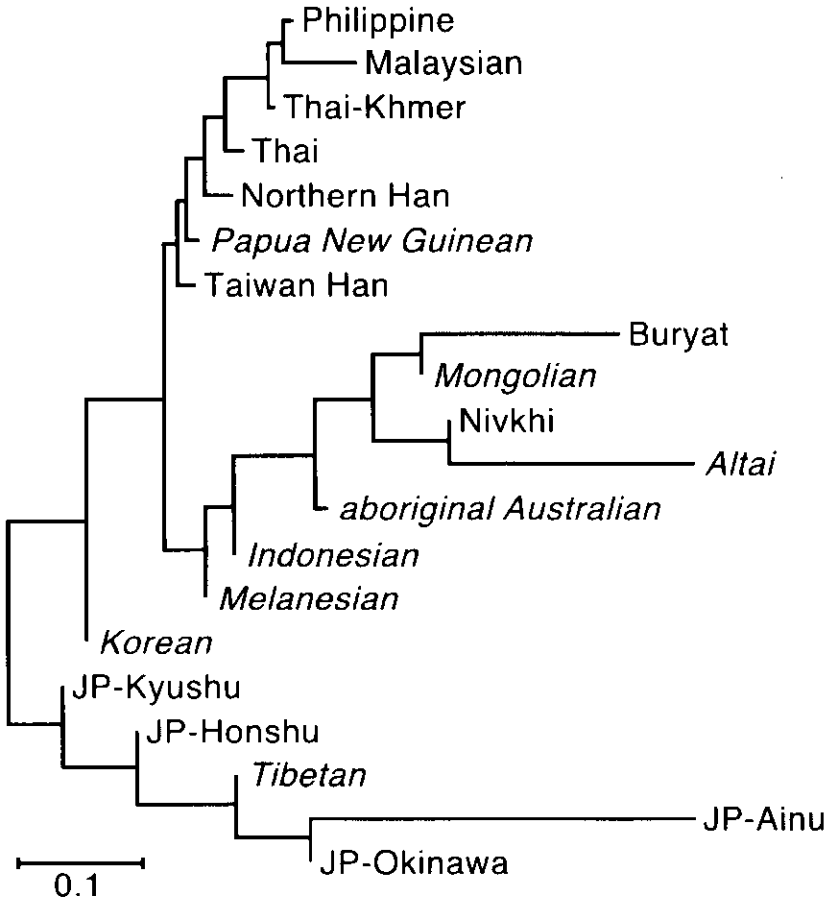


Figure 2-6. Neighbor-joining (NJ) tree for a total of 20 Asian and Australo-Melanesian populations. This tree is based on D_A distances, calculated from the frequencies of seven Y-haplotypes with six polymorphic sites ($DYS257_{108}$, *DYS287*, SRY_{4064}, SRY_{10831}, $RPS4Y_{711}$, and M9). The data for eight populations given in *italics* are from Karafet et al. (1999). The scale for the distance is shown *bottom left*.

2.2.5 Implications for the Peopling of East and Southeast Asia

The three Asian clusters (North Asian, Japanese, and Han Chinese/Southeast Asian clusters; Fig. 2-5) appear to correspond to the contrasting distributions of three haplotypes, viz., ht7, ht1, and ht4, respectively. These may reflect three waves of human migration with different paternal ancestries into East and Southeast Asia. To determine more fully the distributions of the Y-haplotypes, a phylogenetic tree for a total of 20 Asian and Australo-Melanesian populations was constructed, combining the present data with those from Karafet

et al. (1999). Three distinct clusters have been also found in Asian and Australo-Melanesian populations (Fig. 2-6). These clustering patterns are somewhat different from those in the phylogenetic analyses undertaken with other genetic markers, where the geographically close populations in Asia and Australo-Melanesia appear to form a cluster (e.g., Nei and Roychoudhury 1993).

Two North Asian populations (Buryat and Nivkhi) are closely related to Indonesian and Australo-Melanesian populations (aboriginal Australian and Melanesian, but not Papua New Guinean), as well as Central Asian populations (Mongolian and Altai). The observation that ht7 is frequent in the aboriginal Australians (Karafet et al. 1999) supports this unique clustering. The ht7 haplotype is specific to Asian populations and the populations derived from ancestors in Asia, such as aboriginal Australians and Native Americans (Bergen et al. 1999; Karafet et al. 1999; Kayser et al. 2001). On the basis of the dating of fossil remains, the human colonization of Sahul (the single landmass combining Australia and New Guinea during the latest glacial period) started more than 50,000-60,000 years ago (Thorne et al. 1999). The mutational age for $RPS4Y_{711}$ (Table 2-2) is comparable to the early fossil records in Australia, although this mean estimate is somewhat older than those from other studies (Bergen et al. 1999; Karafet et al. 1999; Kayser et al. 2001). These observations suggest that the T allele at $RPS4Y_{711}$ arose in Asia before the separation of modern Asian populations and others of the Asian descent (e.g., aboriginal Australians). These may also reflect an ancient genetic connection between the geographically distant populations, who shared some extent of a common ancestral population in carrying ht7.

The Japanese cluster includes the Tibetan population in the phylogenetic tree for the 20 populations (Fig. 2-6). This coincides with results showing that Tibetan populations are

also marked with high frequencies of YAP$^+$ chromosomes (Hammer et al. 1997; Qian et al. 2000), and with recent findings regarding mtDNA polymorphisms (Qian et al. 2001). In the tree, the Korean population is related to the Japanese/Tibetan group, although YAP$^+$ chromosomes are rare in Koreans (Kim et al. 1998; Karafet et al. 1999).

As described in Chapter 1, there have been at least two major migrations that brought modern humans from the Asian Continent to the Japanese archipelago. The Neolithic people of Japan (the so-called Jomon people) arrived in Japan more than 12,000 years ago, although the exact timing and geographic origin of this migration are still controversial (Turner 1990; Nei 1995). Subsequently, the post-Neolithic migrants (the so-called Yayoi people) came to Japan about 2,300 years ago (for problems regarding the formation of modern Japanese populations, see Chapter 4).

As mentioned in Section 2.2.2, ht1 is predominant in both the Ainu and Okinawans (Fig. 2-3). Furthermore, the Ainu and Okinawans are neighbors in the phylogenetic trees (Figs. 2-5 and 2-6), although the genetic distance between them is longer than those between Okinawans and two mainland Japanese populations (JP-Honshu and JP-Kyushu; Table 2-3). There is some agreement that these two groups represent modern descendants of the Jomon people (Hanihara 1991; Horai et al. 1996; Omoto and Saitou 1997). Thus, the present observations support the hypothesis that the YAP$^+$ chromosome is a marker of Jomon male lineage (Hammer and Horai 1995). Moreover, these suggest that modern humans carrying YAP$^+$ chromosomes migrated into East/Southeast Asia in the late Pleistocene, with some immigrants settling in Japan over 12,000 years ago.

The Papua New Guinean population clusters together with the Han Chinese/Southeast Asian group (Fig. 2-6) because ht4 is frequent in the Papua New Guineans

(Karafet et al. 1999). The phylogeographic analyses of Y-haplotypes have suggested the range expansion of several derivatives of ht4 in East Asia, Southeast Asia and Oceania during the Holocene (Su et al. 2000a, 2000b; Capelli et al. 2001; Kayser et al. 2001; Underhill et al. 2001). Of note is that ht4 is not found in the Ainu (Fig. 2-3), supporting the above inference of historical demography. Also, the relatively short distances between the Han Chinese and Southeast Asians (Figs. 2-5 and 2-6) suggest that they had been subdivided in more recent times of the evolutionary history, although D_A distance is not always linearly related to evolutionary time (Nei et al. 1983). Alternatively, these may imply a relatively recent genetic admixture among them, followed by expansions of their range.

As described in Chapter 1, the Ainu and aboriginal Australians are considered to be relatively pure descendant of people who had settled in the respective area (the Japanese archipelago or Australia) in the late Pleistocene (Turner 1990; Hanihara 1991). If this is the case, both ht1 and ht7 may be regarded as ancient Asian haplotypes, which are mainly distributed in the peripheral area of the Asia-Pacific regions. Thus, the geographic distributions and the age estimates of three Asian Y-chromosome lineages may represent at least three distinct waves of male dispersal into prehistoric East and Southeast Asia, generally consistent with the hypothesis of Underhill et al. (2001). This paternal view more readily explains the close affinities between the populations who are now living far apart (Fig. 2-6), rather than the assumption of a single immigration carrying all the Asian Y-lineages with subsequent genetic drift.

2.2.6 Further Considerations

It should be always cautious to equate genetic lineages with specific migration events. In particular, lack of recombination on the NRY indicates that Y-haplotype is a single locus, which has only paternal history of modern humans. Moreover, it also means that Y-haplotype is susceptible to positive or negative selection as a unit (Brookfield et al. 2000; Jobling and Tyler-Smith 2000), affecting its geographic distribution in human populations. Nevertheless, the phylogeographic analysis in this chapter carries two significant points.

First, the study clarifies a subset of human populations for discussion on the peopling of East and Southeast Asia. Using gene frequency data for many polymorphic loci, phylogenetic analyses of modern worldwide populations have shown the presence of the Asia-Pacific cluster which consists of human populations in East Asia, Southeast Asia, Oceania, and Australo-Melanesia (Nei and Roychoudhury 1993; Omoto 1995) although the statistical support is not so strong. However, most genetic studies on human dispersal in East and Southeast Asia do not have treated the Asia-Pacific regions as a whole. In other words, many researches have separately focused on the genetic relationships (1) between East Asians and Southeast Asians (e.g., Su et al. 1999; Ding et al. 2000), and (2) between Southeast Asians and Pacific Islanders including Australo-Melanesians (e.g., Su et al. 2000a; Capelli et al. 2001). This geographic subdivision may have led to an incomplete understanding of the early migration events to East and Southeast Asia. This chapter presents the possibility that a single hypothesis explains the entire demographic histories of human populations inhabiting the Asia-Pacific regions (a part of North and Central Asia, East and Southeast Asia, and Australo-Melanesia; Fig. 2-6). Therefore, future analyses will be done in this context.

Second, the study offers an important viewpoint regarding the geographic patterns of

human DNA variation. The three clusters in the phylogenetic tree for the 20 populations (Fig. 2-6) imply that there are the *central-peripheral* genetic differences in the geographic distributions of the Y-haplotypes in the Asia-Pacific regions. This interpretation is not so compelling because there has been much evidence showing several kinds of genetic distinction among the Asian-Pacific populations: for example, (1) between continental East Asians (*central*) and the Ainu (*peripheral*; Horai et al. 1996), or (2) between Southeast Asians (*central*) and aboriginal Australians (*peripheral*; Redd et al. 1999). More importantly, the robust Y-haplotype phylogeny (Fig. 2-2) allows to deduce the ancient genetic relationships among the *peripheral* populations, and among the *central* populations in the Asia-Pacific regions. Therefore, further phylogeographic analyses with comparable validity are necessary to verify this paternal view.

Recently, Ingman et al. (2000) have reported that complete mtDNA sequences (excluding the D-loop region) are useful for constructing a robust phylogeny rooted in Africa. They have also revealed that non-Africans (including Asian-Pacific populations) can be clearly divided into two monophyletic clusters, suggesting that this strategy is effective in deducing ancient genetic relationships among the Asian-Pacific populations with a statistical support. Moreover, 1.4 million SNPs in the human genome (The International SNP Map Working Group 2001) may help to identify new haplotype(s) for testing the present hypothesis. At any rate, a combined analysis with several haplotype systems is required.

2.3 Summary

DNA variation on the NRY was examined in 610 male samples from 14 global populations in North, East and Southeast Asia, and other regions of the world. In combination with the previous findings on the geographic distributions of the Y-haplotypes in Asia and Australo-Melanesia, this chapter inferred the peopling of East and Southeast Asia from the phylogeographic point of view.

1. Eight haplotypes were observed by analyses of seven biallelic polymorphic markers (*DYS257*$_{108}$, *DYS287*, *SRY*$_{4064}$, *SRY*$_{10831}$, *RPS4Y*$_{711}$, M9, and M15). The haplotypes were unevenly distributed among the populations. The maximum parsimony tree for the eight haplotypes had no ambiguity, indicating a robust phylogeny which is necessary for a phylogeographic analysis. Moreover, the haplotypes could be classified into four distinct lineages characterized by three key mutations: an insertion of the Y *Alu* polymorphic (YAP) element at *DYS287*, a C-to-G transversion at M9, and a C-to-T transition at *RPS4Y*$_{711}$. Coalescence analysis in the haplotype tree showed that estimated ages for three key mutations ranged from 53,000 to 95,000 years ago, suggesting that the three major lineages (defined by the allele of YAP$^+$, M9-G, and *RPS4Y*$_{711}$-T, respectively) were separated from one another during early stages of human evolutionary history.

2. The three haplotype lineages (designated as YAP$^+$, M9-G, and RPS4Y-T) accounted for 98.6% of the Asian populations studied, indicating that these three paternal lineages have contributed to the formation of modern Asian populations. Phylogenetic analysis revealed three monophyletic Asian clusters, which consisted of North Asian, Japanese, and Han Chinese/Southeast Asian populations, respectively. These three clusters appear to

correspond to geographic distributions of the three lineages: RPS4Y-T, YAP$^+$, and M9-G, respectively. The extensive phylogenetic analysis in 20 Asian and Australo-Melanesian populations also showed the three distinct clusters, which may be correlated with the geographic differences in the haplotype distributions.

3. The distribution patterns of the haplotypes and mutational ages for the key markers suggest that three major groups with different paternal ancestries separately migrated to prehistoric East and Southeast Asia. This hypothesis explains the entire evolutionary histories of human populations inhabiting the Asia-Pacific regions (a part of North and Central Asia, East and Southeast Asia, and Australo-Melanesia). Also, this is amenable to verifying by further phylogeographic studies with other genetic systems.

CHAPTER 3

3 DNA sequence variation in a 12.6-kilobase region on human Y chromosome

As described in Chapter 1, it is necessary to find novel population-specific (or region-specific) polymorphisms because they are most informative to deduce past human migrations such as colonization of a specific region. For this purpose, a 12.6-kb region in the NRY is selected, and sequenced in eight individuals from East and Southeast Asia, Africa, and Europe. Comparisons between the obtained sequences allow to discover variable nucleotide sites (such as SNPs) within the selected DNA region. To clarify the specificity of the newly discovered SNPs, they are genotyped on another DNA samples from worldwide populations with PCR-RFLP methods.

3.1 Materials and Methods

3.1.1 DNA Samples

In Chapter 2, a total of eight Y-haplotypes were observed. To find shared polymorphisms among the haplotypes, and/or polymorphisms specific to a haplotype, a biased sampling strategy was used on the basis of the haplotype affiliations for individuals. The affiliations and the geographic locations of eight males examined are summarized in Table 3-1.

3.1.2 Selection of a Region for SNPs Survey

To select a non-recombining region of the Y chromosome for SNPs survey, the following criteria were employed: the region has (1) a Y-chromosomal sequence-tagged site (STS; Vollrath et al. 1992), (2) no known gene, and (3) no homology to the X chromosome. Searching the electronic database for the human genome sequence (Human Genome Sequencing; http://www.ncbi.nlm.nih.gov/genome/seq/) helped to select a 12.6-kb region (covering positions 44,740 to 57,340 of the contig AC009491.3) in Yp11.2. This region contains *DYS257*, one of the polymorphic STS (see Chapter 2). The current annotation suggests that it is about 20 kb apart from *TTY11* (testis transcript Y11) locus in the direction to the telomere of the short arm.

3.1.3 PCR and DNA Sequencing in SNPs Survey

The 12.6-kb region was amplified in 12 overlapping fragments by PCR method. The nucleotide sequences of the PCR and sequencing primers are shown in Appendix C. Each 25-50 µl of PCR mixture contained 50-100 ng genomic DNA, 0.2-0.4 µM each primer, 0.2 mM each dNTPs, 1 × QIAGEN PCR Buffer (Tris•HCl, pH 8.7; KCl; $(NH_4)_2SO_4$; 1.5 mM $MgCl_2$) with or without Q-Solution, and 0.5-1.0 U QIAGEN *Taq* DNA polymerase (QIAGEN, Germany). The cycling conditions were 94°C for 2 min, and then 30-35 cycles of 94°C for 30 s, 54-62°C for 30 s and 72°C for 30-90 s in GeneAmp PCR System 9600 (Applied Biosystems, USA) (The detailed conditions for PCR are summarized in Appendix D). Female DNA sample was used as a negative control. The PCR products were determined semi-quantitatively by electrophoresis on 1% agarose gel (Bio-Rad Laboratories, USA) in 0.5 × TBE buffer, followed by staining with ethidium bromide.

Direct sequencing method was used to determine the nucleotide sequences with the sequencing primers (same as the PCR primers; see Appendix C). The PCR products were purified with MicroSpin S-400 HR Columns (Amersham Pharmacia Biotech Inc., USA). The purified products were used as templates for sequencing. Sequencing reactions were performed with ABI PRISM BigDye Terminator Cycle Sequencing FS Ready Reaction Kit, or Dye Terminator Cycle Sequencing FS Ready Reaction Kit (Applied Biosystems, USA), according to the manufacturer's instructions. The sequencing products were purified with AutoSeq G-50 or MicroSpin S-200 HR Columns (Amersham Pharmacia Biotech Inc., USA), and were run on 5% Long Ranger gel (BME, USA) in 1 × TBE buffer in ABI PRISM 377 DNA Sequencer (Applied Biosystems, USA).

The nucleotide sequences were automatically obtained with lane tracking and base calling system in ABI PRISM DNA sequence analysis software (ver. 3.0). ABI PRISM AutoAssembler software (ver. 1.4.0) was used to proofread the sequences, and assemble both strands of the sequences for each individual into a consensus sequence. To identify variable site(s) in every overlapping fragment, the consensus sequences for all the individuals were aligned with GENETYX-MAC/MAlign software (ver. 1.0.0). In some experiments for verifying the observed variable sites and obtaining relatively high-quality sequences, internal primer(s) within each fragment was designed to resequence the PCR product (nucleotide sequences of the primers are shown in Appendix C).

3.1.4 Genotyping of New Polymorphic Markers

For further verification, the novel SNPs (named A46812 and A57316) were tested on additional population samples which were also used in Chapter 2. To characterize the A46812

polymorphism, European DNA samples (n=4) were also analyzed. The two polymorphisms were genotyped with PCR-RFLP methods. Each 25-μl PCR mixture contained 20-50 ng genomic DNA, 0.2 μM each primer, 0.2 mM each dNTPs, 1 × GeneAmp PCR Buffer II, 0.5 U Ampli*Taq* DNA polymerase (Applied Biosystems, USA), and 1.5 mM (for A46812) or 3.0 mM (for A57316) of $MgCl_2$.

A 245-bp fragment containing the A46812 polymorphism was amplified with the following primer set: 5'-GGGGCTCTCTGCCAAATTAT-3' and 5'-GGTTCATGGTCTCGTTTTGG-3'. The cycling condition for A46812 was 94°C for 2 min, and then 30 cycles of 94°C for 30 s, 59°C for 30 s and 72°C for 30 s. The PCR products were digested with *Hpy*CH4IV (recognition sequence of this enzyme is 5'-A^{v}[C]G$_{A}$T-3', in which the bordered letter indicates a mutated site for this polymorphism). The digested DNA fragments were separated by electrophoresis on 2% agarose gel (Bio-Rad Laboratories, USA) in 0.5 × TBE buffer, and detected by staining with ethidium bromide for typing the allele.

A 183-bp fragment encompassing A57316 was amplified with the following primer set: 5'-GAGTTGGAAGCACTTTCTGTGTCATCAGTTTTCTCTGACT-3' and 5'-ATTCCTCCATAGTCTTGTGA-3' (the underlined bases in one primer indicate non-complementary nucleotides which generate *Hin*fI recognition sequence; referred to as *mismatch primer*). The recognition sequence for *Hin*fI is 5'-G^{v}A N T$_{A}$[C]-3', where the letter with border represents a polymorphic site for A57316. The cycling condition for A57316 was 94°C for 2 min, and then 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 15 s. The *Hin*fI-digested fragments were separated by electrophoresis on 3% NuSieve 3:1 agarose gel (BME, USA) in 1 × TBE buffer, followed by staining with ethidium bromide for genotyping.

3.1.5 Statistical Analyses of Sequence Data

A total of nine sequences including the available reference sequence (GenBank accession no. AC009491.3) were used to measure the extent of DNA variation in the selected region. The Watterson's θ (Watterson 1975) and its standard error were estimated using Equation 12.52 and 12.53, respectively, in Nei and Kumar (2001). Nucleotide diversity (π) and its standard error (calculated by the bootstrap method with 500 replications) were computed with *MEGA* version 2.1 (Kumar et al. 2001). Tajima's neutrality test (Tajima 1989) for DNA polymorphism was also performed with *MEGA2.1*. The critical values for statistical significance of the test were obtained from Tajima (1989).

3.2 Results

3.2.1 Characterization of Sequence Variations

For eight males, 12,601 nucleotide sites of the selected region (positions 44,740-57,340 of the reference sequence) were sequenced. Among the nine sequences (including the available reference), two SNPs at positions 46,812 and 57,316 were observed, together with one known polymorphism at $DYS257_{108}$ (position 45,469; Table 3-1). All three SNPs had only two alternative nucleotides at each of the polymorphic sites, and were transitions. One length polymorphism of microsatellite DNA (consisting of repeats of CA dinucleotides) was found at positions ranging from 44,934 to 44,961. The number of the repeat lengths for each of the individuals is shown in Table 3-1. There were another two length differences in relation to homopolymeric runs of Ts within the sequenced region. These were not treated as the variations because the present sequencing strategy has usually difficulty in counting the exact number of the stretches for any nucleotide.

Table 3-1. DNA sequence variations in the 12.6-kb region on the NRY for eight individuals

Sample No.	Haplotype affiliation	Geographic location	DNA sequence variation[a] 45469	46812	57316	Microsatellite $(CA)_n$ 44934-44961	Remarks
AC009491			A	T	T	14	
1	ht1	Japan	G	T	C	14	MS23
2	ht1	Japan	G	T	C	14	OK5
3	ht1	Japan	G	T	C	14	HS7
4	ht2	Thai	G	T	C	14	Th143
5	ht3	Africa	G	T	T	14	SB4
6	ht4	Japan	G	C	T	14	MS10
7	ht5	Europe	A	T	T	14	SB22
8	ht7	Japan	G	T	T	13	MS72

[a] The numbering refers to the genomic DNA sequence (accession No. AC009491.3). The polymorphic site at position 45,469 corresponds to the known polymorphism at $DYS257_{108}$. For microsatellite DNA polymorphism, the number of repeat lengths for each individual is shown.

Considerations of the parsimonious relationships among the eight Y-haplotypes led to place the two new SNPs (named A46812 and A57316) on the haplotype tree (Fig. 3-1; see Chapter 2). In other words, A57316 was considered to be a SNP shared by ht1 and ht2; A46812 be specific to ht4. From the root position of the tree, T allele was inferred to be ancestor at each site. Incidentally, the reference sequence (AC009491) could belong to either ht5 or ht6.

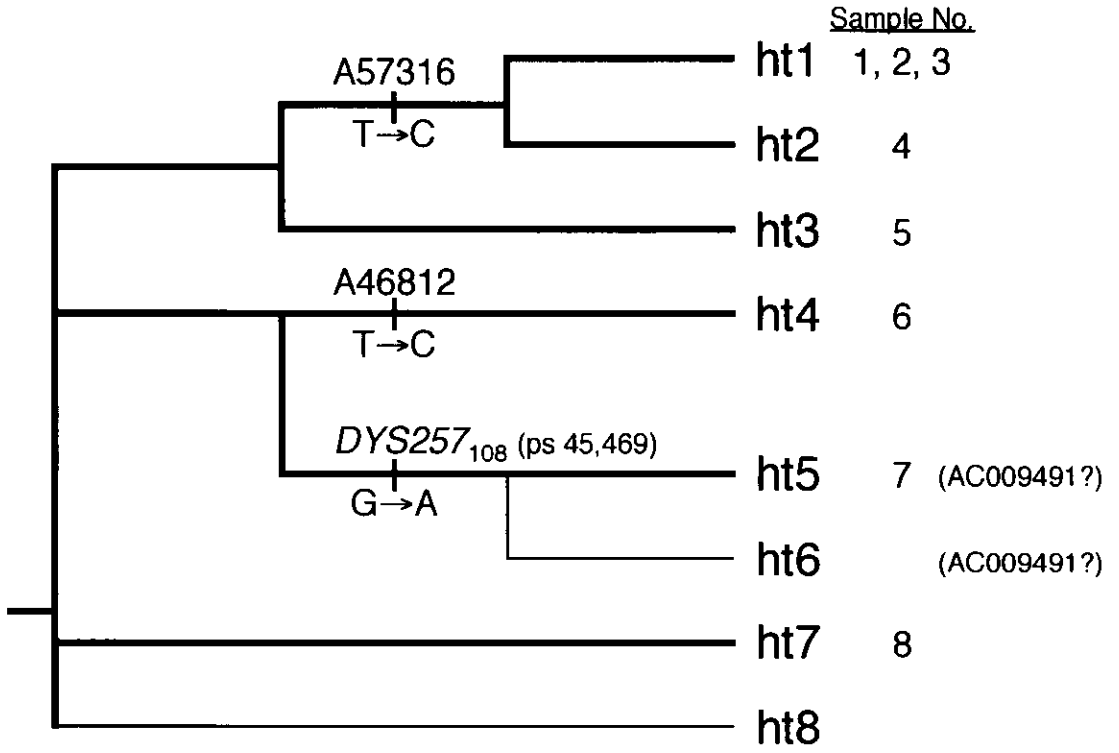


Figure 3-1. Evolutionary inferences of ancestral sequences at the polymorphic sites. The mutational change (*vertical bar* across the branch) and ancestral nucleotide state at each of the polymorphic sites were inferred from the parsimonious relationships among the eight Y-haplotypes (see Chapter 2).

Two estimates of nucleotide variation were calculated from a total of 12,597-bp comparisons excluding four alignment gaps in sequence. The nucleotide diversity (π) and Watterson's θ ($\pm$ standard error) among the nine sequences were estimated to be 0.0093 $\pm$ 0.0051% and 0.0088 $\pm$ 0.0064%, respectively. This π value is comparable to those for several noncoding regions on the NRY (Shen et al. 2000). This is also consistent with the overall estimate for the NRY (0.0151% on using 2,304,916-bp sequence; The International SNP Map Working Group 2001). The Tajima's *D* value for testing the neutrality was 0.22, which was not significantly different from zero (P>0.1). This suggests neutral mutation in the sequenced region under mutation-drift equilibrium.

3.2.2 Characterization of Two Novel SNPs

The A46812 polymorphism was subsequently genotyped with PCR-RFLP method (Fig. 3-2A) in a set of 610 male individuals, which was the same as that in Chapter 2. Only the T allele occurred in the individuals who were affiliated with all the Y-haplotypes except ht4 (n=268; see Fig. 3-1). For the group of people carrying ht4, almost all males (97.7%; 334/342 individuals) showed the C allele, but the rest (8/342) had the T allele, which was distributed in Malaysia, the Philippines, and the Uzbekistan (for detailed geographic distribution of each allele, see Chapter 4). The observations support the aforementioned evolutionary inferences for the A46812 polymorphism (Fig. 3-1). Thus, A46812 is considered to be a real SNP which is effective in subdividing the ht4 haplotype.

When this SNP was also examined in four European males (for their haplotype affiliations, *ht4*, n=2; *ht5*, n=2), they displayed only the T allele. This suggests that the spatial distribution of the C allele (the mutated allele) is restricted in eastern Eurasia, although further analysis on worldwide populations should be required.

For the A57316 polymorphism, a subset of 117 males belonging to the YAP^+ lineage (ht1, ht2, and ht3) were examined by PCR-RFLP analysis with mismatch primer (Fig. 3-2B). The C allele was observed in all of the Asian YAP^+ lineage (ht1 and ht2; n=100), while others outside East and Southeast Asia (ht3; n=17) showed the T allele. The result agrees with the above prediction that A57316 must be shared by ht1 and ht2. Furthermore, the C allele (the mutated allele) may be specific to East and Southeast Asian populations because ht1 and ht2 are distributed in these restricted regions (see Chapter 2; Hammer and Horai 1995; Hammer et al. 2001).

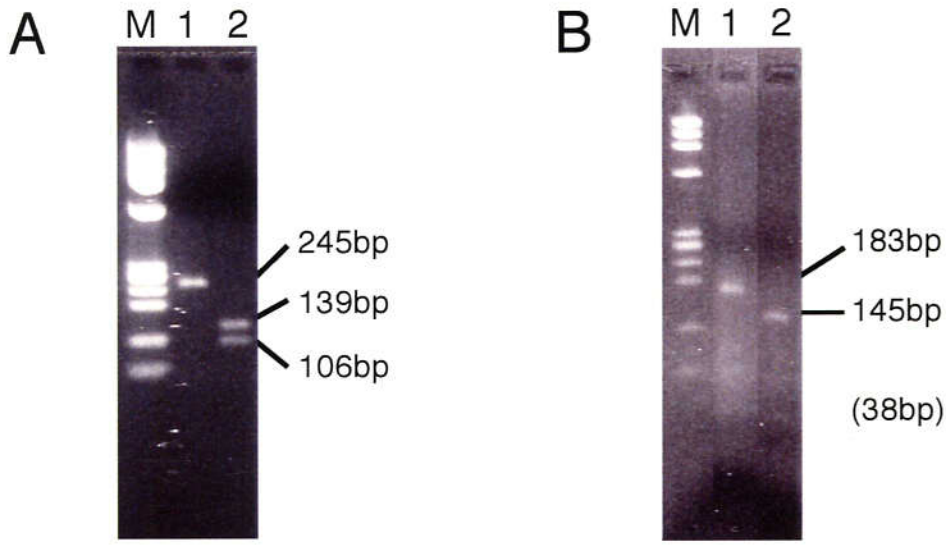


Figure 3-2. Electrophoretic analyses for typing the alleles at each of two novel SNPs. (A) The *Hpy*CH4IV cleavage patterns of the PCR products for A46812 on a 2% agarose gel. *Lane M* DNA size markers; *lane 1* T allele (245 bp); *lane 2* C allele (139 and 106 bp). (B) The *Hin*fI cleavage patterns of the PCR products for A57316 on a 3% NuSieve 3:1 agarose gel. *Lane M* DNA size markers; *lane 1* T allele (183 bp); *lane 2* C allele (145 and 38 bp). The 38-bp fragment for the C allele is usually invisible.

3.3 Discussion

The systematic comparative sequencing in the eight individuals helped to find the two phylogeographically informative SNPs: one (A46812) was polymorphic within ht4, and the other (A57316) was shared by ht1 and ht2. These eight individuals were sampled on the basis of their haplotype affiliations in Chapter 2. As a result of this sampling strategy, the evolutionary characteristics for the new SNPs could be inferred from the parsimonious relationships among the haplotypes (Fig. 3-1). Moreover, the inferences were concordant with the observations of genotyping other worldwide samples, indicating high accuracy of the evolutionary inferences. This is the most important feature in this SNPs survey. In general, it remains unclear whether a newly discovered polymorphism is useful for a phylogeographic analysis until it is examined on several regional populations. Therefore, this sampling strategy

with ascertainment bias may be efficient in finding an informative polymorphic marker.

On the other hand, this strategy is less effective in measuring the extent of DNA polymorphism (such as nucleotide diversity) because the obtained sequences are not randomly sampled. However, the estimated nucleotide diversity (π) is generally consistent with the previous estimates for Y-chromosome DNA polymorphisms from random samples of worldwide populations (Underhill et al. 1996; Bergen et al. 1999; Shen et al. 2000). This indicates that sufficient number of nucleotide differences can be observed in these samples if the published estimates for π are regarded as expected values from DNA polymorphisms in the NRY.

This thesis only focuses on biallelic polymorphisms (such as SNPs) in the NRY to construct a stable Y-haplotype phylogeny for a phylogeographic analysis. This explains why there is no characterization of the discovered polymorphism for microsatellite DNA, which is expected to be multiallelic state owing to its relatively higher mutation rate. Recently, multiallelic Y-chromosome markers with higher mutation rate have been used to measure internal genetic diversity for particular Y-haplotype defined by biallelic markers, leading to estimate time-depth of the haplotype (Jobling and Tyler-Smith 2000). Therefore, future investigation of the observed microsatellite DNA polymorphism will be useful for such combined analysis with the different types of polymorphic markers.

3.4 Summary

To find population-specific (or region-specific) polymorphism, DNA variation in a 12.6-kb region on the NRY was examined in eight individuals from East and Southeast Asia, Africa, and Europe. Moreover, the newly discovered SNPs were analyzed on additional DNA samples from worldwide populations with PCR-RFLP methods.

1. Two novel SNPs (named A46812 and A57316) were found by the comparative sequencing. One length polymorphism of microsatellite DNA (consisting of repeats of CA dinucleotides) was also observed.

2. The A46812 polymorphism was effective in subdividing the ht4 haplotype, while A57316 was shared by ht1 and ht2. The results indicate that both are phylogeographically informative SNPs. The geographic distributions of the mutated alleles (both C alleles inferred) at the two polymorphic sites revealed that these two SNPs may be specific to eastern Eurasian populations.

CHAPTER 4

4 Tracing male-mediated migrations into Japan : Evidence from the Y-chromosome haplotypes

As summarized in Chapter 1, there are at least three major problems in relation to historical formation of modern Japanese populations: they are on (1) the extent of genetic contribution of the Yayoi immigrants to the gene pool of modern Japanese (reviewed by Hanihara 1991), (2) the geographic origin of the Jomon immigrants (Suzuki 1969; Turner 1990; Nei 1995; Omoto and Saitou 1997), and (3) the extent of genetic differences in the Ainu and Okinawan populations (Horai et al. 1996; Dodo et al. 1998; Hatta et al. 1999). All the problems may come down to a question as to how to interpret regional differences within the Japanese archipelago. Therefore, a more complete understanding of human dispersal in East and Southeast Asia must lead to a more precise interpretation on the regional variations in Japan.

In this chapter, geographic distributions of Y-haplotypes in the 14 modern human populations are minutely investigated with 15 polymorphic sites (including the nine biallelic polymorphisms examined in Chapter 2 and 3). The findings promote a better understanding of the evolutionary histories of modern Asian populations, and are also used to test the proposed hypothesis (see Chapter 2) regarding human migrations into East and Southeast Asia. Finally, this chapter discusses the above problems on the formation of modern Japanese populations.

4.1 Materials and Methods

4.1.1 Subjects and DNA Samples

The DNA samples were the same as a set of 610 males in Chapter 2.

4.1.2 Genotyping of Biallelic Polymorphisms

In addition to the nine biallelic polymorphisms (the seven known polymorphisms and the two novel SNPs) which have been already analyzed in Chapter 2 and 3, the following six SNPs on the NRY were genotyped: M8 (G→T), M55 (T→C), M119 (A→C), M122 (T→C), M125 (T→C) and M217 (A→C; Underhill et al. 1996, 2000). Because of limited amounts of genomic DNA for some samples, the six biallelic markers were examined according to a hierarchical approach on the basis of the evolutionary relationships among Y-haplotypes (Underhill et al. 2000; Hammer et al. 2001). From the correspondence of the Y-haplotype phylogeny (Fig. 2-2 in Chapter 2) to that reported in Underhill et al. (2001), a subset of the samples affiliated with each of the three Y-lineages (see Chapter 2) was genotyped only on the following respective markers: for YAP$^-$ lineage, M55 and M125; for M9-G lineage, M119 and M122; for RPS4Y-T lineage, M8 and M217.

All the six polymorphisms were analyzed with PCR-RFLP methods (detailed experimental conditions are summarized in Appendix E). For several polymorphic sites, a mismatch primer with non-complementary nucleotide(s) was used to create a recognition site for an appropriate restriction endonuclease. DNA fragments encompassing each of the polymorphic sites were amplified with the following cycling conditions in GeneAmp PCR System 9600 (Applied Biosystems, USA): 94°C for 2 min, and then 30-40 cycles of 94°C for

30 s, 54-60°C for 30 s and 72°C for 15-30 s. Each 25-µl PCR mixture contained 20-50 ng genomic DNA, 0.2 µM each primer, 0.2 mM each dNTPs, 1.5 mM $MgCl_2$, 1 × GeneAmp PCR Buffer II, and 0.5 U Ampli*Taq* DNA polymerase (Applied Biosystems, USA).

The PCR products for each polymorphism were digested with the appropriate restriction endonuclease (see Appendix E). For typing the allele, the digested fragments were separated by electrophoresis on 2% agarose gel (Bio-Rad Laboratories, USA) in 0.5 × TBE buffer, or 3% NuSieve 3:1 agarose gel (BME, USA) in 1 × TBE buffer, followed by staining with ethidium bromide (typical electrophoretic detection of the allele variation for each of the six polymorphic sites is shown in Appendix F).

4.1.3 Data Analyses

A total of the 15 polymorphisms were used to infer Y-chromosome haplotypes. The phylogenetic and coalescent analyses of Y-haplotypes were performed according to the methods described in Chapter 2 (Section 2.1.3 to 2.1.5). To measure the extent of genetic variation in population, the number of observed haplotypes (k), haplotype diversity (h: equivalent to heterozygosity for diploid data; Nei 1987), and the mean number of pairwise sequence differences among haplotypes (p) were calculated with the ARLEQUIN software ver.2.000 (Schneider et al. 2000; for ARLEQUIN, http://lgb.unige.ch/arlequin/). The haplotype diversity h is defined as

$$h = \frac{n}{n-1}\left(1 - \sum_{i=1}^{q} x_i^2\right)$$

where n is the number of chromosomes sampled, q is the number of haplotypes, and x_i is the frequency of the i-th haplotype in population.

4.2 Results and Discussion

4.2.1 New Phylogenetic Tree for Y-Chromosome Haplotypes

Figure 4-1 shows the evolutionary relationships among 14 haplotypes defined by analyzing the 15 biallelic polymorphic sites on the NRY (for definition and designation of the haplotypes, see legend to Fig. 4-1). The haplotype tree displayed no ambiguous state although it contained the two newly discovered SNPs (see Chapter 3). This indicates a robustness of Y-haplotype analysis on the NRY.

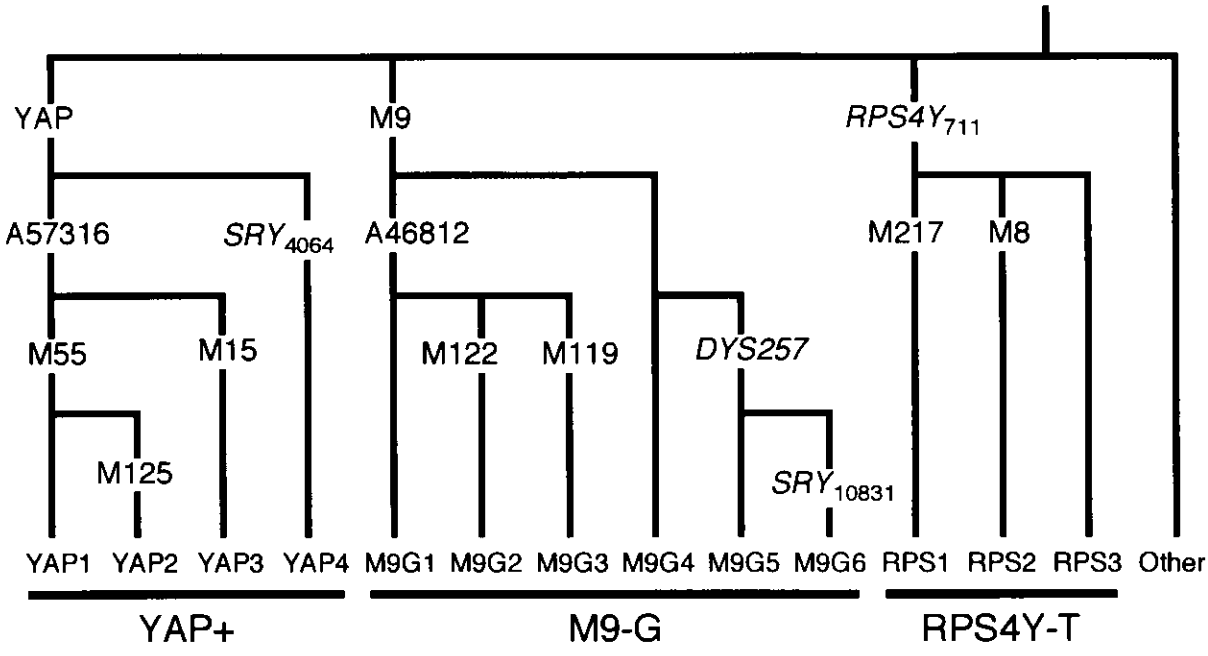


Figure 4-1. Maximum parsimony phylogeny for 14 NRY haplotypes. A total of 14 haplotypes are defined by allelic associations among 15 polymorphic sites, and are referred to as YAP1-YAP4, M9G1-M9G6, RPS1-RPS3, and Other. Each *name of the biallelic marker* across the branch represents an expected mutational event at one of the 15 polymorphic sites. The correspondence of the designation of the previous haplotypes (see Chapter 2) to the newly defined one are as follows: *ht1* YAP1 and YAP2; *ht2* YAP3; *ht3* YAP4; *ht4* M9G1-M9G4; *ht5* M9G5; *ht6* M9G6; *ht7* RPS1-RPS3; *ht8* Other. Three major lineages are indicated *bottom* with the lineage names: YAP^+, M9-G, and RPS4Y-T.

In this chapter, the hierarchical genotyping was used to minimize the DNA samples and the experimental time needed (Underhill et al. 2000; Hammer et al. 2001). Ideally, all polymorphisms should be examined on all samples because some multiple mutations (such as recurrent mutations) possibly remain undetected using this hierarchical approach. As mentioned in Chapter 2, however, no available multiple markers make ambiguity at the Y-haplotype level. In addition, multiple mutation rate at the same nucleotide site in the NRY may not be so high (Underhill et al. 2000). Therefore, the present approach for genotyping would have no serious effect on the reliability of the Y-haplotype phylogeny.

4.2.2 Frequency Distributions of Y-Chromosome Haplotypes

Haplotype frequencies in the 14 populations are shown in Table 4-1. The minute analysis with the 14 haplotypes revealed several geographically restricted distributions of the haplotypes among Asian populations, in addition to those of the YAP$^+$ lineage (see Chapter 2). For the RPS4Y-T lineage, RPS1 (defined by $RPS4Y_{711}$-T/M217-C) was distributed among North and East Asian populations except for JP-Okinawa, while RPS2 (defined by $RPS4Y_{711}$-T/M8-T) was only observed in three Japanese populations (JP-Honshu, Kyushu, and Okinawa). Moreover, RPS3 (defined only by $RPS4Y_{711}$-T) was found in only two males, one each from Thais and Thai-Khmers.

Table 4-1. Frequency distributions of Y-chromosome haplotypes in 14 populations

Population	*N*	Haplotype frequency (%)[b]													
		YAP1[a]	YAP2	YAP3	YAP4	M9G1	M9G2	M9G3	M9G4	M9G5	M9G6	RPS1	RPS2	RPS3	Other
North Asian															
Buryat	61	0.0	0.0	0.0	0.0	9.8	0.0	0.0	0.0	0.0	0.0	83.6	0.0	0.0	6.6
Nivkhi	21	0.0	0.0	0.0	0.0	28.6	0.0	0.0	0.0	19.0	9.5	38.1	0.0	0.0	4.8
East Asian															
JP-Ainu[c]	16	81.3	6.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.5	0.0	0.0	0.0
JP-Honshu[c]	82	19.5	17.1	0.0	0.0	37.8	19.5	0.0	0.0	0.0	0.0	1.2	4.9	0.0	0.0
JP-Kyushu[c]	104	16.3	11.5	0.0	0.0	33.7	24.0	1.9	0.0	0.0	0.0	7.7	3.8	0.0	1.0
JP-Okinawa[c]	45	35.6	20.0	0.0	0.0	22.2	15.6	0.0	0.0	2.2	0.0	0.0	4.4	0.0	0.0
Northern Han	49	0.0	0.0	0.0	0.0	22.4	55.1	6.1	0.0	8.2	0.0	8.2	0.0	0.0	0.0
Taiwan Han	21	0.0	0.0	0.0	0.0	9.5	61.9	9.5	0.0	0.0	4.8	14.3	0.0	0.0	0.0
Southeast Asian															
Thai	34	0.0	0.0	2.9	0.0	44.1	35.3	8.8	0.0	2.9	0.0	0.0	0.0	2.9	2.9
Thai-Khmer	65	0.0	0.0	1.5	0.0	53.8	27.7	3.1	0.0	1.5	10.8	0.0	0.0	1.5	0.0
Malaysian	12	0.0	0.0	0.0	0.0	25.0	25.0	25.0	8.3	0.0	8.3	0.0	0.0	0.0	8.3
Philippine	50	0.0	0.0	0.0	0.0	2.0	38.0	46.0	10.0	2.0	2.0	0.0	0.0	0.0	0.0
Total (Asia)	560	11.1	6.4	0.4	0.0	27.7	25.0	6.8	1.1	2.1	2.1	13.8	1.8	0.4	1.4
Other															
Jewish-Uzbekistan	46	0.0	0.0	0.0	28.3	0.0	2.2	0.0	4.3	2.2	2.2	0.0	0.0	0.0	60.9
African	4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Overall	610	10.2	5.9	0.3	2.8	25.4	23.1	6.2	1.3	2.1	2.1	12.6	1.6	0.3	5.9

[a] The haplotypes are defined in Figure 4-1.

[b] Some rows do not sum to 100% because the frequencies (%) were rounded to one decimal place.

[c] *JP* Japanese

The A46812 polymorphism divided the M9-G lineage into two sub-lineages: M9G1-M9G3, and M9G4-M9G6 (Figure 4-1). For the sub-lineage defined by A46812-C allele, two North Asian populations (Buryat and Nivkhi) exhibited only M9G1 (defined by M9-G/A46812-C). The M9G2 haplotype (defined by M9-G/A46812-C/M122-C) was widely distributed among East and Southeast Asian, and Jewish populations. The highest frequency of M9G2 was found in Taiwan Han (61.9%), followed by Northern Han (55.1%). This supports that the haplotype carrying M122-C allele on a M9-G background has been associated with the range expansion from China (such as the expansions of agriculture and the Sino-Tibetan-speaking people) during the Holocene (Su et al. 2000b; Underhill et al. 2001). The M9G3 haplotype (defined by M9-G/A46812-C/M119-C) was found in Southeast Asian and some East Asian populations. Among them, Philippines and Malaysians were characterized by higher frequencies of 46.0 and 25.0 %, respectively. For the other sub-lineage on a M9-G background, only six Asian individuals (1.1%) in the Philippines and Malaysia belong to M9G4 (defined only by M9-G).

As described in Chapter 3, the two novel SNPs (A46812 and A57316) may be regarded as specific markers for the Asian populations examined. Furthermore, the RPS4Y-T lineage is specific to Asian populations and others of the Asian descent (e.g., aboriginal Australians and Native Americans; see Chapter 2). Therefore, the phylogenetic relationships among the 14 Y-haplotypes (Fig. 4-1) suggest that the above geographic variations among the Asian populations have been mostly caused by evolutionary dynamics (such as mutation, genetic drift, and isolation by distance) after human settlements in East and Southeast Asia.

4.2.3 Coalescent Estimation on Gene Genealogy of Y-Chromosome Haplotypes

Because the data set with the 14 haplotypes had no violation of assumptions regarding stationary (Fig. 4-2) and panmictic populations, the coalescence analysis of the haplotype tree was performed with the GENETREE program as described in Chapter 2. Table 4-2 shows the converted ages in years with stochastic errors, on using the estimated $M_{\text{MIL}} = 2.0$ for the 14 haplotypes. The results are in general agreement with the previous analysis in Chapter 2 (see Table 2-2), which used a subset of these polymorphic sites and $M_{\text{MIL}} = 1.0$. The TMRCA is slightly more recent than the previous estimate of 183,000 ± 70,000 years. In general, a larger M (a higher mutation rate *per haplotype*) leads to a more recent TMRCA in this approach (Griffiths and Tavaré 1994; Hammer et al. 1998). Indeed, the extensive Y-haplotype analysis with large M has shown considerably recent TMRCA (on the order of 50,000 years; Thomson et al. 2000). This is probably because their analysis is carried out under violation of the theoretical assumptions needed. Thus, this estimation method should be used only when neutral variation can be assumed under mutation-drift equilibrium. Within the limits, it may be effective in inferring an evolutionary history of a genealogical tree.

The mean estimates for the two new SNPs were 39,000 years ago for A46812, and 49,000 years ago for A57316, respectively (Table 4-2). Recent studies have revealed that modern humans arrived in the Asia-Pacific regions more than 50,000-60,000 years ago (Quintana-Murci et al. 1999; Thorne et al. 1999; Adcock et al. 2001). Therefore, the estimated mutational ages suggest that the two SNPs (A46812 and A57316) arose during relatively early stages of the Asian history. Also, these estimates are consistent with the spatial distributions of these polymorphisms, which may be specific to eastern Eurasian populations (see Chapter 3).

Table 4-2. Estimation on TMRCA, and the ages of mutations

	Coalescence Time[a] (in years; mean ± SD)
TMRCA	157,000 ± 53,000
Age of mutation	
<YAP$^+$ lineage>	
DYS287 YAP insertion	92,000 ± 42,000
A57316 T→C	49,000 ± 25,000
M55 T→C	25,000 ± 12,000
M125 T→C	8,800 ± 5,700
M15 9-bp insertion	17,000 ± 12,000
SRY_{4064} G→A	35,000 ± 29,000
<M9-G lineage>	
M9 C→G	85,000 ± 42,000
A46812 T→C	39,000 ± 18,000
M122 T→C	18,000 ± 9,400
M119 A→C	9,900 ± 6,600
$DYS257_{108}$ G→A	22,000 ± 15,000
SRY_{10831} G→A	6,400 ± 5,900
<RPS4Y-T lineage>	
$RPS4Y_{711}$ C→T	58,000 ± 32,000
M217 A→C	19,000 ± 11,000
M8 G→T	11,000 ± 9,100

[a] Maximum likelihood estimate of M (M_{ML} = 2.0) was used to estimate the ages in the Y-haplotype tree with the GENETREE program. Several assumptions for this estimation are summarized in Section 2.1.5.

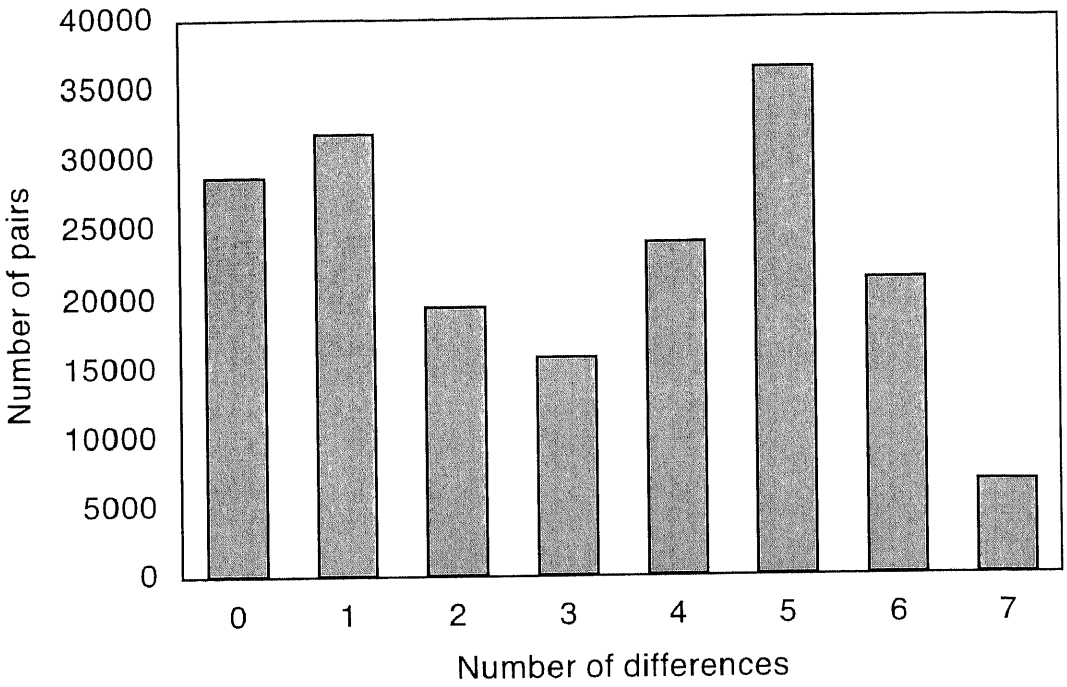


Figure 4-2. Frequency distribution of sequence differences for Y-chromosome haplotypes in the 14 populations. The pairwise differences were calculated in 610 male individuals for the 14 Y-haplotypes with 15 segregating sites.

4.2.4 Y-Chromosome Haplotype Diversity

Among Asian populations, the number of regional haplotypes (k) ranged from three in Buryat and JP-Ainu to eight in JP-Kyushu (Table 4-3), although the sample size for each population was different. The haplotype diversity (h) and the mean number of pairwise differences (p) are also shown in Table 4-3. In combination of the two statistics, three different patterns were observed: (1) low h (less than 0.4)/low p (less than 2), (2) high h/low p, and (3) high h/high p.

Table 4-3. Comparison of Y-chromosome haplotype diversity in 14 populations

Population	k^a	h^b (mean ± SD)	p^c (mean ± SD)
Buryat (N=61)	3	0.292 ± 0.071	0.92 ± 0.64
Nivkhi (N=21)	5	0.762 ± 0.055	2.54 ± 1.42
JP-Ainu[d] (N=16)	3	0.342 ± 0.140	1.29 ± 0.85
JP-Honshu[d] (N=82)	6	0.759 ± 0.024	3.24 ± 1.69
JP-Kyushu[d] (N=104)	8	0.789 ± 0.021	3.23 ± 1.68
JP-Okinawa[d] (N=45)	6	0.775 ± 0.031	3.30 ± 1.73
Northern Han (N=49)	5	0.642 ± 0.060	1.51 ± 0.93
Taiwan Han (N=21)	5	0.605 ± 0.111	1.96 ± 1.16
Thai (N=34)	7	0.690 ± 0.053	1.31 ± 0.84
Thai-Khmer (N=65)	7	0.630 ± 0.047	1.33 ± 0.84
Malaysian (N=12)	6	0.864 ± 0.064	1.73 ± 1.08
Philippine (N=50)	6	0.646 ± 0.040	1.35 ± 0.85
Jewish-Uzbekistan (N=46)	6	0.558 ± 0.062	1.24 ± 0.80
African (N=4)	1	0	0

[a] Number of the observed haplotypes

[b] Haplotype diversity (equivalent to heterozygosity; Nei 1987)

[c] Mean number of pairwise differences among haplotypes

[d] *JP* Japanese

Buryats and the Ainu exhibited the first pattern of low h and low p values. In the respective populations, the low h value is due to the presence of a single predominant haplotype with its frequency of more than 80% (Table 4-1). This pattern suggests a prolonged isolation with small effective population size (Avise 2000). The second pattern of high h and low p was found in two Han Chinese and Southeast Asians, who were marked with high frequencies of the M9-G lineage as described in Chapter 2. This second pattern generally reflects rapid population growth from an ancestral population with small effective population size (Grant and Bowen 1998; Avise 2000), implying a relatively recent range expansion of the

M9-G lineage in East and Southeast Asia.

On the other hand, three Japanese populations (JP-Honshu, Kyushu, and Okinawa) and Nivkhi displayed third pattern of high *h* and high *p*. In general, high values for *h* and *p* are an expected sign for a stable population with large effective population size. Alternatively, they could be also observed in an admixed group of people from historically distinct populations (Grant and Bowen 1998; Avise 2000). Thus, the high values for three Japanese populations may signify that modern Japanese have been shaped by genetic admixture of several populations with different ancestries.

4.2.5 Y-Chromosome Polymorphisms in Asia: Implications for the Peopling of Japan

The detailed haplotype analysis of this chapter reveals that most East and Southeast Asian populations can be explained by the following three haplotype lineages: YAP^+/A57316-C, M9-G/A46812-C, and RPS4Y-T (Table 4-1 and Figure 4-1). Moreover, several variations in geographic distributions of the haplotypes belonging to these three lineages may have been generated after early settlement of the human in these regions as discussed above. Thus, these confirm the usefulness of the three lineages for understanding paternal history of modern humans in East and Southeast Asia.

Chapter 2 hypothesized that there were at least three waves of male dispersal into prehistoric East and Southeast Asia, mainly from phylogeographic analysis of the three haplotypes: ht1 (YAP^+), ht4 (M9-G), and ht7 (RPS4Y-T). In this paternal view, it is important to examine carefully whether early dispersal events from Africa brought these three haplotypes to East and Southeast Asia, because a more recent gene flow (e.g., European gene flow in modern history) into the gene pool of modern Asian populations always leads to an

erroneous inference. As described above, the findings on spatial distributions and coalescent estimation for the three lineages (YAP$^+$/A57316-C, M9-G/A46812-C, and RPS4Y-T) are suggestive of early arrival of populations carrying the three haplotypes in East and Southeast Asia. Therefore, these do not prevent us from considering the formation of modern Japanese populations under the proposed hypothesis for human dispersal into East and Southeast Asia.

On the basis of geographic distributions of the Y-haplotypes, there were several differences in genetic components between the Ainu and Okinawan populations (Table 4-1). For the YAP$^+$ lineage, the Ainu and Okinawan populations shared two haplotypes (YAP1 and YAP2), which were also observed in the mainland Japanese populations (JP-Honshu and Kyushu; Table 4-1). Nevertheless, there were differences in the frequencies for each of the two haplotypes and the combined haplotype (YAP1 + YAP2) among the Japanese populations. As discussed in Chapter 2, YAP$^+$ chromosome is considered as a marker of the Jomon male lineage (Hammer and Horai 1995). Thus, all the modern Japanese populations may share some extent of the Jomon ancestry. Furthermore, the Ainu and Okinawans still retain the Jomon element of male lineages because of the higher frequencies for the YAP$^+$ lineage.

For the M9-G lineage, two haplotypes (M9G1 and M9G2) were shared by Okinawans and the mainland Japanese, while the Ainu exhibited no such haplotypes (Table 4-1). Of note is that M9G2 is found in Okinawan populations because the haplotype carrying M122-C allele (correspond to M9G2) is in association with the range expansion from China (such as agricultural expansion) during the Holocene as mentioned above (Su et al. 2000b; Underhill et al. 2001). This suggests a relatively recent gene flow to Okinawans and the mainland Japanese after population divergence between the Ainu and Okinawans if they had a common ancestor (Dodo et al. 1998; Hatta et al. 1999). There is some agreement that the

Yayoi immigrants came from continental Northeast Asia (probably from northeastern China) ~2,300 years ago (Hanihara 1991). Thus, M9G2 may be one of the genetic elements of the Yayoi immigrants. Further Y-haplotype analysis is necessary to trace its migration route to the Okinawa Islands (e.g., directly from China or from the mainland Japan). At any rate, the findings in this chapter lend support to the hybridization theory regarding the formation of modern Japanese populations if the presence of M9G2 in modern Japanese populations reflect a Yayoi admixture.

For the RPS4Y-T lineage, the Ainu and Okinawans had a different haplotype, viz., RPS1 and RPS2, respectively (Table 4-1). As shown in Figure 4-1, the evolutionary relationship between the two haplotypes is parallel, but not hierarchical (Underhill et al. 2000). In other words, the two haplotypes (RPS1 and RPS2) arose independently on a RPS4Y-T background because of the acquisition of separate mutation (M217 and M8, respectively). As discussed in Chapter 2, range expansion of the RPS4Y-T lineage in the Asia-Pacific regions may have occurred during a relatively ancient time (Underhill et al. 2001). This suggests that the ancestral Japanese in the late Pleistocene (or the Jomon era) had already displayed the regional differences in their paternal lineages. This idea coincides with recent observations of maternal lineages from mtDNA polymorphisms in modern Japanese populations (Horai et al. 1996). Also, estimated mutational ages for M217 (19,000 years ago) and M8 (11,000 years ago; Table 4-2) may support such inference from the phylogeographic point of view.

In sum, the Y-haplotype data from this phylogeographic analysis favor the hybridization theory on the formation of modern Japanese populations, consistent with the inference from other genetic studies with mtDNA haplotypes (Horai et al. 1996) and classic genetic markers (Omoto and Saitou 1997). The data also suggest that the ancient Japanese

populations in the late Pleistocene (or the Jomon era) had already diversified as shown by the presence of two paternal lineages (YAP$^+$ and RPS4Y-T lineages), and regional differences (RPS1 and RPS2). Furthermore, there might be gene flow (M9G2) to Okinawan and mainland Japanese populations during the Holocene after the separation of Okinawans and the Ainu.

This chapter could not clarify geographic origin(s) of the Jomon immigrants. As mentioned in Chapter 1, Southeast Asia (Suzuki 1969; Turner 1990) or Northeast Asia (Nei 1995) is presumed to be an original place of the Jomon immigrants. To elucidate their geographic origin, most genetic studies have made a comparison between modern Japanese and contemporary populations in Southeast Asia and/or Northeast Asia (Nei 1995; Bannai et al. 1996, 2000; Omoto and Saitou 1997). Notably, the studies have implicitly assumed temporal genetic continuity between modern groups of people inhabiting Southeast Asia or Northeast Asia and the respective indigenous inhabitants in the late Pleistocene. The present phylogeographic analysis with Y-haplotypes reveals the past dynamics of human dispersal in East and Southeast Asia (Underhill et al. 2001; Karafet et al. 2001). This suggests that the assumption for the temporal continuity is easily violated. For validating the continuity, it is possible to examine variation in DNAs extracted from ancient human fossils in Southeast and Northeast Asia. This direct analysis of the ancient DNA will also help to attain a deeper understanding of human dispersal in the Asia-Pacific regions.

4.3 Summary

Geographic distributions of Y-haplotypes in the 14 modern human populations were analyzed with a total of 15 polymorphic markers, including the nine biallelic polymorphisms examined in Chapter 2 and 3. This extensive analysis allowed to verify the 'three waves of human migration' hypothesis, which was proposed in Chapter 2 for understanding the entire evolutionary history of human populations inhabiting the Asia-Pacific regions. Furthermore, this chapter discussed the formation of modern Japanese populations from the paternal point of view.

1. Allelic associations among the 15 polymorphic sites in the NRY defined 14 haplotypes. The parsimonious relationships among the 14 haplotypes showed no ambiguity, ensuring high reliability of phylogeographic analysis with Y-haplotypes.

2. In addition to geographically restricted distributions of the YAP$^+$ lineage in East and Southeast Asia, the distributions of several haplotypes tended to be regionally specific. Spatial distributions and estimated ages for three polymorphisms (A57316, A46812, and $RPS4Y_{711}$) suggest that many lines of the regional specificity have mostly resulted from evolutionary forces (such as mutation and isolation by distance) after human colonization in East and Southeast Asia. Also, these demonstrate the effectiveness of phylogeographic analysis with a set of the 14 haplotypes on deducing the evolutionary history of the Asian populations.

3. There were several regional differences in paternal genetic characteristics of modern Japanese populations. In particular, the differences between the Ainu and Okinawan populations helped to infer the formation process of modern Japanese. Ancient Japanese

populations in the late Pleistocene (or Jomon era) may have diversified as evidenced by two paternal lineages (YAP^+ and RPS4Y-T), and regional differentiation within the RPS4Y-T lineage. After population divergence between the Ainu and Okinawans, there might be gene flow to Okinawans and the mainland Japanese during the Holocene, implying the so-called Yayoi migrations to the Japanese archipelago.

4. The Y-haplotype data are consistent with the hybridization theory in relation to the formation of modern Japanese populations. The geographic origin(s) of the Jomon immigrants remains unclear. To address this question, it may be necessary to examine sequence variation in the ancient DNAs from Southeast and Northeast Asian populations in the late Pleistocene.

CHAPTER 5

5 Summary and Concluding Remarks

It is relatively easy to detect geographic variation in the human genome, while there is little knowledge about an evolutionary history of geographic variants. It has been the aim of this thesis to examine the spatial distribution of genetic variation at the haplotype level, and infer valid evolutionary relationships among haplotypes. From the phylogeographic point of view, the findings have been associated with past human migration.

Chapter 1 introduced the phylogeographically useful properties of the human Y-chromosome haplotypes in the NRY. Also, the chapter pointed out the two questions involved in human migrations: (1) human settlements in East and Southeast Asia, and (2) the formation of modern Japanese populations. Although many studies have already investigated these questions, the situations are still confusing. For clearing up the problems, it is necessary (1) to propose a verifiable hypothesis on the peopling of East and Southeast Asia, and (2) to explain the evolutionary history of modern Japanese in a chain of range expansions of the Asian populations, respectively. For these purposes, this thesis has examined geographic variations of Y-haplotypes in 610 male samples from 14 global populations in North, East and Southeast Asia, and other regions of the world. Moreover, the thesis has attempted to find population-specific (or region-specific) polymorphisms for further phylogeographic analysis with Y-haplotypes.

Chapter 2 discussed early human migrations into East and Southeast Asia. First, combined analysis of seven known biallelic polymorphisms ($DYS257_{108}$, *DYS287*, SRY_{4064}, SRY_{10831}, $RPS4Y_{711}$, M9, and M15) defines eight Y-haplotypes, among which the evolutionary relationships are reliable. The haplotypes are subdivided into four separate haplotype lineages characterized by three key mutations. Coalescence analysis with the haplotypes suggests that the three major lineages (defined by the allele of YAP^+, M9-G, and $RPS4Y_{711}$-T, respectively) were separated from one another during early stages of human evolutionary history.

Second, the three paternal lineages (named YAP^-, M9-G, and RPS4Y-T) mainly constitute modern Asian populations. Phylogenetic analysis displays three monophyletic Asian clusters, which consist of North Asian, Japanese, and Han Chinese/Southeast Asian populations, respectively. These clustering patterns may correspond to geographic variation in the haplotype distributions. Similar observations are shown in the phylogenetic tree for 20 Asian and Australo-Melanesian populations, who are analyzed with the combined findings from this thesis and the literature.

Third, this chapter provides an important insight that human populations inhabiting the Asia-Pacific regions (a part of North and Central Asia, East and Southeast Asia, and Australo-Melanesia) may share common genetic backgrounds. Also, it hypothesizes that three major groups with different paternal ancestries separately migrated to prehistoric East and Southeast Asia. Further phylogeographic studies with another haplotype systems are necessary to test this hypothesis.

Chapter 3 took up the subject to find new SNPs for further phylogeographic analysis with Y-haplotypes. For this purpose, DNA variation in a 12.6-kb region on the NRY is examined by comparative sequencing in eight males from East and Southeast Asia, Africa,

and Europe. This examination allows to discover two new SNPs (named A46812 and A57316). Genotyping the two SNPs on additional worldwide samples reveals that the A46812 polymorphism is effective in subdividing the ht4 haplotype (defined only by M9-G), and that A57316 is shared by ht1 (defined only by YAP^+) and ht2 (defined by $YAP^+/M15^+$). Moreover, the two polymorphisms may be specific to eastern Eurasian populations. Further studies on more extensive populations are required to confirm the characterization for each of the two SNPs.

Chapter 4 mainly discussed the formation of modern Japanese populations from the paternal point of view. First, this chapter obtains robust Y-phylogeny for the observed 14 haplotypes which consist of 15 linked polymorphisms in the NRY. Also, phylogeographic analysis with the 14 haplotypes reveals that the distributions of several haplotypes tend to be regionally specific. Taking account of estimated ages for three polymorphisms (A57316, A46812, and $RPS4Y_{711}$), the regional specificity may have been mostly due to evolutionary forces (such as mutation and isolation by distance) subsequent to human settlements in East and Southeast Asia. This indicates that the present analysis is appropriate for deducing the evolutionary history of the Asian populations.

Second, several regional differences in paternal genetic characteristics are observed among modern Japanese populations. From the phylogeographic point of view in the Asia-Pacific regions, the Y-haplotype data suggest that the ancient Japanese populations in the late Pleistocene (or Jomon era) had diversified owing to two paternal lineages (YAP^+ and RPS4Y-T), and regional differentiation within the RPS4Y-T lineage. Okinawan and the mainland Japanese populations might experience some extent of gene flow during the Holocene, implying Yayoi migrations to the Japanese archipelago. Therefore, the data bolster the

hybridization theory that modern Japanese have been caused by distinct genetic contributions involving ancient Jomon people and the Yayoi immigrants. To resolve the problem concerned with the geographic origin(s) of Jomon immigrants, it is meaningful to investigate sequence variation in the NRY in the ancient DNAs from Southeast and Northeast Asian populations in the late Pleistocene. This will also leas to a better understanding of human dispersal in the Asia-Pacific regions.

In conclusion, a phylogeographic analysis with Y-chromosome haplotypes has a great possibility to illuminate past human dispersion. This is mainly because the evolutionary relationships among geographic variants can be inferred without ambiguity. In general, evolutionary history of a single gene (such as Y-haplotype) should be cautiously interpreted as population history. Therefore, the present phylogeographic views need to be further integrated with much knowledge from other genetic systems, and many another disciplines such as paleoclimatology and archaeology. Nevertheless, the significance of the thesis is to present several insights in relation to the evolutionary history of modern human populations in the Asia-Pacific regions.

APPENDICES

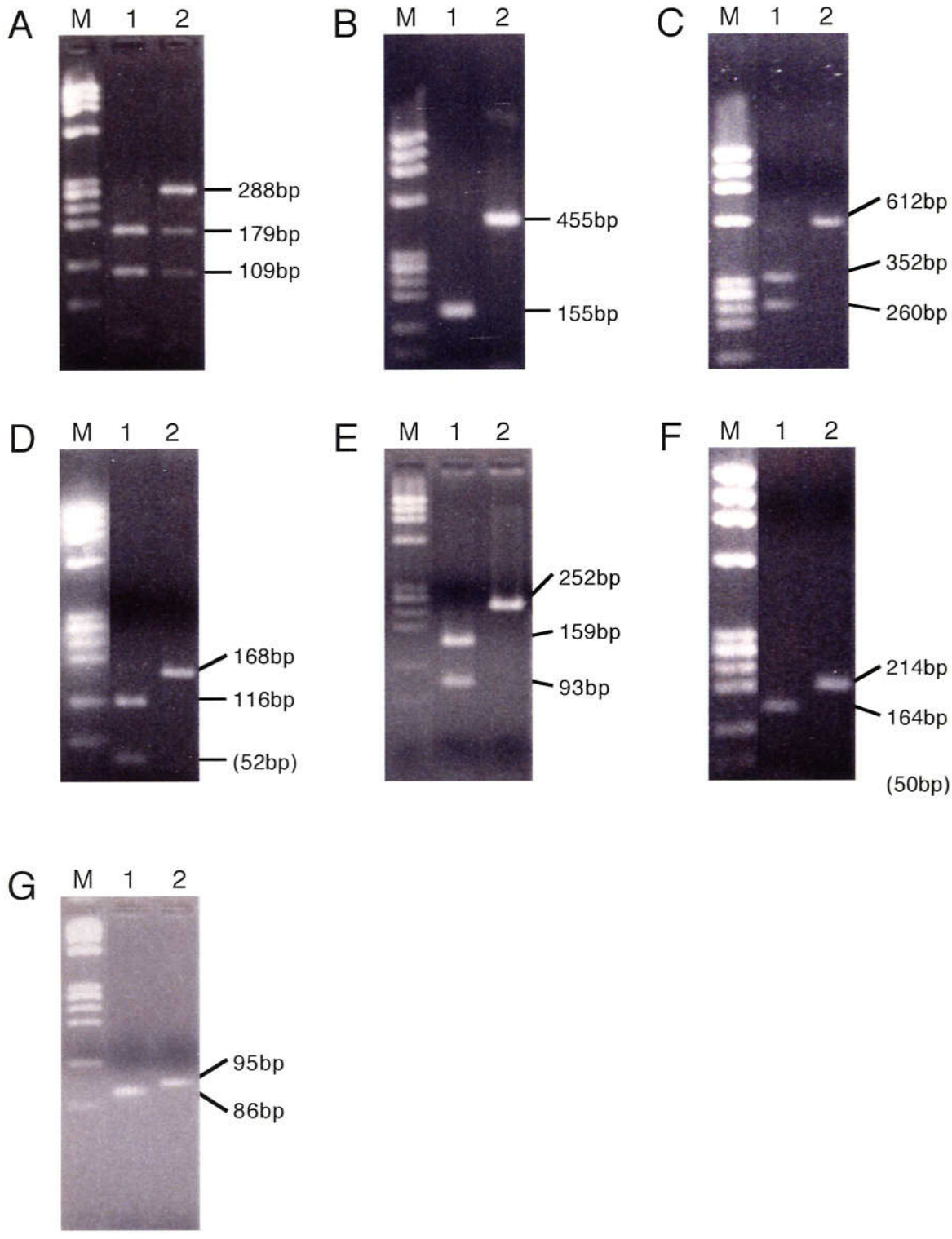
Appendix A. PCR conditions and typing systems of alleles for the seven biallelic polymorphisms

Markers	PCR conditions				Detection Method[b]	Typing of Alleles
	Left and right primers[a]	Concentration		Annealing Temperature (°C)		*upper*: ancestral *lower*: mutated
		$MgCl_2$ (mM)	Each primer (μM)			
$DYS257_{108}$ G→A	5'-GAACTTGTCGGGAGGCAAT-3' 5'-TGATACACTTCCTCCTTTAGTGG-3'	2.5	0.2	60	RFLP (*Ban*I) G^{v}GPyPuC$_{A}$C	G: 179, 109 bp A: 288, 179, 109 bp[c]
DYS287 *Alu* insertion	5'-CAGGGGAAGATAAAGAAATA-3' 5'-ACTGCTAAAAGGGGATGGAT-3'	2.5	0.12	55	Separation on 2% agarose	–: 155 bp +: 455 bp
SRY_{4064} G→A	5'-GCATTTTGTTACCCTTCTCAAC-3' 5'-TGGCAAGACTTACGAGATTTC-3'	1.5	0.2	57	RFLP (*Bsr*BI) GAG^{v}CGG	G: 352, 260 bp A: 612 bp
SRY_{10831} A↔G	5'-AAAATAGCAAAAAATGACACAAGGC-3' 5'-TCCTTAGCAACCATTAATCTGG-3'	3.0	0.12	59	RFLP (*Mae*III) ^{v}GTNAC$_{A}$	G: 116, 52 bp A: 168 bp
$RPS4Y_{711}$ C→T	5'-GATTTTGTTGGGTGGTGGTC-3' 5'-TGCTGCTACTGCAATTTAGCC-3'	1.5	0.2	57	RFLP (*Bsl*I) CCNN$_{A}$NNN^{v}NNGG	C: 159, 93 bp T: 252 bp
M9 C→G	5'-TCAGGACCCTGAAATACAGAACT-3' 5'-TTGAAGCTCGTGAAACAGATTAG-3'	1.5	0.2	57	RFLP (*Hin*fI) G^{v}ANT$_{A}$C	C: 164, 50 bp G: 214 bp
M15 9-bp insertion	5'-CCTCATGCGCATATACAATCA-3' 5'-CCACTGCACCTAGGGAGACA-3'	1.5	0.2	60	Separation on 4% NuSieve 3:1 agarose	–: 86 bp +: 95 bp

[a] Sources for primers are as follows: $DYS257_{108}$, Vollrath et al. (1992); *DYS287*, Hammer and Horai (1995); SRY_{4064}, Qamar et al. (1999); SRY_{10831}, Hammer et al. (1998); and M9, Thomas et al. (1999).

[b] Restriction endonucleases for PCR-RFLP analysis are represented in parenthesis. Recognition sequence for each enzyme is shown in *lower line*. The bordered letter in each sequence represents the respective polymorphic site

[c] There is evidence for two copies of the *DYS257* in the NRY. The G-to-A transition causes the loss of a *Ban*I recognition site in one of the two copies for the *DYS257*, resulting in either a two-fragment pattern (G allele) or a three-fragment pattern (A allele).



Appendix B. Electrophoretic detection of the allele variation for each of the seven polymorphic sites. (A) $DYS257_{108}$, (B) *DYS287*, (C) SRY_{4064}, (D) SRY_{10831}, (E) $RPS4Y_{711}$, (F) M9, and (G) M15. *Lane M* DNA size markers; *lane 1* ancestral allele; *lane 2* mutated allele. The details of the analyses are summarized in Appendix A and Section 2.1.2.

Appendix C. Nucleotide sequences for PCR and sequencing primers

No. of fragments	PCR and sequencing primers		Internal primers for sequencing	
1	44740 5'-GGACACCAGCGAGACTTTGT-3'	44759	45288 5'-CCATGGATCCCACAGAAAAG-3'	45307
	45938 5'-CCAGTCTGAGGGGTGAGAAC-3'	45919	45429 5'-GCACCATGGTTGATTGTGAG-3'	45410
2	45896 5'-AGGCTAGTCAAATGTGGAGCA-3'	45916	46433 5'-AGGAGACCACTGTGGAGGTG-3'	46452
	47090 5'-GCATGACCTTCTCATTGTGG-3'	47071	46603 5'-AATCCTGCGAACATTTTTGG-3'	46584
3	47044 5'-CAGCAGTGGCTCAATGAAAG-3'	47063	47610 5'-CATTTGCTTGCAGGTGAAGA-3'	47629
	48232 5'-CAGAAGCAGCTGGACTGTGT-3'	48213	47720 5'-GGGATTTCATCCTGGGAGAT-3'	47701
4	48163 5'-TGCCACACACAAAAGCAGAT-3'	48182	48935 5'-ACACTCACAACCCTCCCACT-3'	48954
	49381 5'-TGAGCTAGCCTCAGAACATCAG-3'	49360		
5	49354 5'-TCAAGCCTGATGTTCTGAGG-3'	49373	50029 5'-ACTGGGGTGCCAGACTATCA-3'	50048
	50524 5'-TGACTGATCCTCACCTGCAC-3'	50505		
6	50476 5'-TTCTTGCACCTTGGACCTTC-3'	50495	51093 5'-GAAACATTCATGCAACGGATT-3	51113
	51680 5'-CAAACCAGGGGGAGACACTA-3'	51661		
7	51602 5'-CCTACTCAGCAGGGAACCAA-3'	51621	52205 5'-TCCTCAGGGATTTCATGGTT-3'	52224
	52840 5'-TGGCCATCATTCATTGACAC-3'	52821	52281 5'-AAGGTGAGGGCAGACATGAG-3'	52262
8	52777 5'-GGGTTCATAGGGTGATGGAA-3'	52796		
	53963 5'-AGGGATACCCTTCCAAGGTC-3'	53944		
9	53861 5'-CATCCTGCCCTCATCAGATT-3'	53880	54324 5'-CACATTCCTTTGGGCAGACT-3'	54343
	55092 5'-GTTTCCATCGACCAGCTCAT-3'	55073	54502 5'-TTCCCTGAGGTTGCTCTCTC-3'	54483
10	54976 5'-GAAAATCCACTGTGGGAACC-3'	54995		
	56130 5'-ATGGGAGTGAAATGGATGGA-3'	56111		
11	56072 5'-AAAGCAAGAGGCAATCATGG-3'	56091	56706 5'-CCAGCAACAACCCAATGATA-3'	56725
	57364 5'-ATGGGGTTCTGCAAGTTCAC-3'	57345	56752 5'-TGTGGCCTTATCATTTTGGA-3'	56733
12	57233 5'-TCATGGTGTGGTCTGCTTGT-3'	57252		
	58482 5'-ATGATCCCAAGGTCGGTTTT-3'	58463		

The numbering refers to the genomic DNA sequence (GenBank accession No. AC009491.3).

Appendix D. PCR conditions for amplification of 12 overlapping DNA fragments in SNPs survey

No. of fragments	PCR conditions								
	Volume (μl)	Genomic DNA (ng/tube)	Each primer (μM)	Q-Solution[a]	DNA polymerase (units/tube)	Annealing temperature (℃)	Extension time (sec)	No. of cycle	Remarks
1	50	50	0.2	−	1.0	60	30	30	
2	50	50	0.2	+	1.0	62	90	35	
3	50	100	0.2	−	1.0	62	45	35	
4	25	100	0.2	+	0.5	58	60	30	hot start PCR
5	25	50	0.4	+	0.5	62	60	30	
6	25	50	0.2	+	0.5	55	60	30	
7	25	50	0.2	+	0.5	60	30	30	
8	25	50	0.2	+	0.5	60	90	30	
9	25	50	0.2	−	0.5	60	60	30	
10	25	50	0.2	−	0.5	56	60	30	
11	25	50	0.2	+	0.5	56	60	30	
12	25	50	0.3	+	0.5	54	60	30	

[a] Q-Solution is equipped with *Taq* DNA polymerase (QIAGEN, Germany).

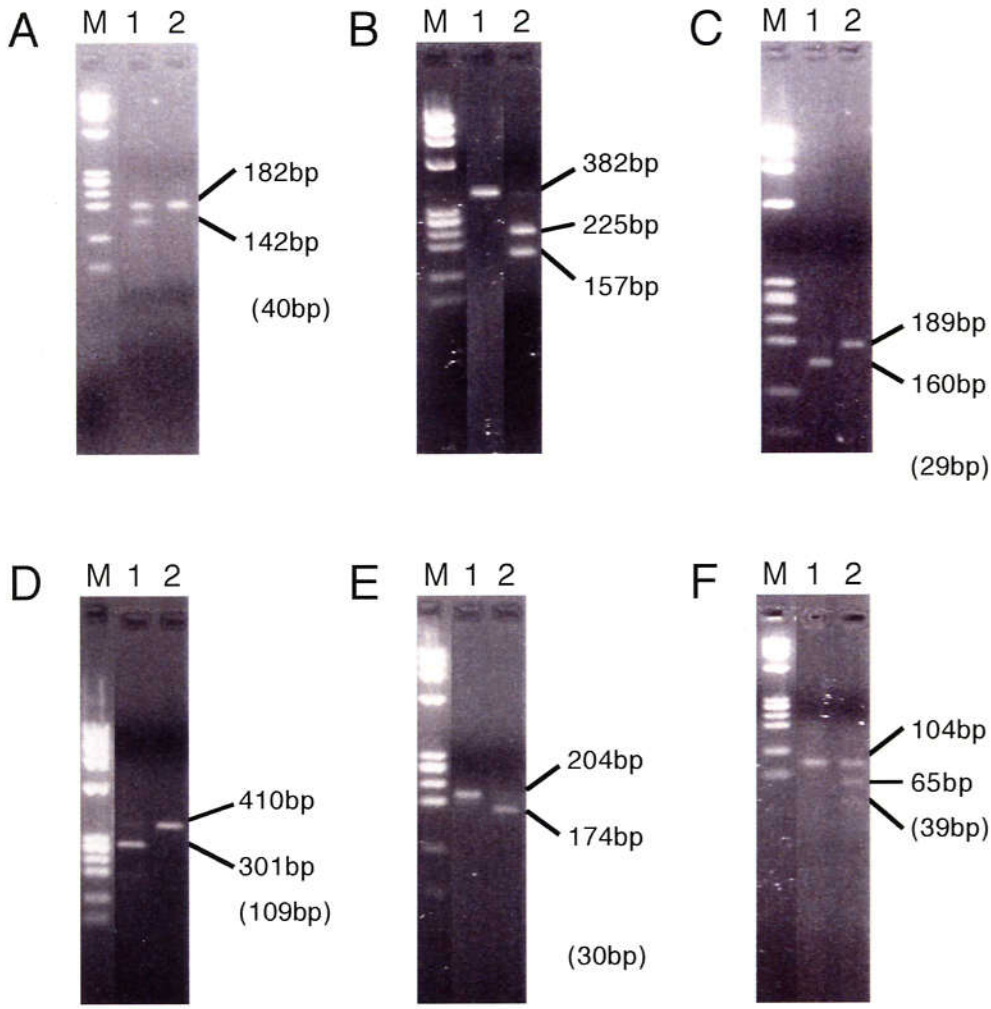
Appendix E. PCR conditions and typing systems of alleles for the six biallelic polymorphisms

Markers[a]	PCR conditions				Digestion Enzyme[c]	Typing of Alleles
	Left and right primers[b]	Annealing Temperature (°C)	Extension Time (sec)	No. of cycle		*upper*: ancestral *lower*: mutated
M8 G→T *DYS263* (Yp11.2)	5'-TGCTCATATGTCTGTGAATCAATAACTTGGACTGGGTΔCA-3' 5'-CTCAGTACTCGAGGCTGACA-3'	56	15	30	*Hpy*CH4III $AC_{\wedge}N^{\vee}\boxed{G}T$	G: 142, 40bp T: 182 bp
M55 T→C *SMCY* (Yq11.2)	5'-CGTAGGCGTTTGACAGCAG-3' 5'-CCTTTCTTCGTAATCCTCCC-3'	60	30	30	*Mae*III $^{\vee}GTNA\boxed{C}_{\wedge}$	T: 382 bp C: 225, 157 bp
M119 A→C Yq11.2	5'-TTATGGGTTATTCCAATTCAGCATACAΔGC-3' 5'-CTTAACTTAAATAGGGAAATGCC-3'	54	30	35	*Alu*I $AG^{\vee}C\boxed{T}$	A: 160, 29 bp C: 189 bp
M122 T→C Yq11.2	5'-TGTGATCAACTTCTTTCCCTCA-3' 5'-TGCAAATGGTATGCAACTCAG-3'	58	30	30	*Nla*III $_{\wedge}C\boxed{A}TG^{\vee}$	T: 301, 109 bp C: 410 bp
M125 T→C *SMCY* (Yq11.2)	5'-CTTAATAAATAGCTGCATACATCTTTTTGTA-3' 5'-GGAGCCAGCATGTGCTGTAGTT-3'	56	15	35	*Rsa*I $GT^{\vee}A\boxed{C}$	T: 204 bp C: 174, 30 bp
M217 A→C *UTY* (Yq11.1)	5'-TCTTTAACTTGTGAAGGAGAATGAAAAAGTTGGGTGGTAC-3' 5'-TTTGATAAAGCTGCTGTGGC-3'	56	15	40	*Acc*65I $G^{\vee}GTAC_{\wedge}\boxed{C}$	A: 104 bp C: 39, 65 bp

[a] Each chromosomal location of the polymorphic marker is shown in *bottom line*.

[b] The underlined base(s) in one of the primers in each pair indicates non-complementary nucleotide(s) which generates the respective recognition sequence for PCR-RFLP analysis. Nucleotide sequences for M55 primers are from Underhill et al. (2001).

[c] Recognition sequence for each enzyme is shown in *lower line*. The bordered letter in each sequence represents the respective polymorphic site.



Appendix F. Electrophoretic detection of the allele variation for each of the six polymorphic sites. (A) M8, (B) M55, (C) M119, (D) M122, (E) M125, and (F) M217. *Lane M* DNA size markers; *lane 1* ancestral allele; *lane 2* mutated allele. The details of the analyses are summarized in Appendix E and Section 4.1.2.

LIST OF ABBREVIATIONS

BP	before present
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
NRY	non-recombining region of the human Y chromosome
mtDNA	mitochondrial DNA
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
RPS4Y	ribosomal protein S4, Y-linked
SNP	single nucleotide polymorphism
SRY	sex determining region Y
STS	sequence-tagged site
TBE	50 mM Tris-borate/EDTA (pH8.3)
TMRCA	time back to most recent common ancestor
Tris	Tris(hydroxymethyl)aminomethane; 2-Amino-2-hydroxymethyl-1,3-propanediol
YAP	Y *Alu* polymorphic

REFERENCES

Adcock GJ, Dennis ES, Easteal S, Huttley GA, Jermiin LS, Peacock WJ, Thorne A (2001) Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. Proc Natl Acad Sci USA 98: 537-542

Avise JC (2000) Phylogeography. Harvard University Press, Cambridge

Avise JC, Arnold J, Ball RM, Bermingham EJ, Lamb T, Neigel JE, Reeb CA, Saunder NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann Rev Ecol Syst 18: 489-522

Bannai M, Ohashi J, Harihara S, Takahashi Y, Juji T, Omoto K, Tokunaga K (2000) Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. Tissue Antigens 55: 128-139

Bannai M, Tokunaga K, Imanishi T, Harihara S, Fujisawa K, Juji T, Omoto K (1996) HLA class II alleles in Ainu living in Hidaka District, Hokkaido, northern Japan. Am J Phys Anthropol 101: 1-9

Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD, Park SC, Tsai SJ, Goldman D (1999) An Asian-Native American paternal lineage identified by *RPS4Y* resequencing and by microsatellite haplotyping. Ann Hum Genet 63: 63-80

Bertranpetit J (2000) Genome, diversity, and origins: the Y chromosome as a storyteller. Proc Natl Acad Sci USA 97: 6927-6929

Brookfield JF (2000) Human evolution: how recent were the Y chromosome ancestors? Curr Biol 10: R722-723

Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. Science 291: 1742-1748

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 31-36

Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. Am J Hum Genet 68: 432-443

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

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

Dodo Y, Doi N, Kondo O (1998) Ainu and Ryukyuan cranial nonmetric variation: evidence which disputes the Ainu-Ryukyu common origin theory. Anthropol Sci 106: 99-120

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791

Fitch WM (1977) On the problem of discovering the most parsimonious tree. Am Nat 111: 223-257

Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46: 115-125

Griffiths RC, Tavaré S (1994) Ancestral inference in population genetics. Stat Sci 9: 307-319

Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. J Hered 89: 415-426

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11: 749-761

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378: 376-378

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56: 951-962

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, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. Genetics 145: 787-805

Hanihara K (1991) Dual structure model for the population history of the Japanese. Jpn Rev 2: 1-33

Harihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, Ellepola SB, Ishida T, Omoto K (1988) Mitochondrial DNA polymorphism among five Asian populations. Am J Hum Genet 43: 134-143

Hatta Y, Ohashi J, Imanishi T, Kamiyama H, Iha M, Simabukuro T, Ogawa A, Tanaka H, Akaza T, Gojobori T, Juji T, Tokunaga K (1999) HLA genes and haplotypes in Ryukyuans suggest recent gene flow to the Okinawa Islands. Hum Biol 71: 353-365

Horai S (1991) Molecular phylogeny and evolution of human mitochondrial DNA. In: Kimura M, Takahata N (eds) New aspects of the genetics of molecular evolution. Japan Science Society Press, Tokyo/Springer, Berlin, pp 135-152

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46: 828-842

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92: 532-536

Horai S, Kondo R, Murayama K, Hayashi S, Koike H, Nakai N (1991) Phylogenetic

affiliation of ancient and contemporary humans inferred from mitochondrial DNA. Philos Trans R Soc Lond B Biol Sci 333: 409-417

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. Am J Hum Genet 59: 579-590

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408: 708-713

The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933

Jin L, Su B (2000) Natives or immigrants: modern human origin in east Asia. Nat Rev Genet 1: 126-133

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11: 449-456

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes the human Y chromosome, disease and selection. Trends Genet 16: 356-362

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. Am J Hum Genet 69: 615-628

Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, Hammer MF (1997) Y chromosome markers and Trans-Bering Strait dispersals. Am J Phys Anthropol 102: 301-314

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

Kato T, Mizokami M, Nakano T, Kondo Y, Ohba K, Orito E, Ueda R, Mukaide M, Gurtsevitch V, Syrtsev A, Ruzibakiev R, Abdurakhnamov M, Yamashita M, Hayami

M (1997) High prevalence of GB virus C/hepatitis G virus infection among the Jewish population in Uzbekistan. Virus Res 48: 81-87

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. Am J Hum Genet 68: 173-190

Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001) African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. Science 292: 1151-1153

Kim W, Shin DJ, You SA, Kim YJ (1998) Y-specific DNA polymorphisms of the *YAP* element and the locus *DYS19* in the Korean population. J Hum Genet 43: 195-198

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108: 381-399

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) *MEGA2: Molecular Evolutionary Genetics Analysis software*. Arizona State University, Tempe, Arizona, USA

Lahr MM, Foley R (1994) Multiple dispersals and modern human origins. Evol Anthropol 3: 48-60

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543-555

Mizoguchi Y (1986) Contributions of prehistoric Far East populations to the population of modern Japan: a Q-mode path analysis based on cranial measurements. In: Akazawa T, Aiken CM (eds) Prehistoric hunter-gatherers in Japan. University of Tokyo Press, Tokyo, pp 107-136

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

Nei M (1995) The origins of human populations: genetic, linguistic, and archeological data. In: Brenner S, Hanihara K (eds) The origin and past of modern humans as viewed from DNA. World Scientific, Singapore, pp 71-91

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nei M, Ota T (1991) Evolutionary relationships of human populations at the molecular level. In: Osawa S, Honjo T (eds) Evolution of life. Springer-Verlag, Tokyo, pp 415-428

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

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19: 153-170

Nei M, Takezaki N (1996) The root of the phylogenetic tree of human populations. Mol Biol Evol 13: 170-177

Omoto K (1995) Genetic diversity and the origins of the "Mongoloids". In: Brenner S, Hanihara K (eds) The origins and past of modern humans as viewed from DNA. World Scientific, Singapore, pp 92-109

Omoto K, Saitou N (1997) Genetic origins of the Japanese: a partial support for the dual structure hypothesis. Am J Phys Anthropol 102: 437-446

Pereira L, Dupanloup I, Rosser ZH, Jobling MA, Barbujani G (2001) Y-chromosome mismatch distributions in Europe. Mol Biol Evol 18: 1259-1271

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

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791-1798

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

Qamar R, Ayub Q, Khaliq S, Mansoor A, Karafet T, Mehdi SQ, Hammer MF (1999) African and Levantine origins of Pakistani YAP$^+$ Y chromosomes. Hum Biol 71: 745-755

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

Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S (2001) Mitochondrial DNA polymorphisms in Yunnan nationalities in China. J Hum Genet 46: 211-220

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23: 437-441

Ramana G, Singh L, Chakraborty R (2001) The SRY-1532 site of the human Y chromosome is subject to recurrent single nucleotide mutations. Hum Biol 73: 71-80

Redd AJ, Stoneking M (1999) Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet 65: 808-828

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: a software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20: 278-280

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science 290: 1155-1159

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

Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol 66:

761-775

Shinka T, Tomita K, Toda T, Kotliarova SE, Lee J, Kuroki Y, Jin DK, Tokunaga K, Nakamura H, Nakahori Y (1999) Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y chromosome lineage. J Hum Genet 44: 240-245

Stoneking M (2001) Single nucleotide polymorphisms. From the evolutionary past. Nature 409: 821-822

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124: 717-733

Stumpf MP, Goldstein DB (2001) Genealogical and evolutionary inference with the human Y chromosome. Science 291: 1738-1742

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000a) Polynesian origins: insights from the Y chromosome. Proc Natl Acad Sci USA 97: 8225-8228

Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, Jin L (2000b) 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, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, 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

Suzuki H (1969) Microevolutionary changes in the Japanese population from the prehistoric age to the present-day. J Fac Sci Univ Tokyo, Sec V 3: 279-308

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595

Takahata N (1993) Allelic genealogy and human evolution. Mol Biol Evol 10: 2-22

Takahata N, Satta Y (2002) Pre-speciation coalescence and the effective size of ancestral populations. In: Slatkin M, Veuille M (eds) Modern developments in theoretical

population genetics. Oxford University Press, New York, pp 52-71

Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. Hum Genet 105: 577-581

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97: 7360-7365

Thorne A, Grün R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, Taylor L, Curnoe D (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton. J Hum Evol 36: 591-612

Tokunaga K, Bannai M, Imanishi T, Juji T (1998) HLA class II alleles and haplotypes in East-Asian populations with special reference to the Ainu. In: Omoto K, Tobias PV (eds) The origins and past of modern humans - towards reconciliation. World Scientific, Singapore, pp 74-87

Tokunaga K, Imanishi T, Takahashi K, Juji T (1996) On the origin and dispersal of east Asian populations as viewed from HLA haplotypes. In: Akazawa T, Szathmary EJE (eds) Prehistoric Mongoloid Dispersals. Oxford University Press, New York, pp 187-197

Tokunaga K, Sideltseva EW, Tanaka H, Uchikawa C, Nieda M, Sideltsev VV, Zhuravleva E, Imanishi T, Itoh K, Akaza T, Takahashi K, Khalturin V, Alexeev LP, Juji T (1995) Distribution of HLA antigens and haplotypes in the Buryat population of Siberia. Tissue Antigens 45: 98-102

Turner CG II (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, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7: 996-1005

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93: 196-200

Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65: 43-62

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26: 358-361

Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 258: 52-59

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7: 256-276

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98: 10244-10249

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378: 379-380

Yamashita M, Miura T, Ibuki K, Takehisa J, Chen J, Ido E, Hayami M (1997) Phylogenetic relationships of HTLV-I/STLV-I in the world. Leukemia 11: 50-51