Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations

Peter A. Underhill[1] and Toomas Kivisild[2]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5120; email: under@stanford.edu

[2]Leverhulme Center of Human Evolutionary Studies, University of Cambridge, Cambridge CB2 1QH, United Kingdom; email: tk331@cam.ac.uk

Annu. Rev. Genet. 2007. 41:539–64

This article's doi:
10.1146/annurev.genet.41.110306.130407

0066-4197/07/1201-0539$20.00

Key Words

population genetics, human evolution, haploid phylogenies

Abstract

Well-resolved molecular gene trees illustrate the concept of descent with modification and exhibit the opposing processes of drift and migration, both of which influence population structure. Phylogenies of the maternally inherited mtDNA genome and the paternally inherited portion of the nonrecombining Y chromosome retain sequential records of the accumulation of genetic diversity. Although knowledge regarding the diversity of the entire human genome will be needed to completely characterize human genetic evolution, these uniparentally inherited loci are unique indicators of gender in modulating the extant population structure. We compare and contrast these loci for patterns of continuity and discreteness and discuss how their phylogenetic diversity and progression provide means to disentangle ancient colonization events by pioneering migrants from subsequent overlying migrations. We introduce new results concerning Y chromosome founder haplogroups C, DE, and F that resolve their previous trifurcation and improve the harmony with the mtDNA recapitulation of the out-of-Africa migration.

FRAMEWORK OF HUMAN GENETIC DIFFERENCES

mtDNA: mitochondrial DNA

NRY: nonrecombining region of the Y chromosome

Homo sapiens can be described as a cosmopolitan species. Despite the wide range of our habitat occupation we are characterized by low intraspecies genetic variation. As our genome shows 1.23% average difference in its nucleotide variation from our closest living sibling species—the chimpanzee (18), the average genomic differences between a pair of humans taken across the world can be one difference per 1000 base pairs or even less, depending on the genetic locus and the particular population we are dealing with. Even though the interpopulation differences are minor compared with within-population differences, it is nevertheless possible, using only a small number of genetic traits, to distinguish, with certain likelihood, individuals of different continental affiliation, and perhaps even to define through genetic markers the actual populations themselves rather than assume their existence (14, 35). Still under debate is how much discontinuity would be observed in genetic patterning between continental population groups (**Figure 1**) were the sampling of populations sufficiently dense and unbiased (87, 88, 94). Whether the continuum of genetic differences between human populations is smooth or bumpy, geography rather than ethnicity seems to be the driving factor in such patterning (65, 78). Some of these differences may have arisen as a consequence of neutral evolution due to random genetic drift, and some others due to selection, the effect of which complicates the task of inferring from the dynamics of genetic patterning in time (90). Nonetheless, if we are able to discern genetic differences between the populations, we may ask the challenging questions of how and when they have arisen. Does the architecture of human genetic differences over vast geographic ranges stem from a long-term segregation of continental gene pools, as suggested by the modern draughtsmen of the multiregional model [e.g., (103)] or has it arisen from small founder demes during the past 100,000 years, as suggested by the replacement theory (11, 101)? Are these differences due to selection of a small number of physically or chemically expressed traits? Or is their existence mainly a natural consequence of stochastic drift that shaped our common ancestral populations in the context of their geographic appellation?

Over the past 20 years or more, approaches to answer these questions in regard to prehistorical range expansions, demographic events, reciprocal gene flows, and contemporary population substructure have been mainly through two uniparentally inherited marker systems—mtDNA (mitochondrial DNA) and the Y chromosome (10, 11, 36, 46, 49, 67, 113). Even though whole-genomic approaches (19, 42, 60, 88, 118) are now opening up new avenues to answer these questions related to the origin and diversification of our species, mtDNA and the Y chromosome, with their unique patterns of inheritance, continue to be important sources of information. The past five years have seen significant progress in reconstructing the detailed genealogical branching order of the tree topologies for both mtDNA (**Figure 2**) and the nonrecombining portion of the Y chromosome (NRY) (**Figure 3**). These phylogenies provide emblematic representations of both the clinality and discreteness embodied by human genetic variation.

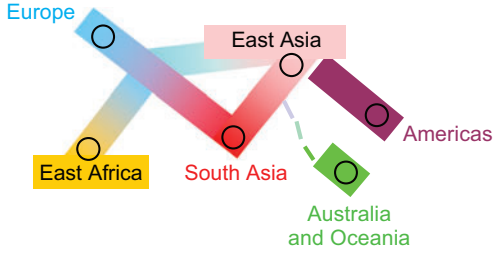


Figure 1

Clusters or clines of genetic diversity? Genetic structure of human populations could well be compatible with both concepts as clines would not be detected if there was no genetic patterning. Circles indicate theoretical sampling points, each with particular allelic frequency pattern. Blended colors between the circles denote the clines.

CONTINUITY AND DISCRETENESS OF HUMAN GENETIC DIVERSITY

Population structure, historically a basic concept in population genetics (124), can be inferred from the distribution of allelic variants in and between populations. By phylogenetic analysis of any particular locus, it is possible to define the hierarchic descent order of the genetic variants and from the tree to infer the level of structure among assessed populations, and, importantly, to infer the order and time of their descent. For example, non-African populations are nested within the African variation both in trees drawn from mtDNA and Y chromosome data, in coherence with the hypothesis of recent African descent of all non-African genetic variation. Down the line of descent, both mtDNA and Y chromosome trees support the distinction of continental gene pools through the low frequency of locally born genetic variants that have been detected through extensive sequencing of worldwide samples (**Figures 2**, **3**). Similarly, the phylogenetic approach to determine evolutionary history of certain genes has been applied in the study of our nuclear genome. Phylogenetic analyses of the X chromosome, for example, have consistently (13 genes out of 16) provided support to the general tree topology of mtDNA and Y chromosome, with Africans showing an older, most recent common ancestry than the rest of the world (37, 96). The few exceptions to the pattern could be explained simply by insufficient sampling of Africans in the original screening sets rather than by more dramatic introgression scenarios (31, 96). However, as discussed in more detail below, the sequence lengths examined to date in nuclear genes generally lack sufficient resolution and power to be informative about further population structure and genetic patterning within and between continents. Alternative approaches to determine population structure include the identification of ancestry informative markers and clustering algorithms applied to a large number of independent genetic markers over the genome (88, 97). Here we review the genetic structure of human populations as revealed through the phylogenetic approach applied to uniparental genomes and compare these findings with population structure inferred from autosomal genes (**Figure 4**).

GEOGRAPHIC STRUCTURE OF HUMAN mtDNA HAPLOGROUPS

The human mtDNA tree (**Figure 2**) splits at its core layers into branches that carry exclusively African sequences and just one, more peripheral haplogroup L3, that the Africans share with the rest of the world (11, 15, 45, 110, 119). All non-African mtDNA lineages derive from just two limbs (M and N) branching out from the root of haplogroup L3 that also has given rise to a number of subclades specific only to African populations (51, 77). The number of extant non-African founder haplogroups can, however, be extended to include a third member, haplogroup R, which is a daughter-clade of N. The transcontinental founder status for non-African populations assumes that the root of the ancient haplogroup considered should also be widely dispersed in non-African populations. In addition to M and N haplogroups, this assumption holds true also for haplogroup R because Europeans from R-derived haplogroups T and H, for example, are as distantly related to each other in their maternal geneaologies as they are with any Asian or Oceanian individual from haplogroup B or any Papuan/Australian aboriginals from haplogroup P. Haplogroup B, on the other hand, despite being almost as ancient as its parental group R, has more localized spread restricted to the eastern hemisphere, and it is more plausible to assume that it had not yet originated in the genetic substrates of the population(s) that left Africa and dispersed worldwide but, more plausibly, that it arose later within the East Asian founder population.

The first informative phylogeographic split in the human mtDNA tree occurs at the level of L3/M,N,R clades (**Figure 2**) and

corresponds thus to the K = 2 level of structuring populations sensu Rosenberg et al. (**Figure 4**), where K refers to the number of distinct clusters into which the data set is partitioned. The next informative split in the mtDNA tree distinguishes all major continents (K = 5) except the Americas beneath the M, N, and R founders. The lack of intermediate-nested K = 3 and K = 4 structures (e.g., into Africans, Southwest Asians, and the rest) has been explained by the fast pace colonization model of the world outside Africa (64, 107). Detection of further regional differences consumes extensive sequencing resources. European and Near-Eastern populations have gained their maternal pedigrees primarily from haplogroups N1, W, X, JT, R0 (including R0a, H, and V), and U *Erratum

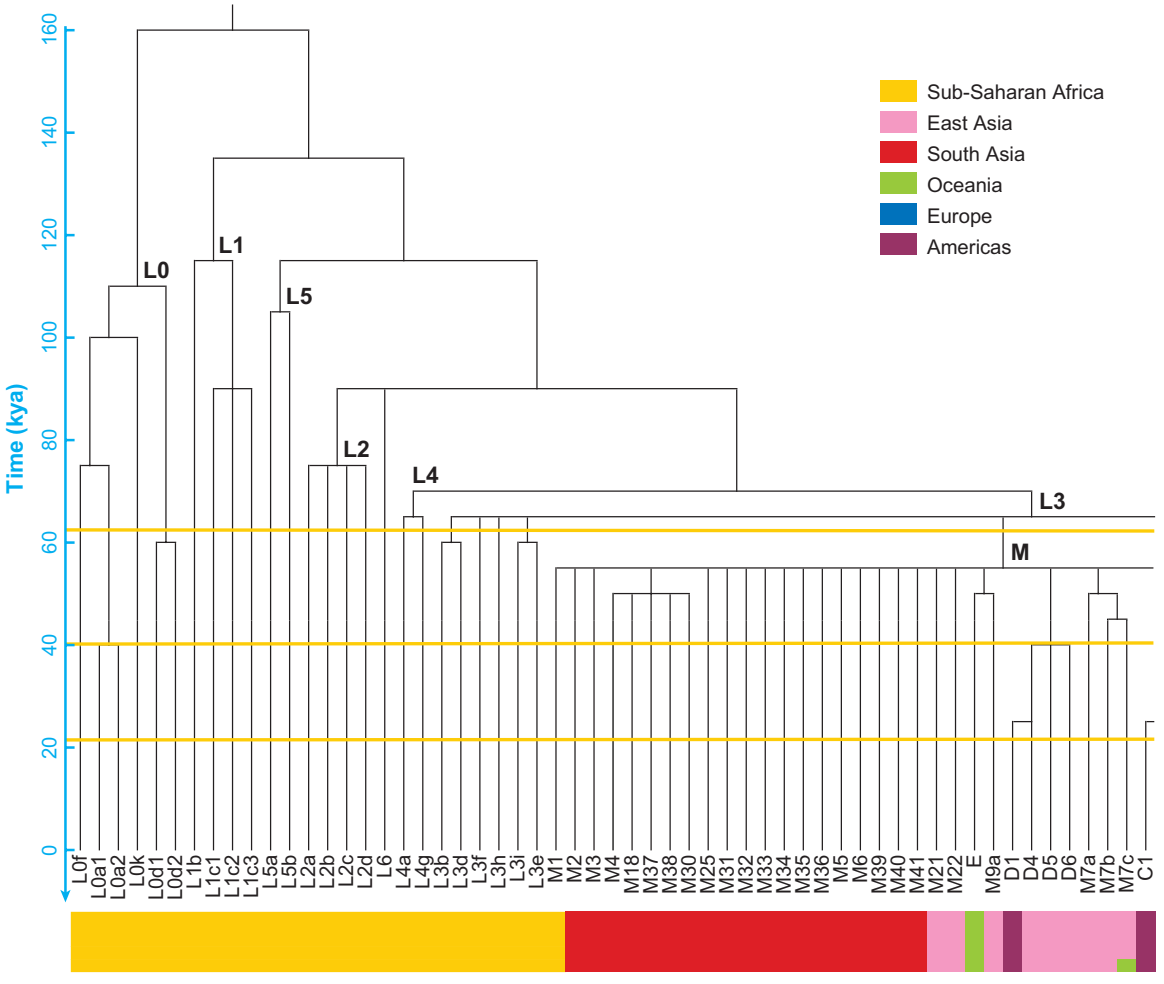


Figure 2

General haplogroup structure of mtDNA global phylogeny. Geographic affiliation of the haplogroups is indicated below the tree by the color bar (33, 43, 53, 56, 64, 69, 74, 102, 105, 110). Structural differences between human continental population groups that arise in the tree are indicated by horizontal lines and assisted with the number of clusters (K) that can be distinguished at the respective level of hierarchy. Each line corresponds to the depth of the tree at which additional region-specific variation can be

*This PDF ammended on (8 Jan. 2008): See explanation at http://arjournals.annualreviews.org/errata/genet

(**Figure 2**). The first three of these coalesce at the root of N, whereas the remaining three share their most recent common ancestor in R, the daughter limb of N. There are no significant frequency differences of these major haplogroups between geographically distinct populations of Europe. Genetic distinction between the geographic subregions within Europe becomes clear only at the fine level of subclades of the high-resolution mtDNA haplogroup tree based on complete sequence data (1, 9, 62, 75). Haplogroups H2a and H3, for example, have contrasting frequency peaks in East and West Europe, respectively.

Using complete sequence information, phylogenetic distinction can also be made between South and East Asian, Malaysian, and Island Southeast Asian branches of the M, N, and R founder groups (53, 56, 64, 102). Similarly, Melanesian, Papuan, and Australian complete sequences derived from these founder clades stand out as unique (27, 28, 43, 69, 117). Central Asian and Native American haplogroup pools, instead, can be seen as subsets of the East Asian mtDNA variation and thus the K = 6 line can be dated as fairly recent (111). Again, as noted for Europe, better microgeographic resolution within these "eastern" lineage groups can be obtained by focusing on the subclades of each haplogroup through increasing sequence and sample resolution: Haplogroup M7a, for example, is restricted in its spread to Japan and South Korea (54), and only one specific branch of B4a can be found in Austronesian-speaking populations of Taiwan and Polynesia (112).

SYNOPSIS OF THE Y CHROMOSOME TREE

The Y chromosome contains the largest nonrecombining block in the human genome and can be considered one of the most informative haplotyping systems, with applications in evolutionary population studies, forensics, medical genetics, and genealogical reconstruction. Since the report of the first Y chromosome polymorphism (13), more than a decade elapsed before a well-resolved phylogenetic tree of Y chromosome binary

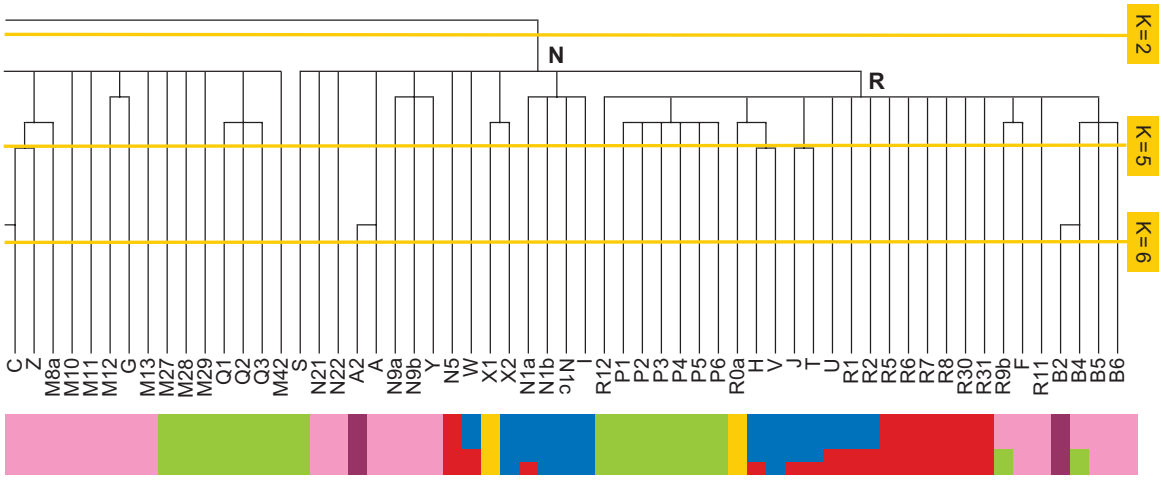


Figure 2 (*Continued*)

identified. The K = 2 line corresponds to the time frame when haplogroups M and N emerge from haplogroup L3, which thus makes the first distinction between Africans and non-Africans. The next line, K = 5, corresponds to the time frame when non-African cluster diverges into European, South Asian, East Asian, and Oceanian variants, while K = 6 distinguishes Native American sub-clades of haplogroups A-D from respective East Asian branches of the tree.

*Erratum

*Erratum posted on (8 Jan. 2008): See explanation at http://arjournals.annualreviews.org/errata/genet

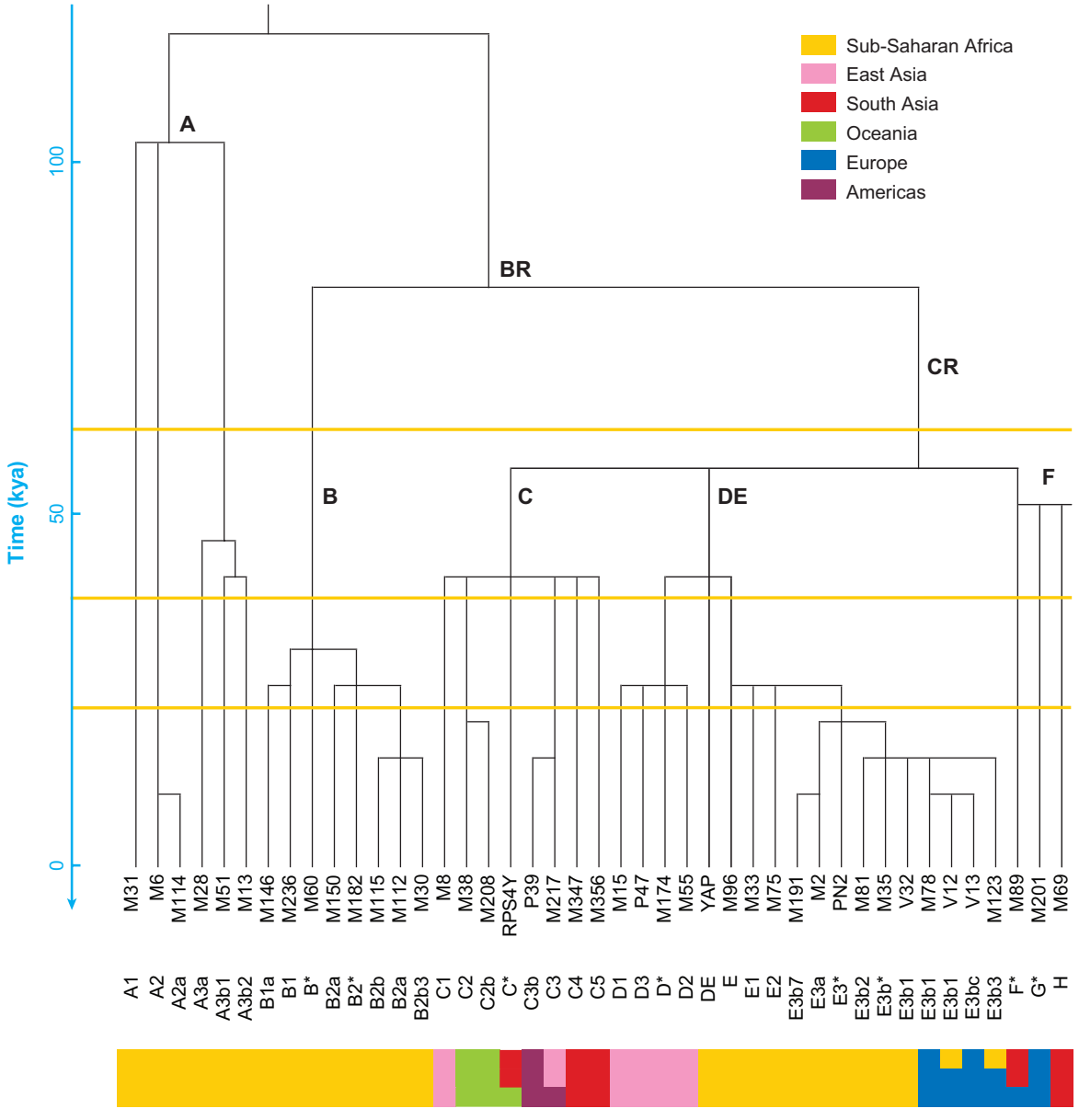


Figure 3

General haplogroup structure of nonrecombining portion of the Y chromosome global phylogeny. Geographic affiliation of the haplogroups is indicated below the tree by the color bar. Both Y Chromosome Consortium nomenclature and defining binary markers are specified beneath respective branches (16, 47, 85, 126, 127). Structural differences between human continental population groups that arise in the tree are indicated by horizontal lines and assisted with the number of clusters (K) that can be

markers emerged and became an important template in human migration studies (39, 116, 125).

By convention, Y chromosomes identified by binary polymorphisms are designated to haplogroups or clades; those that are defined only by short tandem repeats are called haplotypes, and descriptions of data combining both biallelic markers and Y-STRs are referred to as lineages (23). In the Y chromosome haplogroup tree (**Figure 3**), the two primary splits lead to haplogroups, A and B, the spread of which is restricted to Africa. Both primary haplogroups are genetically diverse with subhaplogroups geographically distinct from one another, a pattern consistent with population fragmentation, isolation, and subsequent re-expansions in Africa. The remainder of the deep structure of the phylogeny is characterized by three subclusters that coalesce at the root of the CR-M168 node, which represents the majority of African varieties as well as all the non-African haplogroups (114). This level of structuring of continental pools of Y chromosomes (K = 2) includes: (*i*) the shared presence of haplogroup DE chromosomes in Africa and Asia; (*ii*) the non-African haplogroup C, which is widely distributed in East Asia, Oceania, and North America; and (*iii*) a global distribution of another non-African cluster, haplogroup F-M89, with its most prolific daughter-group haplogroup K. Considerable regionalization of haplogroups is evident

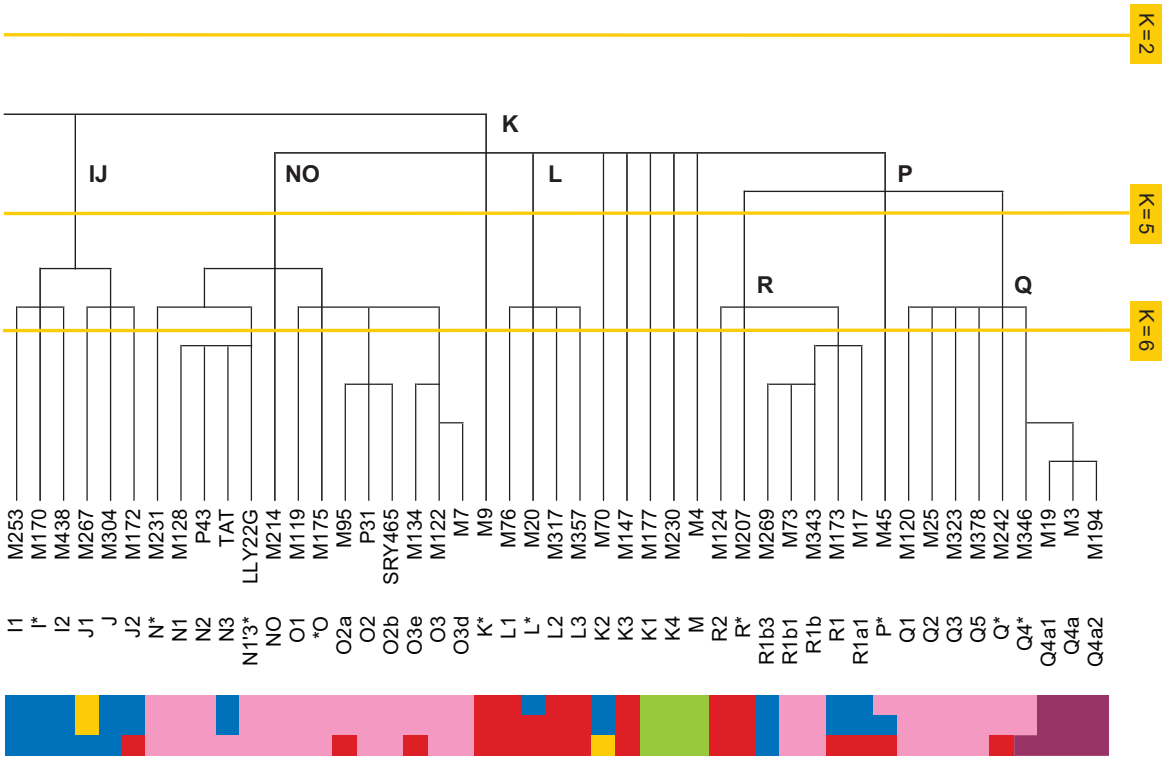


Figure 3 (*Continued*)

distinguished at the respective level of hierarchy. Each line corresponds to the depth of the tree at which additional region-specific variation can be identified. The K = 2 line corresponds to differentiation of non-African haplogroups C and F from the African tree. The next line, K = 5, focuses on further regional differentiation of haplogroups M in Oceania, IJ and R in West, H and L in South, and C, NO, and Q in East Eurasia. The K = 6 line distinguishes Native American sub-clades of haplogroups Q and C from the respective East Asian branches of the tree.

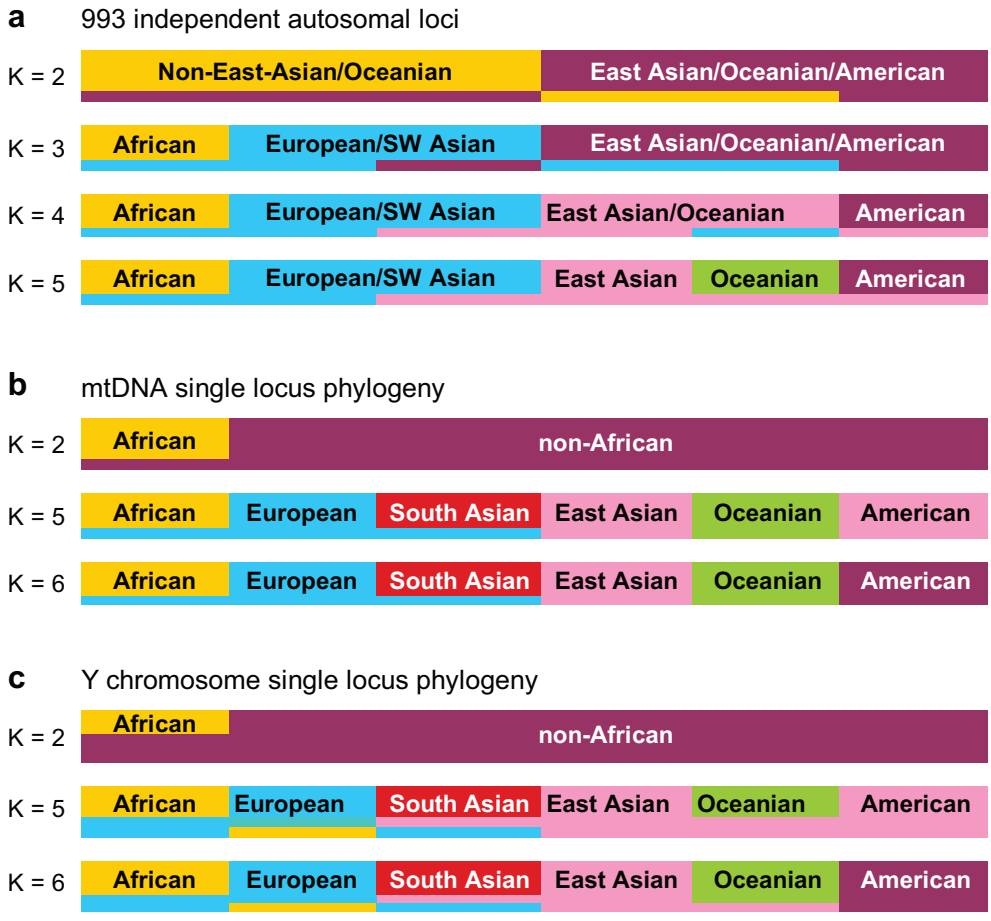


Figure 4

The hierarchy of human population structure as determined by (*a*) 993 autosomal STR markers (87), (*b*) mtDNA phylogeny (see **Figure 2** for details), and (*c*) Y chromosome phylogeny (see **Figure 3** for details). K refers to the number of identified clusters. Although linked haploid loci cannot actually be analyzed by algorithms designed for analysis of independent loci and their frequencies for qualitative comparisons, we placed three K levels of basic structure to approximate the various main hierarchical levels in the haploid trees. For mtDNA and Y chromosome see corresponding breaking lines in **Figures 2** and **3**, respectively.

in the subclades of F and K (**Figure 3**, K = 5 and K = 6). Haplogroups F* and H are quite restricted to the Asian subcontinent whereas the center of gravity for haplogroups I and J is in Europe and the Middle East, respectively (85, 92). In East Asia, haplogroups N and O that arise from the haplogroup K branch are the most frequent. Other important K-affiliated haplogroups include Q in Northeast Asia and the Americas as well as haplogroup R whose phylogeography spans North Africa and West Asia and manifests high frequencies in Europe (116).

CONSISTENCY BETWEEN THE mtDNA AND Y TREES

Genetic admixture can be sex or gene specific because different mating and migration patterns exist in populations and because different genes are subject to different selective forces. Therefore, it is not unexpected

that the two uniparentally inherited marker systems can occasionally provide evidence for different evolutionary histories within the same geographic regions (8, 12, 24, 73). Such well-explained examples seem rather to be exceptions, and ordinarily the whole-genomic composite of independent loci would be expected to initially carry the signatures of the same demographic events. It would be implausible, for example, to imagine the successful colonization of a vacant territory by a single gene or just one sex (83). Many of the statistical tests designed to identify signatures of selection can be inversely applied to detect demographic changes in populations (72). Tajima's D test, one of the most commonly used, for example, typically yields negative values among non-African populations, consistent with the recent out-of-Africa model (30).

The following notable features are examples of consistency between the phylogenies derived from mtDNA and Y chromosome data sets.

1. Only African populations carry the derived lineages of both of the primary branches descending from the root of the tree. The distribution of at least the five first branchings in the mtDNA tree (**Figure 2**) and the two first in Y chromosome-based phylogeny (**Figure 3**) support the African root.
2. Only a small subset of both trees is dispersed outside Africa. Three Y chromosome founder groups (C, D, and F) and three mitochondrial groups (M, N, and R) capture the non-African variation (52).
3. Geographically peripheral regions such as Europe and Australia show more limited founder composition as compared to Asia where all three founder groups have been preserved.
4. Certain well-known episodes of recent gene flow, such as Bantu expansion in Africa, for example, have left well recognizable fingerprints in the genetic composition of both marker systems in African populations (21, 63, 91)
5. Admixture and clines rather than abrupt changes over ethnic boundaries can be observed in regions such as North Africa, and Central and West Asia, which lie in between two or more distinctive pools of mtDNA and Y chromosome varieties (5, 17, 76, 121).

In addition to these general patterns of consistency between the loci, several minor regional differences can be noted:

1. In contrast to the overall homogeneity of mtDNA haplogroup composition in Europe, there are remarkable differences between West and East Europe, for example, in the frequencies of Y chromosome haplogroups R1b, I1a, and I1b (85, 89, 93).
2. East European, and Central and East Asian populations share common Y chromosome genetic components, such as haplogroups N and R1a, which are not recapitulated in mtDNA phylogeography (86).
3. Y chromosomes of the Austronesian-speaking populations do not testify to a well-pronounced founder effect as evidenced from mtDNA data (44, 112).
4. Asymmetric gene flow is detectable between hunter-gatherer and agriculturalist societies in Africa (24).
5. Y chromosome data show a signal for a separate late-Pleistocene migration from Africa to Europe over Sinai as evidenced through the distribution of haplogroup E3b lineages (20, 63), which is not manifested in mtDNA haplogroup distributions.

SPECIATION AND OUT-OF-AFRICA BOTTLENECKS

Palaeontological and molecular genetic evidence continues to accumulate indicating that multiple range expansions by anatomically modern humans leading to Eurasia

and beyond from an African homeland were accomplished within a rather rapid interval followed by subsequent fragmentation and genetic as well as cultural isolation (43, 59, 64, 68, 100). These independent haploid loci with their lower effective population size capture episodes of rapid population divergence better than the autosomes (48), making them the systems of choice for tracing recent migratory events. The variation outside Africa represents only a small subset of African variation (**Figures 2**, **3**) consistent with the out-of-Africa bottleneck hypothesis (25), which is supported now by a substantial body of evidence also from the nuclear genome [for reviews see (30, 66, 109)].

No matter how well resolved the splits in the Y chromosome or mtDNA tree, both remain single-locus trees and are thus subject to large stochastic errors for certain parameters' estimations, including the overall tree depth. Low levels of genetic diversity of mtDNA and Y chromosome show coalescent dates of 100,000–200,000 years (45, 53, 70, 80, 108). These dates have often been implicated in favor of a speciation bottleneck coinciding with the time approximately when anatomically modern humans start showing up in the fossil record (122). Sequence data from the autosomal compartment of the genome have failed to support such an hourglass model and instead continue to provide evidence for a long-lasting low effective population size over the Pleistocene, or the long-neck model (30, 34, 40). Although mtDNA and Y chromosome cannot be considered informative with regards to the speciation event of modern humans, they continue to be the most well resolved genetic loci for the study of population histories since the out-of-Africa migration.

STEPPED CLINAL HAPLOGROUP PROGRESSION

A recent study of L1 and Alu insertion polymorphisms to analyze human population structure in geographic space concluded that human diversity is a combination of clines and clusters and introduced the term "stepped clinal" to describe its composite pattern (123). The unification of both the relatively remote and nearby ancestral relationships inherent in haploid phylogenies resembles the stepped clinal characterization revealed by these autosomal data.

The properties of robustly resolved phylogenies make it possible to track the geographic progression of haplogroup differentiation, the polarity of which can be inferred from the series of increasingly derived character states over a geographic line (6). The inferences from such haplogroup progression patterns are not directly dependent on the frequency of the haplogroups considered. However, the spotting of such lineages over space certainly is frequency dependent inasmuch as genetic drift affects the fate of any genetic variant and the inference depends upon many assumptions of the demographic history of the region and levels of gene flow [discussed in detail in (82)]. **Figure 5** illustrates the principle of coprogression on the example of mtDNA haplogroup B4 and Y chromosome haplogroup Q. As shown in **Figures 2** and **3**, the geographic center of gravity of both of these haplogroups is in East Asia. The progression of haplogroup diversification for both haploid genomes is representative of range expansion events toward different destinations from a region of common provenance that likely were on a population scale rather than reflective of a stochastic accident of a single gene tree. The biogeography of mtDNA haplogroup B4 illustrates the principle of vicariance in which stochastic events (drift) followed by population fragmentation often result in the contrasting geographic distribution of related sub-haplogroups with geography. Thus both North American and Polynesian populations trace their descent to molecular haplogroup B4 ancestors in East Asia. Likewise, the progression of Y chromosome haplogroup Q reveals the directionality of the movement of males from East Asia to North America and subsequently to Central America consistent with the serial founder effect (78).

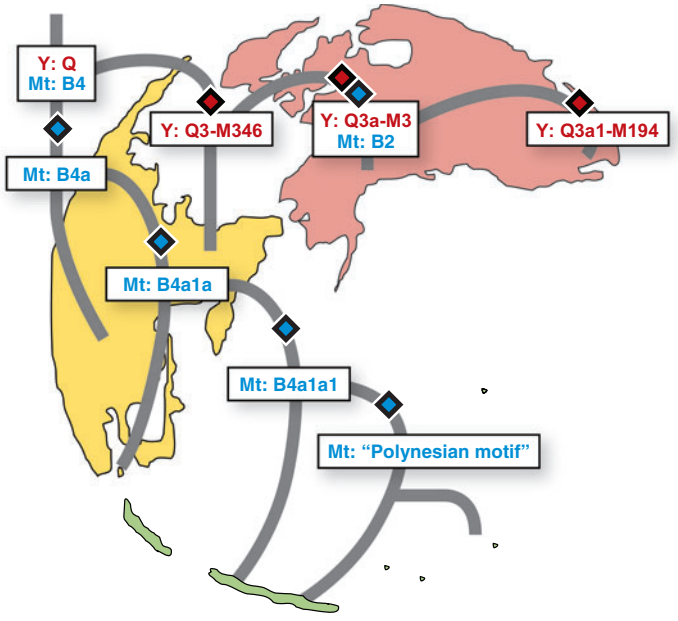


Figure 5

Phylogeographic progression of mtDNA haplogroup B4 and Y chromosome haplogroup Q. The sequential geographic trajectory of mtDNA and Y chromosome subhaplogroup formation is shown by horizontal lines indicating the geographic progression of respective founder lineages in space. Vertical lines denote descent with modification. Mutations occurring in mtDNA are shown with blue and mutations in Y chromosome with red diamonds. Geography: Asia/Siberia (*yellow*); Canada/Americas (*red*); Insular southeast Asia/Polynesia (*green*). The general phylogenetic context of mtDNA haplogroup B4 and Y chromosome haplogroup Q are given in **Figures 2** and **3**, respectively.

WHY DO WE NEED MORE MARKERS?

The specific features of mitochondrial DNA and Y chromosome are the lack of recombination and uniparental inheritance (32, 47), which provide that straightforward genealogical histories can be inferred at maximum molecular resolution as a function only of sequence length examined. Variation in nuclear genes is reshuffled by recombination and therefore tree building from long sequence stretches is complicated. The molecular resolution in the framework of phylogenetics is the measure of informative (variable) positions that can be revealed in contiguous sequence fragments. The second important measure in phylogenetics is the robustness of a branch in a tree, which is a combined measure of the number of independent characters supporting the branch and the average number of their recurrent evolution on the tree. In the nuclear genome, the sequence lengths per gene that have been examined so far through an evolutionary framework have been mostly within the range of 2–10 kb, with the outcome that less than 10 substitutions, normally, per lineage per one million years have been attained. Even though the trees can be fairly robust, provided that mutation and recombination rates are low, the low molecular resolution of such trees allows for a wide range of interpretations, including those discussed in favor of the multiregional model (103). Low resolution leaves us with poor understanding of our genetic history during the pivotal past 100,000 years, which is the time window of interest for most of the human migrations. In contrast, the nonrecombining parts of the Y

chromosome and the completed 16,569 base pairs of highly variable mitochondrial genome have both supplied us with information on the geneaological accumulation of more than 20 substitutions per lineage during the past ~200,000 years. Much more information is yet to be revealed for the NRY, since less than 1% of its sequence has been ascertained so far for the width of global population variation. The efficacy of increased power of molecular resolution in phylogenetic inferences is illustrated in **Figure 6** on the example of mtDNA haplogroup M lineages in South and East Asia. Hypervariable region sequences do not provide sufficient resolution to distinguish between Indian and Chinese varieties of haplogroup M, and occasionally the phylogenetic associations the short hypervariable sequence stretches provide can even be spurious, as, for example, the link of B156 and QD8166 supported by HVS-I motif of two substitutions. The level of phylogenetic ambiguity of the HVS-I is expressed in the complexity of resulting networks, whereas by using

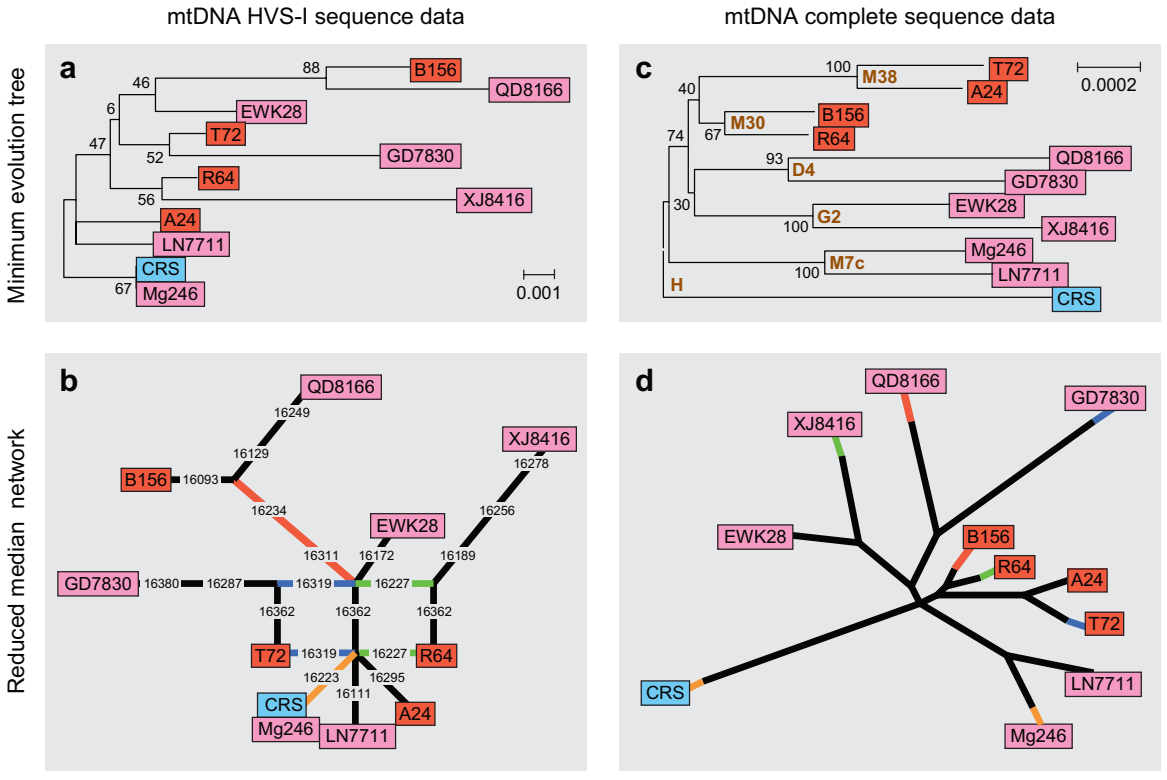


Figure 6

Resolving parallel evolution of mtDNA hypervariable characters by using sequence data from the linked coding regions. Branches defined by conflicting HVS-I characters whose nucleotide position is specified in panel B are highlighted with color to show their position in the resolved trees based on complete sequence data (panel *d*). Haplogroup affiliations of respective sequences are shown in panel *c*. Sequence data for samples shown on pink background are taken from Kong et al. (57), sequences shown on red background from Sun et al. (102), and rCRS shown on blue background refers to the revised Cambridge Reference Sequence (3). Sequences were processed in phylogenetic packages MEGA (**http://www.megasoftware.net/**; panels *a* and *c*) using Minimum Evolution method with 100 bootstrap replicas, and in NETWORK (**http://www.fluxus-engineering.com/**; panels *b* and *d*) using reduced median algorithm (7).

complete sequence information, the network methods tend to produce more fully resolved and robust tree-like structures (41). As illustrated in **Figure 6** panels *c* and *d*, it appears less important which phylogenetic tree-building algorithms are being used when the tree structure is robustly supported by the data.

However, certain inadequacies in the phylogenies may remain, both at internal and external branches, that reflect consequences of ascertainment as well as experimental effort. Although complete sequencing of mtDNA genomes in worldwide populations initiated by Ingman et al. (45) and expanded upon by others [reviewed by (110)] has greatly improved the phylogeny, the innately relatively small size of the mitochondrial genome limits the capacity for retention of a more complete record of prehistoric cladistic relationships. Thus many mtDNA clades are paraphyletic, in which numerous haplogroups radiate from a common node and their precise cladistic relationships remain uncertain. Such paraphyletic limitations also characterize the current state-of-the-art Y chromosome phylogeny. However, the much larger physical size of the NRY genome, although less practical to survey than mtDNA, offers the potentiality that more ancient drift events will be recoverable in the form of structurally deep intermediate binary markers that will unify some currently paraphyletic clades. These anticipated NRY bifurcations may ultimately illuminate a more complete cladistic branching order and improve the usefulness of this locus in retrieving signals of genetic affinities. An example of such ancient relationship is NRY marker M429 (126) that now unifies haplogroups I and J (**Figure 3**).

RATES OF MUTATION, FIXATION, AND HOMOPLASY

Substitution rate in the nuclear genome is generally low as compared to mtDNA. On average, the probability of observing a substitution at a nucleotide pair in one generation is considered to be $3–5 \times 10^{-8}$ (18, 108) for the Y chromosome. The Y chromosome appears as a striking outlier in comparison with other chromosomes, showing the highest divergence rate of 1.9% from the chimpanzee as compared to only 0.94 of X chromosome (18). The higher male germline mutation rate might be explained by its lower effective population size, whereby slightly deleterious mutations would appear as neutral more often than they would in autosomal genes and thereby have a higher chance of fixation. Alternatively, the difference might be due to the fact that there are 5–6 times more cell divisions in the male germline and therefore more mutations resulting from DNA damage such as deamination of methyl CpG to TpG (61). Compared to mitochondrial DNA, the still conservative-enough substitution rate of Y provides that the trees inferred from Y chromosome SNP (single nucleotide polymorphism) data are fairly robust (or "bullet-proof") and show a low level of homoplasy (47, 116). Gene conversion rather than multiple hits at the SNP position may be the main mechanism of parallel mutations within multicopy sequences in the Y chromosome, as shown for the substitution P25 (2). Similarly, large-scale insertions and deletions occur more often recurrently and cannot thus be used as stable markers for phylogenetic inferences (80).

Due to its high mutation rate, which can substantially vary over sites, mtDNA variation is characterized by excessively high levels of intraspecies homoplasy when compared to the variation in nuclear genes. The site-specific mutation rate varies in different mammalian groups, so that the same sites might not be hypermutable in distinct lineages (29). Homoplasy introduces complications to tree building and is most problematic in data sets of hypervariable region sequences, as illustrated in **Figure 6**. A number of coding region mutational hotspots have been identified in human populations as well; nevertheless, the phylogenetic support for the general architecture and for most of the internal branches of human mtDNA phylogeny is

robust when sequence information from the whole molecule is taken into account (53).

SIGNIFICANCE OF ASCERTAINMENT

Any description of genetic diversity in populations is influenced by the ascertainment status of the polymorphisms that are being used (71, 84). Effective population size modulates levels of genetic variation, genetic drift, and linkage disequilibrium in populations (104). The lower effective population size of haploid genomes relative to the other constituents of the genome accentuates the significance of ascertainment as haploid genomes show (e.g., Fst) more rapid between-population divergence rates (48). The sensitivity of diversity within sample ascertainment panels used to discover haploid polymorphisms therefore profoundly influences the detection of phylogenetically informative markers. Although the NIH dbSNP database presently catalogs several thousands of Y chromosome SNPs contributed by various resequencing projects, despite some exceptions (80, 95), most were ascertained in panels only partially reflective of the known spectrum of haplogroup diversity. Also, the bulk of the Y chromosome SNPs in public databases remains undercharacterized with regard to their Y chromosome specificity, frequency, and phylogenetic relationships. The significance of ascertainment is illustrated concerning Y SNPs reported as part of a genome-SNP discovery effort (42). Of the 18 major Y chromosome haplogroups described, the ascertainment panel used was composed of only haplogroup E, I, J, O, and R (**Figure 7*a***) members composed of 12, 4, 1, 8, and 8 individuals each, respectively. This partiality in the ascertainment panel resulted in an abbreviated phylogeny (**Figure 7*b***) constructed from 295 phylogenetically consistent markers that underrepresented the known tree structure (**Figure 3**) while leading to an excessive number of redundant characters reinforcing the same branches (**Figure 7*b***). Nonetheless, beside fractionating some haplogroups (98), the 26 polymorphisms that define the E branch in **Figure 6*b*** also provide an opportunity to explore the primary deep-rooted relationships of haplogroups C, DE, and F.

BIFURCATIONS AND MIGRATION MODELS: A CASE STUDY

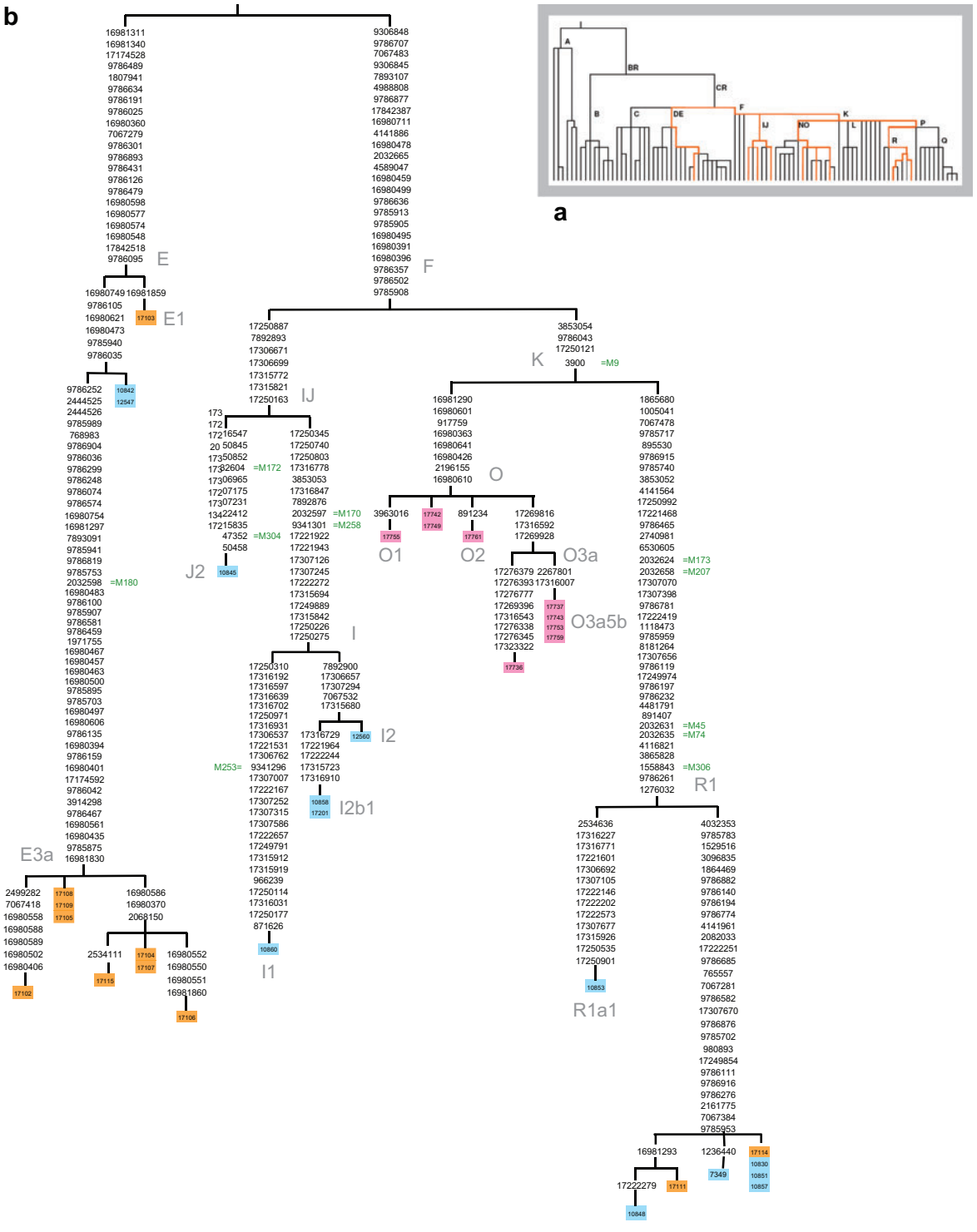
With the exception of African-specific haplogroups A and B, all other Y chromosome haplogroups descend from one ancestral node of the tree termed CDEF, which is defined by mutations M168 and M294 (**Figure 8**). This previously unresolved trifurcation of this node into haplogroups C, DE, and F comprises the majority of African- and all non-African-affiliated chromosomes (**Figure 8*a***). There are three possible solutions to this tripartite structure as presented in **Figure 8*b*–*d***. Using the principle of phylogeographic

*Erratum

Figure 7

Improved resolution of Y chromosome phylogeny in a small ascertainment panel. (*a*) Established Y chromosome haplogroup structure as shown in **Figure 3**. The phylogenetic affiliation of the 33 male samples used in the ascertainment panel of the study by Hinds et al. (42) is indicated by orange lines. See text for more details. (*b*) The single most parsimonious tree relating 295 Y chromosome markers assessed in these 33 individuals. The tree is constructed using reduced median algorithm (7). The original data set (42) including 334 Y chromosome markers was subjected to network analysis. The total of 39 markers were found to be in character conflict with known Y chromosome markers (identified in this figure in *blue*), and as they were ascertained within sequences showing high homology to human X chromosome, we excluded them from further analyses. The reference sequence (rs#) numbers of the markers defining the tree structure are shown along the branches. Sample codes are as given in the original study. Samples of African-American origin are shown on orange, those of Asian origin on pink, and those of European-American origin on blue background.

*This PDF ammended on (8 Jan. 2008): See explanation at http://arjournals.annualreviews.org/errata/genet



parsimony, which minimizes the number of inferred migrations and the fact that the deepest clades (A and B) occur solely in Africans, an African origin of haplogroup CDEF-M168, M294 node was proposed (39, 116) and supported in a survey of over 12,000 Asian men (50). Although the initial proposal (38) of an Asian origin of haplogroup DE was first neutralized by the recognition of the haplogroup D-M174 (115) and further eroded by the detection of DE* chromosomes in Nigeria (120), the previous inability to resolve the earlier tripartite structure left an element of uncertainty because the Asian origin of haplogroup DE could be resurrected using the same principle of parsimony [e.g., consider the parallel example of catarrhine evolution (99)] if the trifurcation were resolved in favor of a common ancestor of haplogroups DE and F (**Figure 8*b***). Such an ancestral node would imply that DE is a subset of Eurasian variation and therefore the African YAP (Y-chromosome Alu polymorphism) chromosomes could be considered as due to a back-migration from Asia. Second, if haplogroups C and F were to share a common recent ancestor apart from the DE clade (**Figure 8*d***), the distribution of Y chromosome haplogroup D in Asia could be explained by an evolutionary history separate from that of the other two clades. Haplogroup D is particularly enigmatic because of its widely separated disjunctive distribution in Asia suggestive of an ancient (perhaps independent) range expansion to Asia followed by fragmentation and considerable isolation. The absence of haplogroup D in Oceania and its relic peripheral distribution in Asia is in contrast to that observed for haplogroup C and F chromosomes.

We resolved this discrepancy by using improved phylogenetic resolution in the Y chromosome phylogeny. This was achieved by leveraging knowledge contained in some of the phylogenetically consistent Y chromosome SNPs reported by Hinds et al. (42) (**Figure 7*b***). By experimentally haplogrouping the same 33 males that were used to ascertain these Y chromosome polymorphisms, it was possible to infer that 22 of the SNPs were derived in all haplogroup E chromosomes and 24 in all F chromosomes (125) when the individuals in the ascertainment panel were subjected to phylogenetic analysis. Since haplogroup E and F chromosomes were present in the ascertainment panel but haplogroup C and D representatives were not, the possibility existed that some of these 46 SNPs might be positioned upstream of either the E or F node in the phylogeny. A total of 18 of these were designed as successful PCR- (polymerase chain reaction) based assays and genotyped by DHPLC (denaturing high-performance liquid chromatography) (113) in samples belonging to haplogroups A, B, C-M216, D-M174, E-M96, and F-M89. The results of these haplogrouping experiments indicated that one (**Table 1**) of the 18 SNPs evaluated shared derived alleles in haplogroups C and F while being at an ancestral state in the other haplogroups. These results hold up the phylogenetic senario shown in **Figure 8*d***, which is consistent with two independent founder types, D and CF, evolving outside Africa, and thus weakens the other two possible interpretations discussed above. However, the common ancestry of C and F founder types is supported by a short branch, defined by a single mutation, implying the diversification of CF from DE was shortly followed by the split of C from F. Although extinction events within Africa offset by haplogroup survival

*Erratum

Table 1 Primers and specifications of haplogroup CF related node Y chromosome marker

rs No.	Nucleotide change	Amplicon size (bp)	SNP Position from 5′ end	Primer forward 5′–3′	Primer reverse 5′–3′
4141886	G to A	344	216	cctgaggagacatagccata	atagctagattctggtcccg

*This PDF ammended on (8 Jan. 2008): See explanation at http://arjournals.annualreviews.org/errata/genet

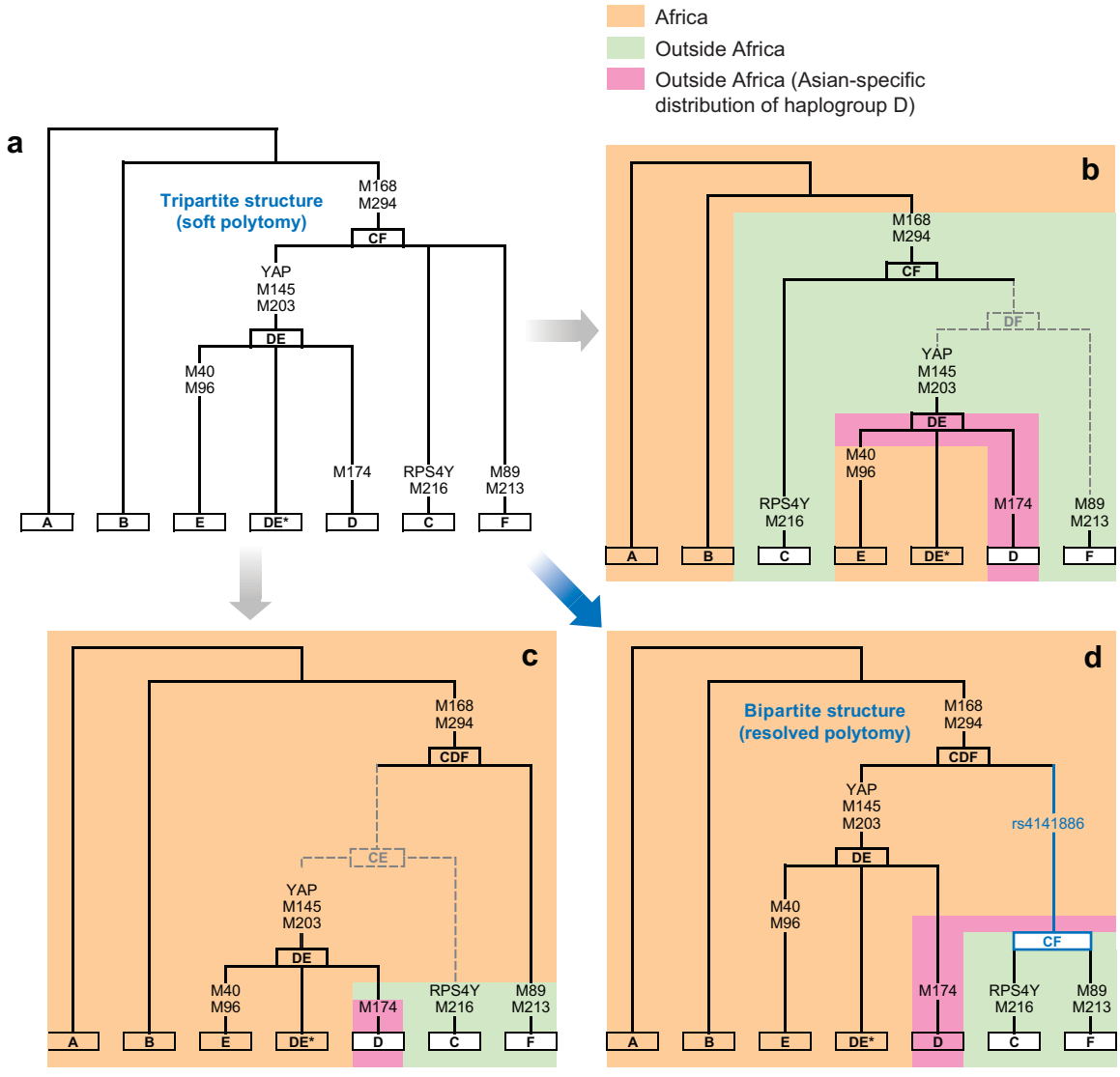


Figure 8

Improved phylogenetic resolution of the Y chromosome tree. (*a*) Tripartite structure of the YCC 2003 tree (47), which allows for an interpretation (*b*) that one founder haplogroup CDF colonized Eurasia and DE was brought back to Africa, or (*c*) a model of one migration from Africa involving three founder haplogroups (C, D, and F), or a model (*d*) implying that the strictly Asian-specific distribution of haplogroup D (highlighted in *pink*) reflects a relic distribution of an early out-of-Africa settlement of Eurasia that was overwhelmed by a separate, demographically successful migration from Africa marked by a putative of CF clade, derived lineages of which are now widely spread from Europe to Australia. A new SNP marker from the study by Hinds et al. (42) (see **Figure 7**; **Table 1**) provides a resolved bipartite structure that supports this scenario and thus weakens interpretations *b* and *c*. See text for further details. The structure of the resolved tree, however, is not informative on whether the descendants of C and F evolved and subsequently went extinct in Africa, or whether the C and FR clades emerged outside Africa.

AZF: Azoospermia Factor

of descendents in Asia cannot be empirically demonstrated, both the refutation of the option shown in **Figure 8*b*** and the apparent absence of deep-rooted haplogroups for either CF or D chromosomes in Africa bolsters the model that haplogroup CF and DE molecular ancestors first evolved inside Africa and subsequently contributed as Y chromosome founders to pioneering migrations that successfully colonized Asia. While not proof, the DE and CF bifurcation (**Figure 8*d***) is consistent with independent colonization impulses possibly occurring in a short time interval.

Although haplogroups D and E share common ancestry, a geographic gap exists between the frequent occurrence of haplogroup E in Africa and the relic distribution of D in Asia, suggestive of long-term isolation and extinction of descendents in the geographic intermediary zone to Asia. Although it is difficult to distinguish the influence of positive selection from demographic expansion, this anomaly in distribution could also be explained, in part, by negative selection affecting haplogroup D carriers. A slightly deleterious mutation may become fixed in some peripheral populations after its increase in frequency while surfing the wave of population advancement (26, 55). A possible candidate locus to be considered in the Y chromosome is copy-number variation in the aZFc gene as a possible haplogroup-associated risk factor regarding male fertility (47). Although such partial deletions of <2MB that occur across the spectrum of Y chromosome haplogroups (79, 80) are indicative of parallel mutations, this feature is significantly common in some Y chromosome haplogroup backgrounds (4, 127), especially those that were successful in recent demographic expansions such as haplogroup N (81, 86) in northern Asia. Individuals with haplogroup D affiliation have been reported to have such partial deletions in association with lower sperm concentrations in Japan (58). This partial deletion may in some measure explain the absence of haplogroup D in India, a zone implicated in the southern coastal migratory route, whereas haplogroup D is fixed in some tribal populations of the Andaman Islands (106). However, not all haplogroup D carriers have been reported to have such partial deletions (22). Nonetheless, the question remains to what extent, if any, susceptibility to a potential reproductive liability has influenced the phylogeography of D and other haplogroups in the Y chromosome and mtDNA phylogenies. Imbalance of the proportions of nonsynonymous mutations among and within the old and young clades of the mtDNA tree (53) further suggests that the outcome of some puzzling phenomena in uniparental haplogroup distributions might be the result of intertwining factors such as drift, selection, founder effect, and migration.

SUMMARY POINTS

1. Broad-spectrum features of global population structure as deduced from phylogenetic analysis of mtDNA and Y chromosome markers are generally consistent with inferences based on analyses of multiple independent autosomal loci. In addition, a phylogenetic approach allows us to get relative estimates of temporality and polarity for the structural differentiation among the populations.
2. The mtDNA and Y chromosome trees coalesce at shallow time depth and are uninformative about the speciation event leading to our species. Instead, at high molecular resolution, these loci provide detailed information about both the out-of-Africa migration and further population differentiation, at substantially finer detail than the recombining domains of our genome.
3. Maternal and paternal phylogenies display signals of both discreteness and continuity in respect to the geographic differentiation and distribution of their hierarchic

component haplogroups. Three founder haplogroups in both trees describe the non-African variation as a small subset of African genetic diversity. Discrete subclades of these founders show frequency patterns characteristic of the regions of their origin, whereas clinal blending of their frequency occurs across their geographic boundaries.

4. The sample composition of ascertainment panels used to discover haploid polymorphisms profoundly influences the detection of phylogenetically informative markers. Although many thousands of Y chromosome markers have been described to date, only a fraction appear informative in populations. Different world populations used for ascertainment are often strongly biased in this respect.

FUTURE ISSUES

1. Identification of additional informative markers by considering ascertainment criteria is required to improve the haploid phylogenetic structures of human populations.
2. To differentiate demographic history from locus-specific natural selection, comparisons are needed between the haploid genomes and genetic systems such as the X-chromosome and autosomes within the same population surveys.
3. Exposure of regional population structure and inference of underlying histories requires more intensive sampling.
4. Improvements in the empirical data sets must be matched with equivalent improvements in modeling demographic histories by analytical methods and computational simulations.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Rob Foley for his valuable criticism and comments on the manuscript. We also thank Professor L. Luca Cavalli-Sforza for his thoughtful comments. We express appreciation to Ms Alice A. Lin for her technical assistance towards resolving the deep trifurcation in the Y chromosome tree.

LITERATURE CITED

1. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* 75:910–18
2. Adams SM, King TE, Bosch E, Jobling MA. 2006. The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. *Forensic Sci. Int.* 159:14–20
3. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23:147

4. Arredi B, Ferlin A, Speltra E, Bedin C, Zuccarello D, et al. 2007. Y-chromosome haplogroups and susceptibility to azoospermia factor C microdeletion in an Italian population. *J. Med. Genet.* 44:205–8
5. Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, et al. 2004. A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am. J. Hum. Genet.* 75:338–45
6. Ashlock P. 1974. The use of cladistics. *Annu. Rev. Ecol. Syst.* 5:81–99
7. Bandelt H-J, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–53
8. Bolnick DA, Bolnick DI, Smith DG. 2006. Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol. Biol. Evol.* 23:2161–74
9. Brandstatter A, Salas A, Niederstatter H, Gassner C, Carracedo A, Parson W. 2006. Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27:2541–50
10. Brown WM. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA* 77:3605–9
11. Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36
12. Carvajal-Carmona LG, Soto JD, Pineda N, Ortziz-Barrientos D, Duque C, et al. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.* 67:1287–95
13. Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, et al. 1985. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230:1403–6
14. Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33(Suppl.):266–75
15. Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* 57:133–49
16. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–48
17. Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, et al. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* 12:495–504
18. Consortium CSaA. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87
19. Consortium TIH. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
20. Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am. J. Hum. Genet.* 74:1014–22
21. Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* 70:1197–214
22. de Carvalho CM, Zuccherato LW, Fujisawa M, Shirakawa T, Ribeiro-dos-Santos AK, et al. 2006. Study of AZFc partial deletion gr/gr in fertile and infertile Japanese males. *J. Hum. Genet.* 51:794–99
23. 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.* 67:1055–61

24. Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, et al. 2004. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol.* 21:1673–82
25. Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1597–601
26. Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. *Proc. Natl. Acad. Sci. USA* 101:975–79
27. Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, et al. 2005. Expanding southwest pacific mitochondrial haplogroups P and Q. *Mol. Biol. Evol.* 22:1506–17
28. Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, et al. 2007. Melanesian mtDNA complexity. *PLoS ONE* 2:e248
29. Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. 2006. Mutation hot spots in mammalian mitochondrial DNA. *Genome Res.* 16:215–22
30. Garrigan D, Hammer MF. 2006. Reconstructing human origins in the genomic era. *Nat. Rev. Genet.* 7:669–80
31. Garrigan D, Mobasher Z, Kingan SB, Wilder JA, Hammer MF. 2005. Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics* 170:1849–56
32. Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 77:6715–19
33. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* 24:757–68
34. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444:330–36
35. Guillot G, Mortimer F, Estoup A. 2005. Geneland: a computer package for landscape genetics. *Mol. Ecol. Notes* 5:712–15
36. 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.* 11:749–61
37. Hammer MF, Garrigan D, Wood E, Wilder JA, Mobasher Z, et al. 2004. Heterogeneous patterns of variation among multiple human x-linked loci: the possible role of diversity-reducing selection in non-Africans. *Genetics* 167:1841–53
38. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, et al. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* 15:427–41
39. Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, et al. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* 18:1189–203
40. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* 95:1961–67
41. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, et al. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* 70:1152–71
42. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, et al. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* 307:1072–79
43. Hudjashov G, Kivisild T, Underhill P, Endicott P, Sanchez JJ, et al. 2007. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci. USA* 104:8726–30
44. Hurles ME, Matisoo-Smith E, GR D, Penny D. 2003. Untangling Oceanic settlement: the edge of the knowable. *Trends Ecol. Evol.* 18:531–40

45. Ingman M, Gyllensten U. 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J. Hered.* 92:454–61
46. Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, et al. 1996. Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum. Mol. Genet.* 5:1767–75
47. Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 4:598–612
48. Jorde LB, Bamshad M, Rogers AR. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *BioEssays* 20:126–36
49. Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, et al. 1997. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* 110:125–33
50. Ke Y, Su B, Song X, Lu D, Chen L, et al. 2001. African origin of modern humans in East Asia: a tale of 12000 Y chromosomes. *Science* 292:1151–53
51. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, et al. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the Gate of Tears. *Am. J. Hum. Genet.* 75:752–70
52. Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72:313–32
53. Kivisild T, Shen P, Wall DP, Do B, Sung R, et al. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–87
54. Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, et al. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* 19:1737–51. Erratum. 2003. *Mol. Biol. Evol.* 20:162
55. Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.* 23:482–90
56. Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, et al. 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum. Mol. Genet.* 15:2076–86
57. Kong QP, Yao YG, Sun C, Zhu CL, Zhong L, et al. 2004. Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. *J. Hum. Genet.* 49:414–23
58. Kuroki Y, Iwamoto T, Lee J, Yoshiike M, Nozawa S, et al. 1999. Spermatogenic ability is different among males in different Y chromosome lineage. *J. Hum. Genet.* 44:289–92
59. Lahr M, Foley R. 1994. Multiple dispersals and modern human origins. *Evol. Anthropol.* 3:48–60
60. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
61. Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Curr. Opin. Genet. Dev.* 12:650–56
62. Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, et al. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* 21:2012–21
63. Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, et al. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* 74:532–44
64. Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–36

65. Manica A, Prugnolle F, Balloux F. 2005. Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* 118:366–71
66. Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, et al. 2003. Sequence variations in the public human genome data reflect a bottlenecked population history. *Proc. Natl. Acad. Sci. USA* 100:376–81
67. Mathias N, Bayes M, Tyler-Smith C. 1994. Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.* 3:115–23
68. Mellars P. 2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313:796–800
69. Merriwether DA, Hodgson JA, Friedlaender FR, Allaby R, Cerchio S, et al. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl. Acad. Sci. USA* 102:13034–39
70. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* 100:171–76
71. Mountain JL, Cavalli-Sforza LL. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. *Proc. Natl. Acad. Sci. USA* 91:6515–19
72. Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218
73. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* 29:20–21
74. Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, et al. 2004. Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. *Am. J. Hum. Genet.* 75:966–78
75. Pereira L, Richards M, Goios A, Alonso A, Albarran C, et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15:19–24
76. Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, et al. 2004. Where West meets East: the complex mtDNA landscape of the Southwest and Central Asian corridor. *Am. J. Hum. Genet.* 74:827–45
77. Quintana-Murci L, Semino O, Bandelt H-J, 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.* 23:437–41
78. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. USA* 102:15942–47
79. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, et al. 2003. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat. Genet.* 35:247–51
80. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, et al. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat. Genet.* 38:463–67
81. Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, et al. 2004. A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor C region. *Genomics* 83:1046–52
82. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1251–76

83. Richards MB, Bandelt H-J, Kivisild T, Oppenheimer S. 2006. A model for the dispersal of modern humans out of Africa. In *Human Mitochondrial DNA and the Evolution of Homo sapiens*, ed. H-J Bandelt, V Macaulay, M Richards, pp. 225–65. Berlin, New York: Springer
84. Rogers AR, Jorde LB. 1996. Ascertainment bias in estimates of average heterozygosity. *Am. J. Hum. Genet.* 58:1033–41
85. Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* 75:128–37
86. Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, et al. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* 15:204–11
87. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* 1:e70
88. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. 2002. Genetic structure of human populations. *Science* 298:2381–85
89. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* 67:1526–43
90. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, et al. 2006. Positive natural selection in the human lineage. *Science* 312:1614–20
91. Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, et al. 2002. The making of the African mtDNA landscape. *Am. J. Hum. Genet.* 71:1082–111
92. Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, et al. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* 74:1023–34
93. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–59
94. Serre D, Pääbo S. 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* 14:1679–85
95. Shen P, Wang F, Underhill PA, Franco C, Yang WH, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* 97:7354–59
96. Shimada MK, Panchapakesan K, Tishkoff SA, Nato AQ Jr, Hey J. 2007. Divergent haplotypes and human history as revealed in a worldwide survey of x-linked DNA sequence variation. *Mol. Biol. Evol.* 24:687–98
97. Shriver MD, Kittles RA. 2004. Genetic ancestry and the search for personalized genetic histories. *Nat. Rev. Genet.* 5:611–18
98. Sims LM, Garvey D, Ballantyne J. 2007. Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum. Mutat.* 28:97
99. Stewart CB, Disotell TR. 1998. Primate evolution—in and out of Africa. *Curr. Biol.* 8:R582–88
100. Stringer C. 2002. Modern human origins: progress and prospects. *Philos. Trans. R. Soc. London Ser. B* 357:563–79
101. Stringer CB, Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–68

102. Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, et al. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.* 23:683–90
103. Templeton A. 2002. Out of Africa again and again. *Nature* 416:45–51
104. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, et al. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17:520–26
105. Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, et al. 2005. Reconstructing the origin of Andaman Islanders. *Science* 308:996
106. Thangaraj K, Singh L, Reddy A, Rao V, Sehgal S, et al. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* 13:86–93
107. Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, et al. 2005. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum. Genet.* 116:507–17
108. 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* 97:7360–65
109. Tishkoff SA, Kidd KK. 2004. Implications of biogeography of human populations for 'race' and medicine. *Nat. Genet.* 36:S21–27
110. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* 22:339–45
111. Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, et al. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591–608
112. Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, et al. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol.* 3:e247
113. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, et al. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 7:996–1005
114. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, et al. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65:43–62
115. Underhill PA, Roseman CC. 2000. *The case for an African rather than an Asian origin of the human Y-chromosome YAP insertion*. Presented at Recent Adv. Hum. Biol., Yunnan Univ., China
116. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26:358–61
117. van Holst Pellekaan SM, Ingman M, Roberts-Thomson J, Harding RM. 2006. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *Am. J. Phys. Anthropol.* 131:282–94
118. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
119. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–7
120. Weale ME, Shah T, Jones AL, Greenhalgh J, Wilson JF, et al. 2003. Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. *Genetics* 165:229–34

121. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, et al. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98:10244–49
122. White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, et al. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* 423:742–47
123. Witherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, et al. 2006. Human population genetic structure and diversity inferred from polymorphic L1(LINE-1) and Alu insertions. *Hum. Hered.* 62:30–46
124. Wright S. 1949. Population structure in evolution. *Proc. Am. Philos. Soc.* 93:471–78
125. YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–48
126. Underhill PA, Myres NM, Rootsi S, Chow C-ET, Lin AA, et al. 2007. Chapter 3. New phylogenetic relationships for Y-chromosome haplogroup I: Reappraising its phylogeography and prehistory. In *Rethinking the Human Revolution*, ed. P Mellars, K Boyle, O Bar-Yosef, C Stringer, pp. 33–42. Cambridge: McDonald Inst. Archaeol. Res.
127. Zhang F, Lu C, Li Z, Xie P, Xia Y, et al. 2007. Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. *J. Med. Genet.* 44:437–44