

# Inferring Human History: Clues from Y-Chromosome Haplotypes

P.A. UNDERHILL

*Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120*

DNA molecules are organic elements of information storage imbued with imperfect copying processes. Thus, they are fundamental repositories of an organism's evolutionary history. The field of human molecular evolution is predicated on the concept that patterns of DNA sequence variation in living populations encode aspects of human heritage shaped by a constellation of evolutionary influences. The framework of genetic variability in the genome reflects both evolutionary adaptive processes that are locus-specific and population-level forces that affect all the components of the genome equally. Genetic research often focuses on distinguishing inconsistencies in patterns of variation between genomic regions to help bridge the gap between particular genes and traits, including matters of function and malfunction. Alternatively, genomic DNA is also an archive of those aspects of human evolutionary processes reflective of population-level forces like drift, subdivision, size fluctuation, and migration. By studying the degree of genetic molecular variation, one can, in principle, reconstruct past events such as expansions and settlements from which origins of specific populations can be predicted (Cavalli-Sforza et al. 1994). However, since the bulk of common variation in the genome occurs between individuals, the difference between populations is low, making it more challenging to investigate ambiguities concerning affinities and origins of populations. It is the component of between-population variance that best provides insights into the evolution of the spectrum of extant populations (Cavalli-Sforza and Feldman 2003). In addition, determining the migratory patterns of our ancestors and their timing, as well as the amount of population admixture due to such migrations, has been experimentally less tractable. Consequently, such issues are often simply ignored when reconstructing population phylogenies where little or no migration is implicitly assumed.

Progress in understanding the spectrum of human DNA sequence variation and its causes, especially when integrated with other knowledge from historians, archaeologists, anthropologists, and linguists, can help recapitulate human population histories (Owens and King 1999; Cann 2001). In particular, informative haplotypes, immune from the scrambling effects of recurrent mutation and recombination, provide a promising way forward. In the early genetic studies, both protein and matrilineal inherited mitochondrial DNA (mtDNA) loci have provided much of the initial evidence. However, considerable re-

cent progress in elucidating Y-chromosome sequence variation from the nonrecombining region (NRY) has made it possible to more fully investigate the parallel paternal heritage that underlies the central theme of this paper. Although extrapolating variation associated with a single gene to population history must be done cautiously, the phylogeographic reconstruction of haplotypes offers one such interpretation that is amenable to testing by further studies from a number of disciplines. Although nongenetic evidence provides constraints that improve confidence in phylogeographic-based interpretations, the development of statistical methods to objectively evaluate such geographic and phylogenetic patterns remains incomplete (Knowles and Maddison 2002). The growing accumulation of Y-chromosome haplotype data in geographic context will provide a substantial test bed for future statistical modeling.

## Y-CHROMOSOME HAPLOTYPES CONVEY SPATIAL AND TEMPORAL INFORMATION

Although explorations into prehistory have been traditionally archeological, additional perspectives have been provided by linguistic and genetic studies. Although the records of these nonrecombining sex-specific loci may diverge because of natural selection or differences between male and female behaviors, the accumulation of sequence variation during the lineal life spans of these haplotypic systems provides a powerful way to recover genetic prehistory. Specifically, the particular distinctive clinal patterns of NRY haplotypes, together with patterns of associated genetic diversification with geography, mark trajectories of gene flow (and by inference, the movement of populations). The lower effective population size of Y chromosomes relative to other components of the human genome make the Y chromosome particularly sensitive to the influences of drift and founder effect. Whatever the causes of this property (e.g., localized natural selection, gender-based differential reproductive success, and/or migratory behavior), it is particularly useful since it explains the characteristic high stratification of NRY diversity with geography relative to other genes including mtDNA. Although nearly all Y chromosomes have shared common ancestry, the succession of accumulating genetic markers reveals a cascade of differentiation that randomly coincides with various population

origin episodes, each with specific temporal and geographical context (Underhill et al. 2001b).

The Y chromosome is now the most informative haplotyping system known, making this locus an attractive reservoir of gene lineages permeated with nonrandom geographic structure. The determination of evolutionary stable haplotypes with geographic appellation provides clues to geographic ancestry. Considerable potential for additional informative haplotype resolution remains, since only a small fraction of the Y chromosome has been surveyed for informative DNA sequence variants and only a fraction of populations have been surveyed. Consequently, Y-chromosome haplotypes are ideal for assessing the origins of contemporary population affinities.

### GEOGRAPHIC PATERNAL ANCESTRY

Binary DNA sequence variants like single nucleotide substitutions (SNPs) and small insertions or deletions (DIPs) associated with the nonrecombining portion of the haploid Y chromosome provide a unique metric into population affinity and substructure. These evolutionarily stable mutations accrue throughout the lineal life span of the molecule such that their sequential accumulation across the generations can be deduced. Since most of the Y chromosome escapes recombination, one can construct an unequivocal genealogy and observe the geographic relationships of various lineages (Fig. 1). The tree is rooted using great ape sequences to deduce ancestral allele status. The lower effective population size of the Y chromosome in relation to other components in the gene pool translates into increased levels of subdivision creating the strongest geographic and greatest diversification signal among populations.

What emerges is a molecular narrative of population demography, relatedness, and genetic substrata reflecting possible signatures of initial colonists, novel micro-evolutionary differentiations during isolation, and subsequent dispersals overlaying previous ranges. These patterns of Y-chromosome lineal origins, affinity, and substructure provide a portrait of human paternal history.

The aim here is to present examples of Y-chromosome affinity and diversification. These interpretations of Y-chromosome heritage provide independent perspectives to theories of prehistorical events and resemblances based on material culture, linguistic, and other genetic knowledge. The inference of putative origins of lineages should be possible, as well as deducing their subsequent dispersal routes, by localizing regions of highest associated diversity. When high-resolution binary lineages are coupled to more rapidly mutating microsatellites or short tandem repeat loci, the combination of linked polymorphic systems provides a powerful system for understanding diversity across both intermediate and recent time frames (de Knijff 2000). These geographic patterns of genetic affinity and diversification provide intriguing insights into human history, especially the population dynamics associated with migration, population subdivision, fluctuations in population size, and subsequent gene flow episodes.

An overall picture, reflective of modern human origins, affinity, differentiation, and demographic history, has been reconstructed. Components of the Y-chromosome binary haplotype phylogeny are described according to a recently formulated standard nomenclature (Y Chromosome Consortium 2002).

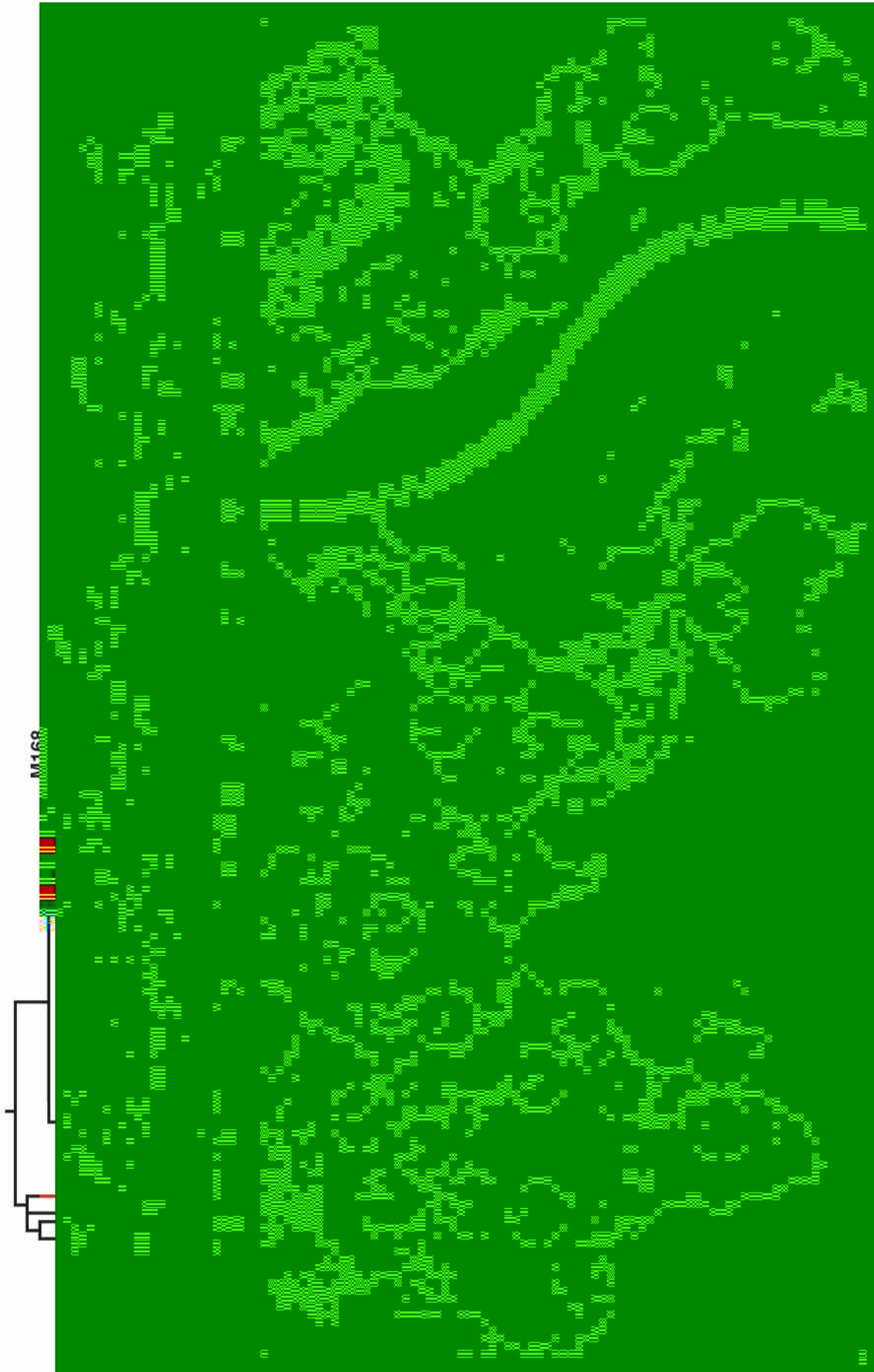
### AFRICAN HERITAGE

Over 400 binary polymorphisms currently describe the Y-chromosome tree. Several mutually reinforcing binary mutations divide the Y-chromosome haplotype phylogeny into two distinctive components, haplogroup A and the remainder of all other haplogroups, specifically B through R. The ancestral (i.e., nonhuman primate) alleles associated with these ancient polymorphisms are localized exclusively to a minority of both extant north African and sub-Saharan populations, whereas the majority of other Africans, and all non-Africans, carry only the derived mutant alleles. This mode indicates that almost all modern Y chromosomes trace their ancestry to a common primogenitor, as expected in a stable genealogy. These Y-chromosome data contradict the possibility that early hominids contributed significantly, if at all, to the gene pool of anatomically modern humans of the region (Capelli et al. 2001; Ke et al. 2001). This is evidence that all modern extant human Y chromosomes trace their ancestry to Africa and that the descendants of the derived lineage left Africa and eventually completely replaced previous archaic human Y-chromosome lineages.

A second distinctive monophyletic haplogroup called B, defined by several binary polymorphisms, is also restricted to African populations. Both A and B lineages are diverse and suggest a deeper genealogical heritage than other haplotypes. Representatives of these lineages are distributed across Africa, but generally at low frequencies. Populations represented in A and B clades include some Khoisan and Bantu speakers from South Africa, pygmies from central Africa, and lineages in Sudan, Ethiopia, and Mali (Underhill et al. 2000; Semino et al. 2002). One group B-associated lineage is shared by click speaking San of South Africa and Hadzabe of Tanzania. The genetic distance between these lineal representatives indicates that both populations have been separated for a considerable time but share ancient genetic and possibly linguistic heritage (Knight et al. 2003). The phylogenetic position of A and B lineages nearest the root of the Y tree, their survivorship in isolated populations, and accumulated variation are suggestive of an early diversification and dispersal of human populations within Africa, and an early widespread distribution of human populations in that continent. The discovery of *Homo sapiens* fossils in Ethiopia dating to 160,000 years ago is consistent with an African origin of our species (White et al. 2003).

### OUT OF AFRICA

At least three mutations lie at the root of all the remaining Y-chromosome haplotypes that compose the majority of African and non-African lineages, namely



**Figure 1.** Geographic patterns of Y-chromosome haplogroups according to recently standardized nomenclature in 22 global geographic regions. (Adapted, with permission, from Underhill et al. 2001b [copyright Blackwell Scientific].)

haplogroups C through R (Underhill et al. 2001b). The mutations that define this node deep within the interior of the Y tree reflect descendants of males who successfully left Africa and formed the scaffold on which all other Y-chromosome diversification with geography has accumulated. The geographical distribution of this diversification allows us to try to understand some of the major movements that occurred after anatomically modern humans left Africa. The original founders diversified into important lineages that display an irregular geographic distribution. The majority of Y lineages in the world are composed of a tripartite assemblage consisting of (1) haplogroup C, (2) haplogroups D and E, and (3) overarching haplogroup F that defines the internal node of all remaining haplogroups G through R. These geographic patterns of genetic affinity and diversification provide insights into the population dynamics associated with migration, population subdivision, fluctuations in population size, and more recent gene flow episodes. A synopsis of the relevant features of these haplogroups is discussed below.

### ASIAN HAPLOGROUP C

Since the mutations that define haplogroup C have not been observed in any African populations, it has been postulated that this haplogroup likely arose somewhere in Asia on an M168 lineage sometime after an early departure event prior to the arrival of modern humans to Sahul in Southeast Asia (Capelli et al. 2001; Underhill et al. 2001b). Haplogroup C comprises a collection of sublineages that display irregular geographic patterning. Most notable are C lineages that carry the M217 transversion mutation which are common in eastern Asia and Siberia with representatives in north America (Bergen et al. 1999; Karafet et al. 2001; Lell et al. 2002). Interestingly, these M217-derived lineages are absent in haplogroup C lineages seen in Indonesia (Underhill et al. 2001a), Oceania (Kayser et al. 2000, 2001), and Yunnan, China, where numerous minority populations reside (Karafet et al. 2001). Recently, such related C lineages have been observed in India at 4.6% (Kivisild et al. 2003). Their persistence in India is consistent with the model of an early coastal migration route via southwest Asia to insular southeast Asia and Oceania (Stringer 2000). The phylogeography suggests that early male colonizers to Australia were haplogroup C descendants, and the presence of the M217 sublineage in Siberia (Karafet et al. 2001, 2002; Lell et al. 2002) is consistent with diversification and northward migration since the last ice age.

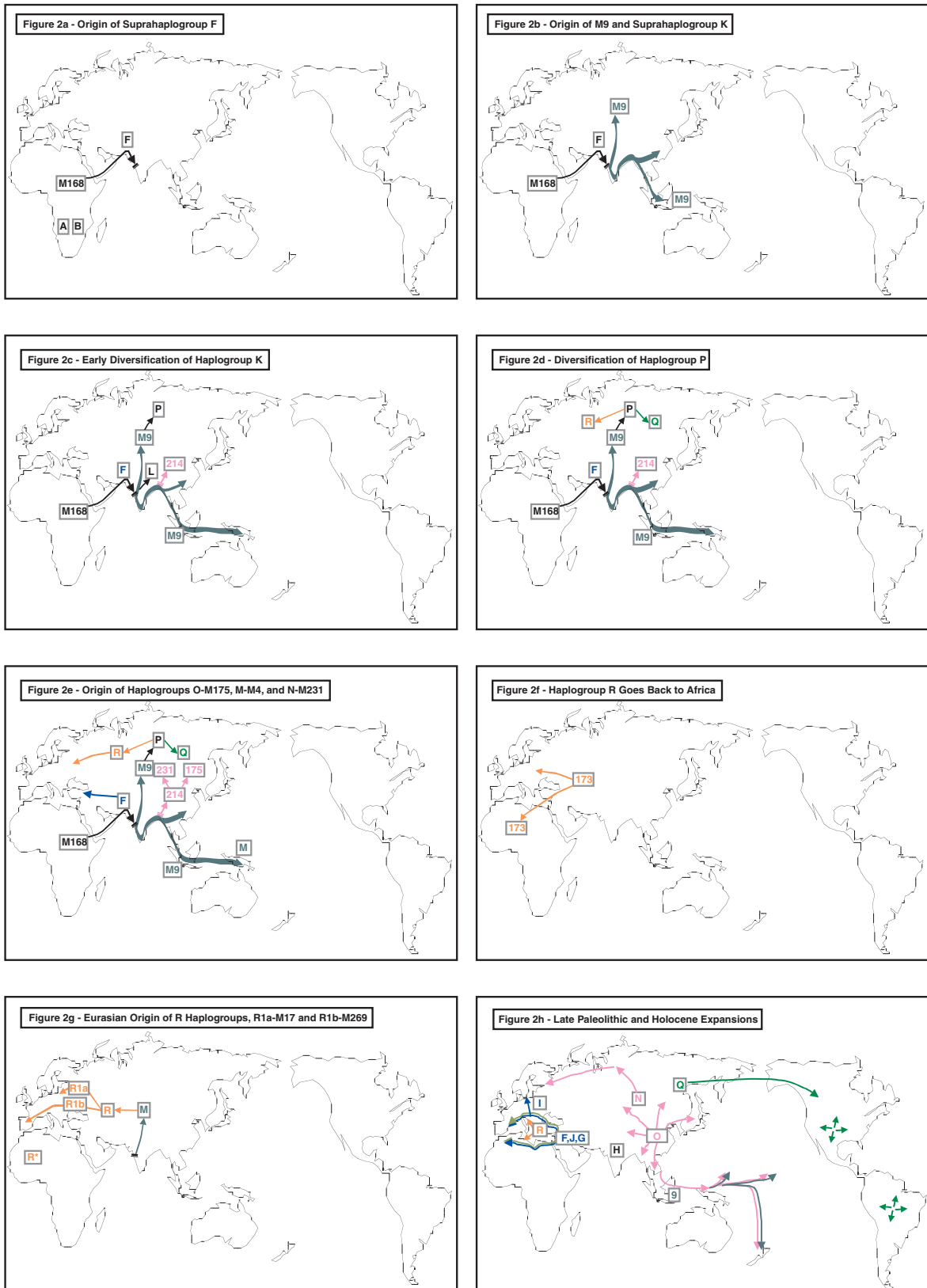
### ASIAN AND AFRICAN HAPLOGROUPS D AND E

Although each monophyletic haplogroup displays continental separate distributions, both share three phylogenetically equivalent binary markers indicative of unequivocal shared heritage. The ancestors who accumulated these three mutations could have just as well arisen in Africa as in Asia. Despite the apparent absence of any intermediate haplotypes based on these three binary poly-

morphisms, Africa remains the most plausible geographic origin of these three relatively old polymorphisms (Underhill and Roseman 2001). It appears that some descendants with these three mutations remained and some left Africa to become part of the gene pool of the early successful colonizers in Asia. Following geographic separation, subsequent continent-specific mutations arose creating the two monophyletic D and E clades in Asia and Africa, respectively. Haplogroup E lineages are the most frequent in Africa and display subsequent binary and microsatellite diversification. Conversely, Asian haplogroup D generally occurs at low frequencies throughout eastern Asia, except in peripheral locations like Tibet, Japan, and the Andaman Islands, where significant frequencies have been observed, most likely because of founder effects (Underhill et al. 2001b; Thangaraj et al. 2003). The Ainu of Japan are composed of both C and most likely D representatives (Tajima et al. 2002). The phylogeography suggests that Asian D lineages are likely the descendants of early Asian colonizers who arrived from Africa. To a large degree, they have been subsequently displaced to geographic margins by pressures from demic expansions by ensuing peoples.

### HAPLOGROUP F ACROSS THE ENTIRE WORLD EXCEPT SUB-SAHARAN AFRICA

The third major and most peripatetic subcluster of M168 lineages is characterized by at least three mutations (one of which is M89) that define the root of haplogroup F from which all other haplogroups (G through R) deploy. This F subcluster (Fig. 2a) is suggested to have evolved outside Africa early in the diversification and migration of modern humans (Kivisild et al. 2003). Early Upper Paleolithic peoples throughout Eurasia provide sources from which later populations derive. The differentiation of haplogroup F (Fig. 2a–h) within Eurasia helps to begin understanding this complex period of the peopling of the world (Underhill et al. 2001b). The Middle East has major representatives of haplogroups G, J, and R. India also has J, which may have arrived with agriculturalists. Interestingly, it has some F and H lineages seldom observed elsewhere. An expansion of F lineages toward central Asia or the Caucasus would have given rise to a population that acquired the M9 mutation which defines a major bifurcation in the phylogeny (Fig. 2b). Haplogroup K, L, and M lineages all descend from an M9 ancestor and are widespread with some distinctive K lineages being observed in India, the Middle East, and Europe. Haplogroup L has greatest frequency in southwest Asia, and distinctive K and M lineages are restricted to Oceania (Fig. 2c). The M9 mutation also lies at the root of haplogroups O through R, all of which are Eurasian and African in distribution. The population carrying the M9 mutation must have expanded widely (Fig. 2b), with one in north Asia characterized by the haplogroup P, which encompasses distinctive eastward expanding Q (Siberian and American) and westward expanding R (Eurasian) lineages (Fig. 2d), and another one in eastern Asia characterized by monophyletic haplogroups N and



**Figure 2.** Phylogeographic inferences of the origin of suprahaplogroup F and its subsequent diversification across the world. (Adapted, with permission, from Underhill et al. 2001b [copyright Blackwell Scientific].)

O (Fig. 2e) that share a common unifying mutation. In summary, the early diversification of a haplogroup F population in Eurasia between 40,000 and 30,000 years ago would have given rise to at least six Y-chromosome populations (Underhill et al. 2001b). Thus, there were multiple independent formations and fragmentations of populations carrying F-related lineages throughout most of Asia, displacing the earlier haplogroup C and D lineages toward the margins.

More recent expansion events following population contraction associated with the last Ice Age 18,000–16,000 years ago are detectable in the Y-chromosome phylogeny (Fig. 2h). These include two main R lineages, R1a and R1b, with distinctive European geography (Semino et al. 2000), and Neolithic farmers from the Near East (haplogroups E and J). Interestingly, the diversification and phylogeographic patterns of Y chromosomes indicate a rather old back-to-Africa migration of Euroasian R lineages (Fig. 2f) prior to the widespread dispersion of high-frequency R1a and R1b lineages (Fig. 2g) that are not observed in Africa (Cruciani et al. 2002). Haplogroup O and N lineages are common in eastern Asia and may reflect the impact of millet and rice (Cavalli-Sforza et al. 1994). The apparent success of O and N lineages in eastern Asia appear analogous to certain widespread African E lineages that are common in Bantu-speaking agriculturalists (Diamond and Bellwood 2003).

### CONCLUSIONS

The haploid Y chromosome is unusual in that it is depauperate in genes relative to other nuclear chromosomes. However, the other unusual innate properties of being largely nonrecombining as well as having a low effective population size relative to other loci, combine both to preserve haplotypes over evolutionary time scales and to record numerous episodes of population divergence, even on micro-geographic scales, making it perhaps the single most insightful haplotype system known to characterize population affinity, substructure, and history. Some Y-chromosome polymorphisms could become part of a genome-wide inventory of genomic control markers useful in assessing the influences of population stratification. Both the Y chromosome and autosomes can be evaluated as SNPSTR systems with the empirical determination of phase providing an index of haplotype deterioration (Mountain et al. 2002). The Y chromosome provides a comparative model for evaluating haplotypes from other regions of the genome. The recovery of complex scenarios can be best advanced via an integrative approach, since the totality of the evidence should be reflective of an overall history and some correlation should be expected. When the story lines from multiple genes reinforce one another, overall population histories are revealed. Conversely, when different genes yield different haplotype patterns, locus-specific forces are in play. The recent and ongoing progress in deciphering the Y-chromosome structure in contemporary populations provides new opportunities to formulate specific testable hypotheses involving human evolutionary popu-

lation genetics. Although the genetic legacy of *Homo sapiens* remains incomplete, the recent ability to unearth new levels of shared Y-chromosome haplotypic heritage and subsequent diversification provide not only an index of contemporary population structure, but also a preamble to human prehistory and substantial foundation for comparisons with other genomic regions.

### ACKNOWLEDGMENTS

Although many colleagues have helped to accelerate progress in understanding Y-chromosome variation in human populations, the contributions of Drs. P.J. Oefner and L.L. Cavalli-Sforza deserve particular recognition.

### REFERENCES

- Bergen A.W., Wang C.-Y., Tsai J., Jefferson K., Dey C., Smith K.D., Park S.-C., Tsai S.-J., and Goldman D. 1999. An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* **63**: 63.
- Cann R.L. 2001. Genetic clues to dispersal in human population: Retracing the past from the present. *Science* **291**: 1742.
- Capelli C., Wilson J.F., Richards M., Stumpf M.P.H., Gratrix F., Oppenheimer S., Underhill P.A., Pascali V.L., Ko T.-M., and Goldstein D.B. 2001. A predominantly indigenous paternal heritage for the Austronesian speaking peoples of insular South East Asia and Oceania. *Am. J. Hum. Genet.* **68**: 432.
- Cavalli-Sforza L.L. and Feldman M.W. 2003. The application of molecular genetic approaches to the study of human diversity. *Nat. Genet.* **33**: 267.
- Cavalli-Sforza L.L., Menozzi P., and Piazza A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, New Jersey.
- Cruciani F., Santolamazza P., Shen P., Macaulay V., Moral P., Olckers A., Modiano D., Destro-Bisol G., Holmes S., Coia V., Wallace D.C., Oefner P.J., Torroni A., Cavalli-Sforza L.L., Scozzari R., and Underhill P.A. 2002. An Asia to Sub-Saharan Africa back migration is supported by high-resolution analysis of human Y chromosome haplotypes. *Am. J. Hum. Genet.* **70**: 1197.
- de Knijff P. 2000. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* **67**: 1055.
- Diamond J. and Bellwood P. 2003. Farmers and their languages: The first expansions. *Science* **300**: 597.
- Karafet T.M., Osipova L.P., Gubina M.A., Posukh O.L., Zegura S.L., and Hammer M.F. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.* **74**: 761.
- Karafet T., Xu L., Du R., Wang W., Feng S., Wells R.S., Redd A.J., Zegura S.L., and Hammer M.F. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* **69**: 615.
- Kayser M., Brauer S., Weiss G., Schiefenhövel W., Underhill P.A., and Stoneking M. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**: 173.
- Kayser M., Brauer S., Weiss G., Underhill P.A., Roewer L., Schiefenhövel W., and Stoneking M. 2000. Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**: 1237.
- Ke Y., Su B., Son X., Lu D., Chen L., Li H., Qi C., Marzuki S., Deka R., Underhill P., Xiao C., Shriver M., Lell J., Wallace D., Wells R.S., Seielstad M., Oefner P., Zhu D., Jin J., Huang W., Chakraborty R., Chen Z., and Jin L. 2001. African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes. *Science* **292**: 1151.

- Kivisild T., Rootsi S., Metspalu M., Mastana S., Kaldma K., Parik J., Metspalu E., Adojaan M., Tolk H.-V., Stepanov V., Gölge M., Usanga E., Papiha S.S., Cinnioglu C., King R., Cavalli-Sforza L., Underhill P.A., and Villems R. 2003. The genetic heritage of earliest settlers persist in both the Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**: 313.
- Knight A., Underhill P.A., Zhitovovskiy L.A., Mortensen H.M., Ruhlen M., and Mountain J.L. 2003. African Y chromosome and mtDNA diversity and the antiquity of click languages. *Curr. Biol.* **13**: 464.
- Knowles L.L. and Maddison W.P. 2002. Statistical phylogeography. *Mol. Ecol.* **11**: 2623.
- Lell J.T., Sukernik R.I., Starikovskaya Y.B., Su B., Jin L., Schurr T.G., Underhill P.A., and Wallace D.C. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* **70**: 192.
- Mountain J.L., Knight A., Jobin M., Gignoux C., Miller A., Lin A.A., and Underhill P.A. 2002. SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* **12**: 1766.
- Owens K. and King M.-C. 1999. Genomic views of human history. *Science* **286**: 451.
- Semino O., Santachiara-Benerecetti A.S., Falaschi F., Cavalli-Sforza L.L., and Underhill P.A. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**: 265.
- Semino O., Passarino G., Oefner P.J., Lin A.A., Arbuzova S., Beckman L.E., De Benedictis G., Francalacci P., Kouvatsi A., Limborska S., Marcikiae M., Mika A., Mika B., Primorac D., Santachiara-Benerecetti A., Cavalli-Sforza L.L., and Underhill P.A. 2000. The genetic legacy of Palaeolithic *Homo sapiens sapiens* in extant Europeans: A Y-chromosome perspective. *Science* **290**: 1155.
- Stringer C. 2000. Coasting out of Africa. *Nature* **405**: 24.
- Tajima A., Pan I.-H., Fucharoen G., Fucharoen S., Matsuo M., Tokunaga K., Juji T., Hayami M., Omoto K., and Horai S. 2002. Three major lineages of Asian Y chromosomes: Implications for the peopling of east and southeast Asia. *Hum. Genet.* **110**: 80.
- Thangaraj K., Singh L., Reddy A.G., Rao V.R., Underhill P.A., Pierson M., Frame I.G., and Hagerberg E. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13**: 86.
- Underhill P.A. and Roseman C.C. 2001. The case for an African rather than an Asian origin of the human Y-chromosome YAP insertion. In *Recent advances in human biology*, vol. 8: *Genetic, linguistic and archaeological perspectives on human diversity in Southeast Asia* (ed. L. Jin et al.), p. 43. World Scientific, River Edge, New Jersey.
- Underhill P.A., Passarino G., Lin A.A., Marzuki S., Cavalli-Sforza L.L., and Chambers G. 2001a. Maori origins, Y chromosome haplotypes and implications for human history in the Pacific. *Hum. Mutat.* **17**: 271.
- Underhill P.A., Passarino G., Lin A.A., Shen P., Foley R.A., Mirazón Lahr M., Oefner P.J., and Cavalli-Sforza L.L. 2001b. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**: 43.
- Underhill P.A., Shen P., Lin A.A., Jin L., Passarino G., Yang W.H., Kauffman E., Bonn -Tamir B., Bertranpetit J., Francalacci P., Ibrahim M., Jenkins T., Kidd J.R., Mehdi S.Q., Seielstad M.T., Wells R.S., Piazza A., Davis R.W., Feldman M.W., Cavalli-Sforza L.L., and Oefner P.J. 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**: 358.
- White T.D., Asfaw B., DeGusta D., Gilbert H., Richards G.D., Suwa G., and Howell F.C. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**: 742.
- Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**: 339.

