Haplotype Trees and Modern Human Origins

Alan R. Templeton*

Department of Biology, Washington University, St. Louis, Missouri 63130-4899

ABSTRACT   A haplotype is a multisite haploid genotype at two or more polymorphic sites on the same chromosome in a defined DNA region. An evolutionary tree of the haplotypes can be estimated if the DNA region had little to no recombination. Haplotype trees can be used to reconstruct past human gene-flow patterns and historical events, but any single tree captures only a small portion of evolutionary history, and is subject to error. A fuller view of human evolution requires multiple DNA regions, and errors can be minimized by cross-validating inferences across loci. An analysis of 25 DNA regions reveals an out-of-Africa expansion event at 1.9 million years ago. Gene flow with isolation by distance was established between African and Eurasian populations by about 1.5 million years ago, with no detectable interruptions since. A second out-of-Africa expansion occurred about 700,000 years ago, and involved interbreeding with at least some Eurasian populations. A third out-of-Africa event occurred around 100,000 years ago, and was also characterized by interbreeding, with the hypothesis of a total Eurasian replacement strongly rejected ($P < 10^{-17}$). This does not preclude the possibility that some Eurasian populations could have been replaced, and the status of Neanderthals is indecisive. Demographic inferences from haplotype trees have been inconsistent, so few definitive conclusions can be made at this time. Haplotype trees from human parasites offer additional insights into human evolution and raise the possibility of an Asian isolate of humanity, but once again not in a definitive fashion. Haplotype trees can also indicate which genes were subject to positive selection in the lineage leading to modern humans. Genetics provides many insights into human evolution, but those insights need to be integrated with fossil and archaeological data to yield a fuller picture of the origin of modern humans. Yrbk Phys Anthropol 48:33–59, 2005.   ©2005 Wiley-Liss, Inc.

INTRODUCTION

Evolutionary trees of genetic variation have played a prominent role in studies of human evolution ever since the publication of Cann et al. (1987) and the popularization of "mitochondrial Eve." There are many methods of measuring genetic variation, but the focus of this paper is upon haplotypes. A haplotype is the multisite haploid genotype at two or more polymorphic sites on the same chromosome in a defined DNA region. The polymorphic sites that determine the haplotype state could be single-nucleotide polymorphisms (SNPs), insertion and deletion polymorphisms (indels), or any other polymorphisms scored on the same chromosome in the DNA region of interest. The defined DNA region could be an entire genome, such as the entire mitochondrial genome, or it could be a small portion of a large chromosome. All homologous copies of the DNA molecule in such a region that share the same genetic state at all scored polymorphic sites are members of the same haplotype.

Haplotypes are simple to obtain for haploid (or effectively haploid) DNA regions such as mitochondrial DNA (mtDNA), most Y-chromosomal DNA (Y-DNA), or X-chromosomal DNA scored in males. Haplotypes are more difficult to obtain for autosomal regions, because the phase at two or more heterozygous sites is often not observable with current DNA sequencing technology. For example, suppose an individual is heterozygous for two SNPs in an autosomal region. A standard DNA sequencing gel displays heterozygous sites as double-nucleotide scores, say, nucleotides A and T at one SNP and G and C at the second SNP, but does *not* indicate if the A at SNP 1 is located on the same DNA molecule as the G or the C at SNP 2. As a result, a double-heterozygote individual such as A/T, G/C could either have the haplotype pair AG and TC (where underlining indicates the pair of nucleotides found on a single DNA molecule), or AC and TG. Additional molecular and statistical techniques are needed to obtain autosomal haplotypes in cases such as this, although the cost and effort can sometimes be considerable (Templeton et al., 1988; Stephens et al., 2001). As a result, the haplotype literature is strongly biased toward studies of mtDNA, Y-DNA, and X-linked DNA, although autosomal regions are becoming increasingly common.

Given a sample of haplotypes that arose solely from mutations, an evolutionary tree of the haplotypes exists that describes the history of mutational accumulation in DNA lineages that yield the current array of haplotype variation. This evolutionary tree is called a haplotype tree. If the DNA region being studied is also subject to genetic recombination, new haplotypes can be generated by recombining parts of preexisting haplotypes without mutation. Such recombinant haplotypes do not have a single evolutionary history, and define a reticulating network rather than a single evolutionary tree. Recombination therefore undermines the very existence of a tree-like structure as a descriptor of haplotype variation. Hence, haplotype trees do not exist for all DNA regions. In general, haplotype diversity in X-linked and autosomal DNA

can be generated by a combination of mutational accumulation in DNA lineages and recombination (Templeton et al., 2000a,b), but human mtDNA and Y-DNA are generally not subject to recombination. As a consequence, much of the literature on haplotype trees has been biased toward mtDNA and Y-DNA. However, recent studies showed that recombination in the X-linked and autosomal portions of the human genome is often concentrated into recombination hotspots separated by DNA regions with no to little recombination (Templeton et al., 2000a; Crawford et al., 2004; McVean et al., 2004). As a consequence, haplotype trees exist for many regions of the human genome, and are not confined just to mtDNA and Y-DNA.

Haplotype trees have a variety of uses in studies of human evolution. Before discussing these uses, it is first necessary to briefly review some of the results from coalescent theory, the population genetic theory underlying haplotype trees. Given this theoretical background, the estimation of haplotype trees will be discussed, followed by their various uses in studies of modern human origins.

BASIC COALESCENT THEORY

DNA replicates. In a forward sense, this means that one molecule of DNA can become two or more molecules in the future. In a backward sense, this means that two or more molecules of DNA observed today can coalesce into a single copy of DNA in the past. A "coalescent event" occurs when two lineages of DNA molecules merge back into a single DNA molecule at some time in the past. Hence, a coalescent event is simply the time inverse of a DNA replication event.

To illustrate the forward and backward implications of DNA replication, consider a hypothetical population consisting of just six haploid individuals (Fig. 1). This population has six copies of any homologous DNA segment at any given time, and it is assumed that there is no recombination. As this population reproduces, DNA replicates. By chance alone (genetic drift), some molecules get more copies into the next generation than others. Fixation occurs in this hypothetical population at generation 10. Fixation, which is inevitable in all finite populations, means that all present-day homologous DNA copies are now descended from a single DNA molecule in the past (Fig. 1).

Now consider the process of fixation as observed backwards in time. Suppose generation 10 is the current generation, and all six DNA molecules are surveyed. Note first that any DNA lineage that went extinct prior to generation 10 (parts in grey in Fig. 1) is unobservable, and there is no way of knowing about the DNA lineages that no longer exist. (In some cases, information about past genes can be obtained from fossils, as will be discussed later.) Moreover, once all our current DNA lineages have coalesced back to a common ancestor (generation 3 in Fig. 1), there is no longer any genetic variation in the coalescent process. All population genetic inference requires genetic variation. Consequently, the time to ultimate coalescence places a limit on how far back inferences can be made from a haplotype tree. The most that can be observed about the coalescent process is shown in Figure 2.

In general, consider taking a sample of n genes from a population. The word "genes" in coalescent models refers to the different copies of a homologous stretch of DNA, regardless of their genetic state. Because genetic drift inevitably causes fixation in the future sense, this means in the backward sense that all of the genes can be traced back in time to a common ancestral gene from which all current copies are descended. Figure 2 is an example of a "gene tree" that portrays how all the different present-day copies at a homologous gene locus are "related" by ordering coalescent events through time until there is only a single, ancestral molecule of DNA. The most recent common-ancestor DNA molecule is often denoted by the acronym MRCA, and the time at which this occurs is denoted by TMRCA (time to the MRCA). For example, all copies of mitochondrial DNA (mtDNA) found in living humans must eventually coalesce into a single ancestral mtDNA. Because mtDNA is inherited as a maternal haploid, this ancestral mtDNA must have been present in a female. Some scientists and much of the popular media have dubbed this bearer of our ancestral mtDNA "mitochondrial Eve," and have treated this as a startling discovery about human evolution. However, the existence of a mitochondrial Eve is trivial under coalescent theory. Finite population size (and all real populations are finite) ensures that *all* copies of *any* homologous piece of DNA present in *any* species have been derived from a single common ancestral DNA molecule in the past; indeed, this is the very definition of genetic homology. To say that humans have a mitochondrial Eve is to say only that all human mtDNA is homologous. When "Eve" is called the ancestor of us all, it only means that our mtDNA is descended from her mtDNA, and not necessarily any other piece of the human genome.

All genes that are homologous have a common ancestor in the past. Assuming that all the genes are neutral, consider the case of a random sample of just two genes from an idealized random-mating population. The probability that these two genes coalesce in the previous generation is the probability that both genes came from the same individual in the previous generation, which is $1/N$, where N is the population size in this idealized random-mating population. Coalescence requires that the two copies in the present generation were replicated from the same DNA molecule in the ancestral individual. If the genes being sampled are haploid, $1/N$ is also the probability that they coalesce in the previous generation. If the genes are autosomal, the probability that two genes are both copies of the same gene in the ancestral individual is one-half, because the ancestral individual has two genes for an autosomal locus. Hence, the total probability of an autosomal DNA molecule coalescing in the previous generation is now $(\frac{1}{2})(1/N) = 1/(2N)$. In general, the probability of coalescence in the previous generation is $1/(xN)$, where x is the ploidy level. Real populations deviate from the idealized assumptions of this simple random-mating model, so the probability of coalescence in the previous generation is often expressed as $1/(xN_{ef})$, where N_{ef} is the inbreeding effective size (a quantity that can differ substantially from N, the census size). Therefore, the probability that the two genes did *not* coalesce in the previous generation is $1 - 1/(xN_{ef})$. The probability of coalescence exactly t generations ago is the probability of no coalescence for the first $t - 1$ generations in the past, followed by a coalescent event at generation t:

$$\text{Prob.(Coalesce at } t) = \left(1 - \frac{1}{xN_{ef}}\right)^{t-1}\left(\frac{1}{xN_{ef}}\right). \quad (1)$$

The average time to coalescence is then:

$$\text{Expected(Time to Coalesce)} = \sum_{t=1}^{\infty} t\left(1 - \frac{1}{xN_{ef}}\right)^{t-1}\left(\frac{1}{xN_{ef}}\right)$$
$$= xN_{ef}. \quad (2)$$
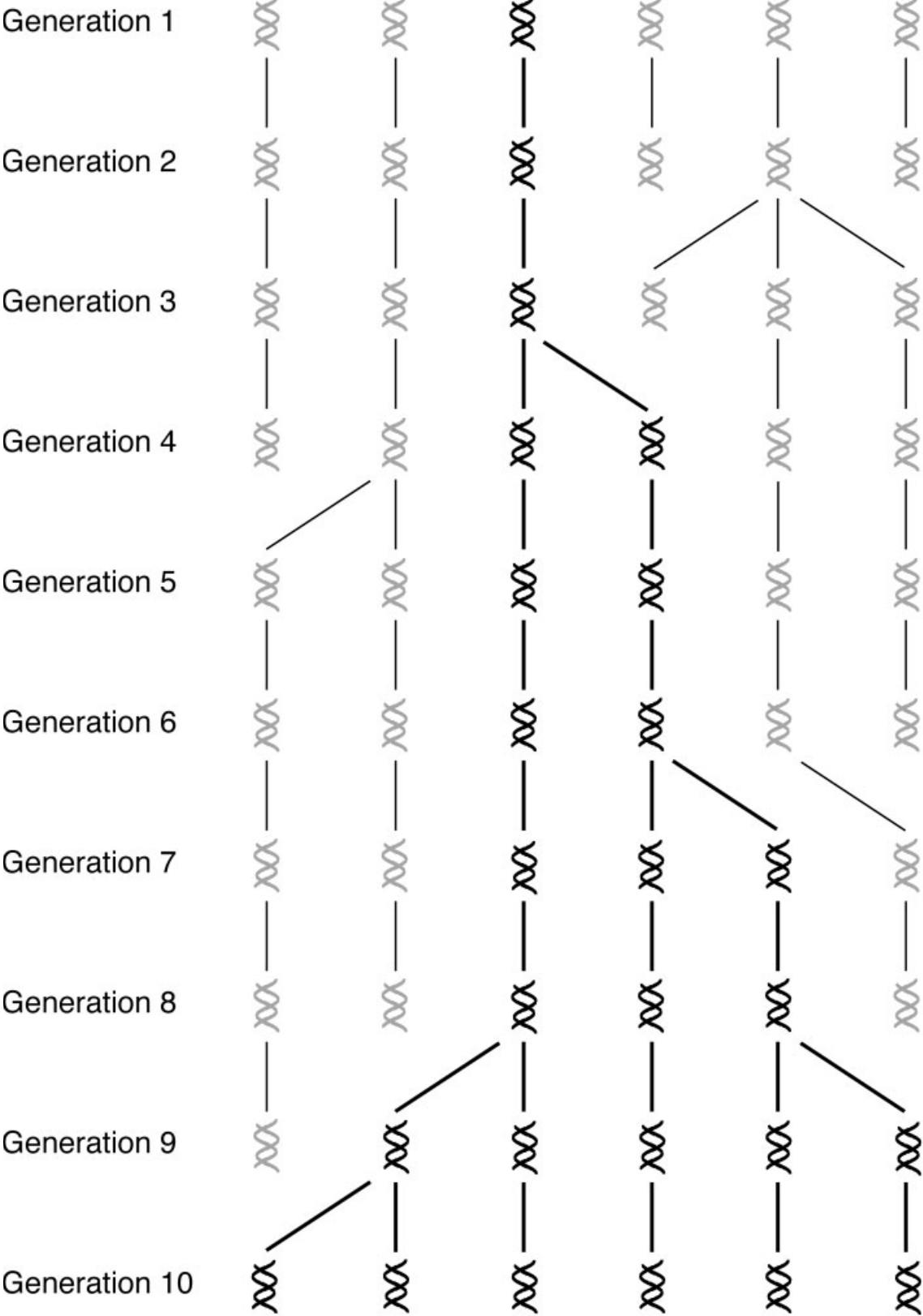
Fig. 1. Hypothetical case of genetic fixation in population with only six copies of homologous gene. Each vertical or diagonal line indicates a DNA replication event, going from top to bottom. DNA molecule with no lines coming from its bottom did not pass on any descendants. By generation 10, all six copies are descended from a single DNA molecule. This common ancestral molecule and all its descendants are shown in bold, compared to all other DNA lineages that went extinct by generation 10.
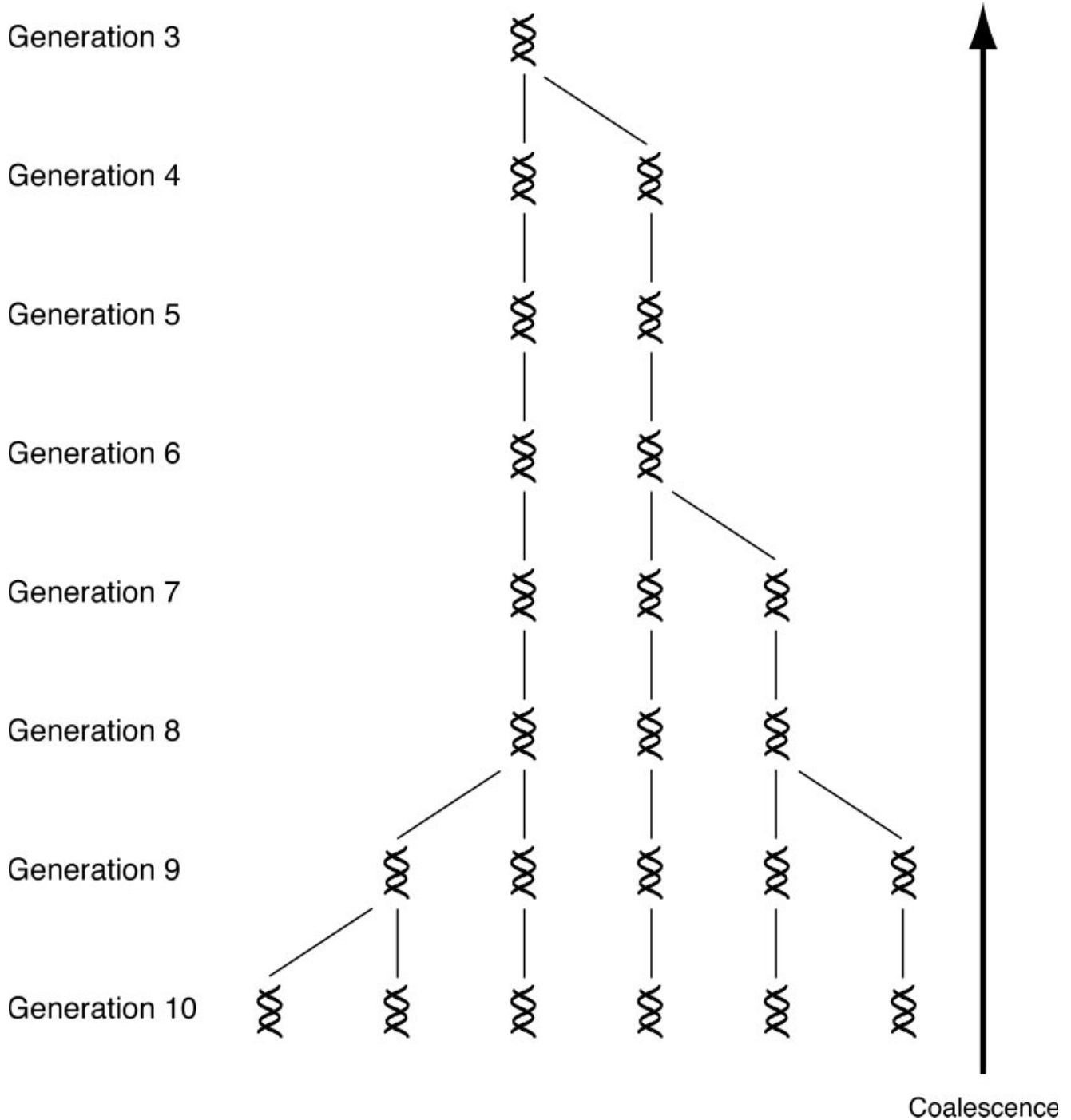
Fig. 2. Hypothetical case of a coalescent process in a population with only six copies of a homologous gene. Each combination of vertical with diagonal line indicates coalescent event, going from bottom to top. Shown is same population as in Figure 1, but only that part of genealogical structure associated with variation present at generation 10 back to generation of coalescence of all six copies present at generation 10.

These results can be generalized from a sample of two genes to a sample of n genes from a large population (Ewens, 1990; Hudson, 1990; Hein et al., 2005). The average coalescent time to the common ancestor of all n genes is $2xN_{ef}(1 - 1/n)$. Note that the expected time for ultimate coalescence approaches $2xN_{ef}$ as the sample size (n) increases.

The fact that a large sample of genes has a TMRCA of $2xN_{ef}$ and a sample of just two genes has a TMRCA of xN_{ef}

has important implications for the information contained in haplotype trees. A sample of n genes requires $n - 1$ coalescent events to yield the MRCA, and $n - 2$ of these events are expected to occur in the first half of this coalescent process, reducing the number of DNA lineages to just two. These last two DNA lineages take as much time to coalescence to the MRCA (xN_{ef} generations) as the first $n - 2$ coalescent events in order to yield the total TMRCA of $2xN_{ef}$ generations. Population genetic inference
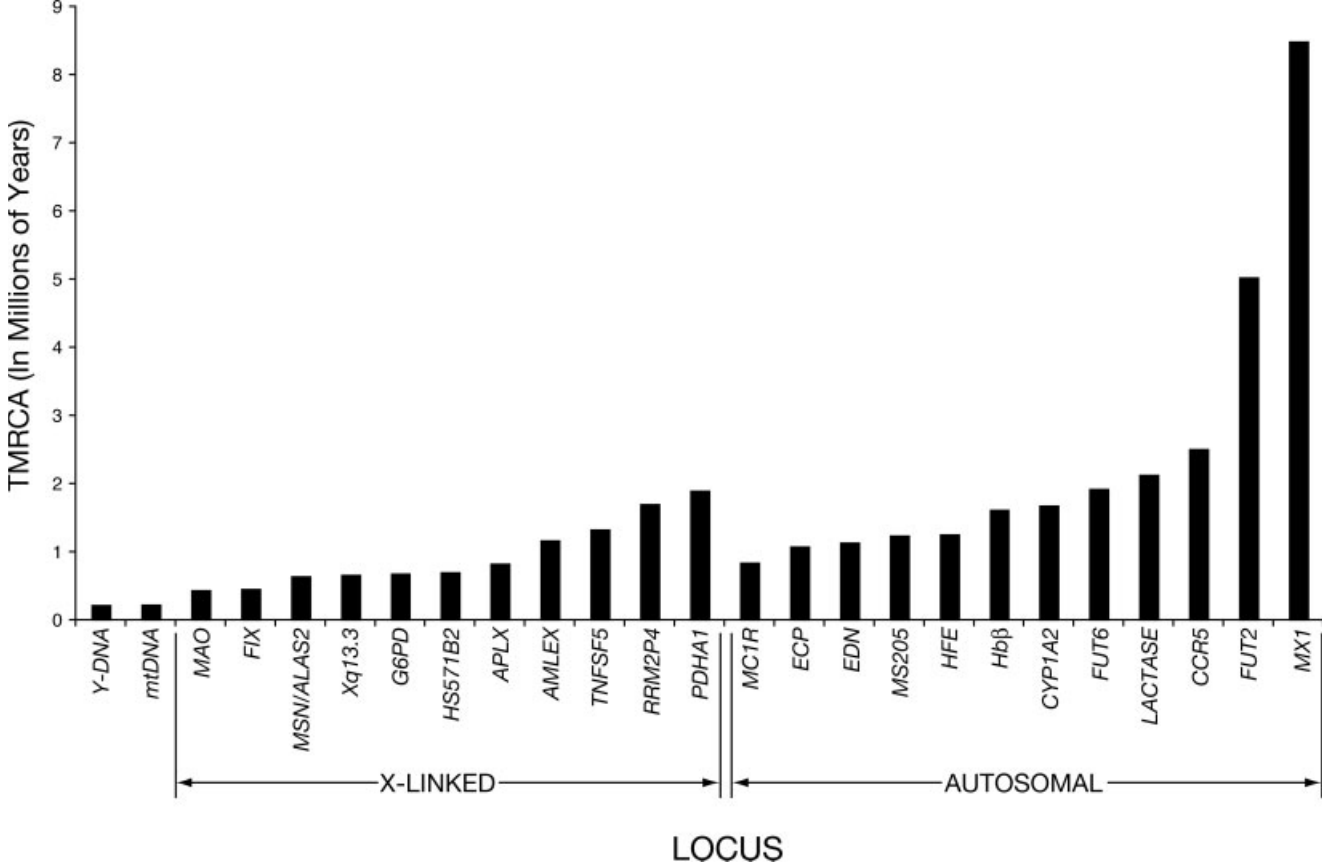
Fig. 3. Estimated coalescent times for 25 human DNA regions. Details and references for DNA regions studied are given in Templeton (2005).

requires genetic variation, and just two DNA lineages are often insufficient for many types of inferences. Hence, haplotype trees are generally most informative of evolutionary events during the more recent half of the TMRCA, with all information lost at the TMRCA.

These equations reveal that ploidy level places restrictions on the expected informative time period of a haplotype tree. For an autosomal region, $x = 2$, so the expected TMRCA is $4N_{ef}$ in a large sample. X-linked DNA is haploid in males and diploid in females, so in a population with a 50:50 sex ratio, $x = 1.5$, and the expected TMRCA is $3N_{ef}$. MtDNA is inherited as a haploid element in mammals, so $x = 1$. Moreover, mtDNA is maternally inherited, so only females pass on their mtDNA. Thus, the inbreeding effective size for the total population of males and females, the N_{ef} that is applicable to autosomal and X-linked DNA, is not applicable to mtDNA. Instead, the expected TMRCA of mtDNA is influenced only by the inbreeding effective size of females, say, $N_{ef♀}$. Thus, with $x = 1$, the expected coalescence time of mtDNA is $2N_{ef♀}$. Similarly, Y-chromosomal DNA is inherited as a paternal haploid, so its expected TMRCA is $2N_{ef♂}$, twice the inbreeding effective size for males. Because the sex ratio is close to 50:50 in humans, it is commonplace to approximate the sex-specific inbreeding sizes by $\frac{1}{2}N_{ef}$. Thus, a 1:1:3:4 ratio is expected for the relative coalescence times of mtDNA, Y-DNA, X-linked DNA, and autosomal DNA, respectively. However, inbreeding effective sizes are affected by many factors, including the variance of reproductive success. In humans, and indeed most mammals, the variance of

reproductive success is generally larger in males than in females. The higher this variance, the lower the effective size, so in general, $N_{ef♀} > N_{ef♂}$. Therefore, Y-DNA is expected to coalesce the most rapidly of all genetic elements in the human genome.

Equation 1 can also be used to calculate the variance in ultimate coalescence time. For example, the variance of time to coalescence of two genes (σ_{ct}^2) is the average or expectation of $(t - xN_{ef})^2$:

$$\sigma_{ct}^2 = \sum_{t=1}^{\infty}(t - xN_{ef})^2\left(1 - \frac{1}{xN_{ef}}\right)^{t-1}\left(\frac{1}{xN_{ef}}\right)$$
$$= xN_{ef}(xN_{ef} - 1) = x^2N_{ef}^2 - xN_{ef}. \quad (3)$$

Note that the variance in Equation 3 is proportional to N_{ef}^2, and this is also true for a sample of n genes (Ewens, 1990; Hudson, 1990; Hein et al., 2005). Hence, the time to coalescence is expected to display much variation from gene to gene, even for the same ploidy level. This variance is an inherent property of the evolutionary process itself, and is known as "evolutionary stochasticity." This stochasticity greatly limits how accurately times or effective sizes can be estimated from any single gene.

These theoretical predictions on means and variances of coalescence times are confirmed in Figure 3, which presents the estimated TMRCAs for human Y-DNA, mtDNA, 11 X-linked loci, and 12 autosomal loci using a calibration point of 6 million years ago (MYA) for the human/chim-

panzee split (Templeton, 2005). As expected, Y-DNA has the smallest coalescence time, with the mtDNA time not being much longer. The X-linked loci all have larger coalescence times, but with considerable variation. On average, the autosomal loci have the largest coalescence times of all, but with much variation and much overlap with the X-linked loci. In addition, note that one autosomal locus has a coalescence time older than 6 MYA. Because polymorphic DNA lineages can be carried across speciation events, coalescence times are sometimes older than the species (Hein et al., 2005). This situation is called transpecific polymorphism, and illustrates the danger of automatically equating an evolutionary tree of haplotypes to an evolutionary tree of species or populations. Often the two are concordant, but not always.

The coalescent model given above ignores mutation and thereby haplotype variation. Consider adding mutation to the case of a sample of two genes. As before, the probability that these two genes coalesce in the previous generation is $1/(xN_{ef})$, and the probability that they do not coalesce in the previous generation is $1 - 1/(xN_{ef})$. Assuming an infinite-alleles model of mutation in which each mutation creates a distinct haplotype, the probability that two genes are identical by descent is the probability that the two gene lineages coalesce *before* a mutation occurred in either lineage. If the two genes coalesced t generations ago, this means that there were $2t$ DNA replication events at risk for mutation (two gene lineages each undergoing t replication events). Hence, the probability that neither gene lineage experienced any mutation over this entire time period is $(1 - \mu)^{2t}$. Putting this probability together with Equation 1 yields:

Prob.(coalescence before mutation)

$$= \left(1 - \frac{1}{xN_{ef}}\right)^{t-1}\left(\frac{1}{xN_{ef}}\right)(1-\mu)^{2t}. \quad (4)$$

Now consider the probability that a mutation occurred t generations ago and before coalescence. This means that in the $2t$ DNA replication events being considered, only one replication event experienced a mutation, and the other $(2t - 1)$ replication events did not. Because there are two gene lineages, either one of them could have mutated at generation t, so the total probability of having a single mutation at generation t in $2t$ DNA replication events is $2\mu(1 - \mu)^{2t-1}$. The probability of no coalescence in these t generations is $[1 - 1/(xN_{ef})]^t$. Putting these two probabilities together yields:

Prob.(mutation before coalescence)

$$= \left(1 - \frac{1}{xN_{ef}}\right)^{t} 2\mu(1-\mu)^{2t-1}. \quad (5)$$

If μ is very small and N_{ef} is very large, then the occurrence of both coalescence and mutation during the same generation can be ignored. Therefore, the conditional probability of a mutation before coalescence given either mutation or coalescence is:

Prob.(mutation before coalescence | mutation or coalescence)

$$= \frac{2\mu(1-\mu)^{2t-1}(1-1/xN_{ef})^t}{2\mu(1-\mu)^{2t-1}(1-1/xN_{ef})^t + 1/xN_{ef}(1-\mu)^{2t}(1-1/xN_{ef})^{t-1}}$$

$$= \frac{2xN_{ef}\mu - 2\mu}{2xN_{ef}\mu - 3\mu + 1}. \quad (6)$$

If $\mu \ll N_{ef}\mu$ (i.e., a large inbreeding effective size) and defining $\theta = 2xN_{ef}\mu$, Equation 6 simplifies to:

Prob.(mutation before coalescence

$$| \text{ mutation or coalescence}) = \frac{\theta}{\theta + 1}. \quad (7)$$

When mutation occurs before coalescence, the two gene lineages being compared must be different haplotypes under the infinite-alleles model. Since the two genes were drawn at random from the population, Equation 7 is equivalent to the expected heterozygosity under random mating. Note also that Equations 2, 3, and 7 provide a direct relationship between the coalescent process and the demographic parameter of inbreeding effective size. Thus, the coalescent process contains information about past demography.

With mutation in the model, a distinction can be made between gene trees vs. haplotype trees. As noted before, gene trees describe how different copies of DNA sampled at a homologous gene locus are "related" by ordering coalescent events (Fig. 2). Figure 4A is a repeat of the gene tree shown in Figure 2, but now some of the DNA replication events are assumed to have experienced mutation. The gene tree shows precise information here about the gene genealogy, including cases in which two genes are closer genealogically even though they are not identical (e.g., A and one of the copies of B), compared to two genes that are more distant genealogically yet identical in sequence (the two B genes). Such precise information about gene genealogy is usually not available. For example, from sequence data alone, there is no way of knowing that one B gene is actually genealogically closer to the A gene than to its indistinguishable B copy. The only branches in the gene tree that can be observed from sequence data are those marked by a mutation. Therefore, the tree observable from sequence data retains only those branches that mark the transition from one haplotype to another. This lower-resolution tree is called a "haplotype tree." The haplotype tree is the gene tree in which all branches not marked by a mutational event are collapsed together. Figure 4B show the haplotype tree corresponding to Figure 4A. The haplotype frequencies in the current sample are also observable (e.g., haplotype B in Fig. 4 has two copies in our sample, and all other haplotypes are present in only one copy), as are the spatial distributions of the haplotypes when sampling includes more than one location. Coalescent theory is applicable to both gene and haplotype trees, but the observable coalescent process deals only with the evolutionary history of the haplotypes and their current frequencies and spatial distributions in the sample.

ESTIMATION OF HAPLOTYPE TREES

In most modern studies, haplotypes appear in the form of DNA sequence data, and hence haplotype trees can be estimated through the same algorithms and programs used by molecular systematists to estimate trees of species from DNA sequence data (Felsenstein, 2004). However, there are some problems and issues that are unique to intraspecific haplotype trees that need special attention.

Recombination

Most algorithms for estimating evolutionary trees from DNA sequence data assume that all variation arose only
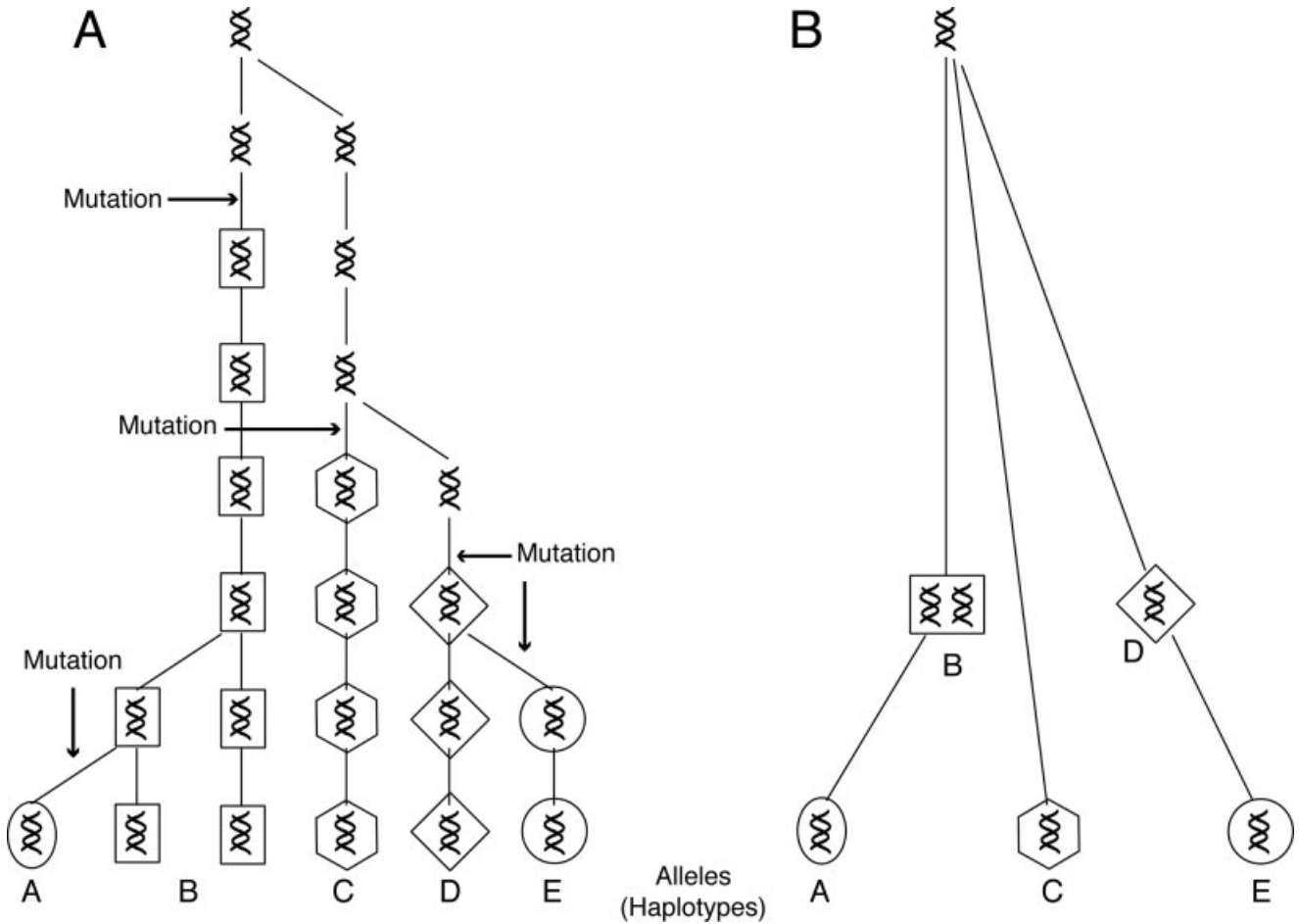
Fig. 4. Gene trees vs. haplotype trees. **A:** Same gene tree portrayed in Figure 2, but now with some mutational events added on. Each mutation creates a new, distinguishable allele, as indicated by a change in the shape of the box containing DNA molecule. The extinct, ancestral haplotype is shown by an unboxed DNA motif. **B:** Haplotype tree associated with gene tree in A. The only observable coalescent events are those associated with mutational change, so each line in this tree represents a single mutational change. Letters correspond to different allelic categories that exist in the current population.

through mutation. However, recombination can place mutations that arose in different DNA lineages onto the same DNA molecule, and thus produce a novel haplotype that represents a fusion of two (or more) distinct genealogical histories. Recombination undermines the fundamental premise of most tree-estimation programs, and can create phylogenetic ambiguities and phantom "branches" and "mutations" (Crandall and Templeton, 1999; Templeton et al., 2000a; Posada and Crandall, 2002). Recombination can occur in the nuclear DNA, with the exception of most of the Y-chromosome. Consequently, it is important to check for the presence of recombination *before* (or simultaneously with) estimating a phylogenetic tree in a nuclear DNA region, and a variety of programs are available for this purpose (Posada and Crandall, 2001; Posada, 2002; Posada et al., 2002; Tsaousis et al., 2005).

If no recombination is detected, a standard phylogenetic analysis is possible. If a few, rare recombinants are detected, they can be excluded from the data set, and a tree can be estimated for the nonrecombinant haplotypes (Templeton et al., 1987). Sometimes recombination is common but concentrated into a hotspot with regions of little or no recombination on either side of the hotspot. In this case, recombination hotspots are used to subdivide the DNA region, and separate haplotype trees are estimated

for each low-recombination block (Templeton et al., 2000b). If recombination is common and uniformly distributed in the DNA region, no haplotype tree exists.

Mutation

Tree-estimation algorithms such as neighbor-joining, maximum likelihood, and Bayesian procedures have to specify a model for nucleotide substitution to estimate evolutionary trees from DNA sequence data. Phylogenetic inferences from the same human haplotype data sets can vary, and have varied, under different models of mutation (Gutierrez et al., 2002). Moreover, many of the coalescent-based procedures for extracting information from haplotype trees and for making demographic inferences using coalescent theory are also sensitive to the underlying mutational model (Palsboll et al., 2004). Hence, it is important to specify an appropriate mutational model for a given DNA region, and programs are available for this purpose (Posada and Crandall, 1998; Posada and Buckley, 2004).

Testing the DNA sequence data for its fit to various models of substitution is important, because the default model in many programs and coalescent simulations is a model of uniform mutation (all sites are equally likely to

mutate), with all mutational transitions equally likely, given a mutation. Many coalescent statistics couple the uniform-mutation model with the infinite-sites model that assumes that each mutation occurs at a different nucleotide site, thereby making it impossible for the same site to experience multiple mutational hits. Such a mutational model is unrealistic for much human DNA (Templeton et al., 2000a; Bacon et al., 2001; Lercher and Hurst, 2002; Zhao and Boerwinkle, 2002; Tomso and Bell, 2003; Tomatsu et al., 2004; Buzin et al., 2005; Fryxell and Moon, 2005), and can lead to incorrect conclusions (Templeton et al., 2000a).

The infinite-sites model is sometimes mistakenly equated with the infinite-alleles model (Innan et al., 2005). If the infinite-sites model is true, then all mutations must result in new haplotypes, so the infinite-alleles model is also true. However, if a site mutates more than once during the coalescent process, it will only violate the infinite-alleles assumption if the exact same mutation occurs at the same site *and* on exactly the same ancestral haplotype background. This additional restriction means that the infinite-alleles model is more robust to nonrandom mutation than the infinite-sites model.

Multifurcation and branch lengths

Most interspecific tree-estimating algorithms assume a bifurcating tree structure. However, coalescent theory indicates that the probability of a mutation hitting a particular haplotype is proportional to the frequency of that haplotype in the population: the more copies of a haplotype, the more likely that at least one copy will experience a mutation. This means that a common haplotype is likely to have many mutational hits, resulting in multifurcation. Hence, a bifurcating tree structure is an inappropriate assumption for intraspecific haplotype trees.

Another difference between intra- vs. interspecific trees is the expected branch lengths. When a haplotype is hit by a mutation, it generates a derived haplotype that is only one mutational step from the ancestral haplotype. Both the ancestral and derived haplotypes often persist as polymorphisms within a species. Hence, branch lengths of one mutation are common between haplotypes in many human haplotype trees. Branches defined by a single mutation are often assigned low measures of confidence in interspecific phylogenies. Yet, small branch lengths define the branches that have the greatest statistical confidence in intraspecific trees (Templeton et al., 1992).

Statistical parsimony (Templeton et al., 1992; Crandall, 1994) is one of the few tree-estimation methods specifically designed for intraspecific haplotype trees. This method allows multifurcations, and assigns the greatest statistical confidence to those linkages defined by the least number of mutational changes. Moreover, this method can be combined with other information (such as haplotype frequencies) that, under coalescent theory, allows tree ambiguities to be resolved (Crandall and Templeton, 1993).

USING HAPLOTYPE TREES FOR PHYLOGEOGRAPHIC ANALYSES

Intraspecific phylogeography is the study of the geographical distribution of genetic variation within a species, and the underlying evolutionary processes and events that created that geographical distribution. Haplotype trees have been a common tool for phylogeographic inference about human evolution ever since the publication of the human mtDNA haplotype tree (Cann et al., 1987). As was the norm at the time, Cann et al. (1987) did not perform any formal phylogeographic analysis, but instead made their inferences from a visual inspection of the mtDNA tree and the geographical distribution of its branches. For example, they inferred that the mtDNA tree had its root in Africa, and therefore, humanity had its root in Africa. They saw that all the oldest clades (branches on the tree) were strictly African, whereas the younger clades were found in both Africa and outside Africa. Therefore, they inferred that human populations expanded out of Africa into Eurasia. Because they found no old branches of the mtDNA tree in Eurasia and the out-of-Africa expansion was dated, using a molecular clock, to about 100,000 years ago, they concluded that Eurasian populations were replaced or driven to complete genetic extinction by expanding African populations. This method of visual inspection equates the haplotype tree to a tree of human populations, with no statistical quantification of the confidence in the observed spatial/temporal patterns as a function of sample size and number of sample locations.

Coalescent theory warns us that haplotype trees should not be equated with trees of populations. Populations can contain many different haplotype lineages as polymorphisms at any given time, and coalescent theory indicates that the transfer of shared polymorphic lineages across historical events or gene-flow processes, followed by subsequent genetic drift, can frequently lead to haplotype-tree topologies that are discordant with population-tree topologies (Avise, 2000). Moreover, population trees need not exist at all within a species. If a species has extensive gene flow throughout all parts of its geographic range, the species would evolve as a single evolutionary lineage, with no intraspecific population-tree whatsoever. Instead, local populations in such a species would be genetically interrelated by a trellis or lattice-like structure, not distinct branches on a tree. Nevertheless, such a species would still have haplotype trees for all its DNA regions with no to little recombination. Consequently, haplotype trees should never automatically be equated with trees of populations, which need not even exist. Population-level evolutionary forces influence haplotype trees, but the information about population-level processes and events found in haplotype trees needs to be extracted carefully and in light of coalescent theory, and *never* equated with the haplotype tree itself.

Three different methods for extracting information about past human evolution from haplotype trees will be considered here. These methods tend to be most informative on different time scales, and they will be considered in temporal order from the shallowest to the deepest in time.

Tracing recent haplotypes

After a new haplotype has been created by mutation, it exists initially as a single copy at a single point in space under the infinite-alleles model. Many successful DNA replication events are needed before this newly arisen haplotype can become common and spread through space. As a consequence, many newly arisen haplotypes tend to be rare in the species as a whole (even though they may be common in some local demes), and are restricted in their geographical range. When individuals or populations bearing such new, globally rare haplotypes move through space, their movements are traced by these rare haplo-

types (Richards et al., 2000). Finding such recently arisen, globally rare, geographically restricted haplotypes is no longer difficult. When genetic diversity is scored at the DNA sequence level, haplotypes with these properties are frequently found in most species. At first, this may seem counterintuitive. Obviously, by definition, any particular rare haplotype is rare, but the category of rare haplotypes is collectively common. Large samples of genes in most species will reveal many rare haplotypes. When placed into a haplotype tree, most rare, geographically restricted haplotypes are located near the tips of the haplotype tree, indicating that they are indeed newly arisen, as expected from coalescent theory (Castelloe and Templeton, 1994). Consequently, finding recently arisen and globally rare haplotypes is not difficult in most species, and such haplotypes provide a powerful marker for recent movements through space of individuals or populations (Richards et al., 2000).

For example, Thomas et al. (2000) identified several Y-chromosomal haplotypes from the tips of the Y-DNA haplotype tree that were also of restricted geographical distribution. They surveyed Y-chromosomal variation in the Lemba, a southern African group who speak a variety of Bantu languages but claim Jewish ancestry (Wilson and Goldstein, 2000). According to their oral traditions, the Lemba are descended from a group of Jewish males who centuries ago came down the eastern coast of Africa by boat. Many were lost at sea, but the remainder interbred with local Bantu women, thereby establishing the ancestors of the current Lemba, who are now found mostly in South Africa and Zimbabwe. Thomas et al. (2000) showed that about two-thirds of the Lemba Y chromosomes have a Middle Eastern origin, and one-third a Bantu origin. Moreover, one particular Y-chromosomal haplotype is found in frequencies ranging from 0.100–0.231 in various Jewish populations, but is very rare or absent from most other human populations. Yet the frequency of this newly arisen, globally rare haplotype is 0.088 in the Lemba of southern Africa, consistent with a genetic interconnection between the Lemba and Jews of Middle Eastern origin. In contrast to the Y haplotypes, there is no evidence of Semitic admixture with the maternally inherited mtDNA (Soodyall 1993), a pattern also consistent with the oral traditions of the Lemba that the original admixture involved Jewish men and Bantu women.

These studies illustrate the richness of detailed inference that studies on newly arisen haplotypes can provide about recent movements of individuals and populations through space. However, what about more ancient movements? Because these haplotypes are young in an evolutionary sense, they offer little to no insight into older movements and historical events. Rare, tip haplotypes are useful in humans only for inferences going back a few thousand years at the most, and often less.

Founder analysis

Richards et al. (2000) described a formalized procedure known as founder analysis that extends haplotype analyses further back in time, and that can detect both migration and backmigration. Less formal but similar applications of founder analysis appeared earlier in the human genetic literature (Stoneking et al., 1990; Torroni et al., 1993a,b; Sykes et al., 1995). Richards et al. (2000) focused on the problem of genetic interchange involving two locations, and specifically the Near East and Europe over the

last 50,000 years. The idea is to identify "founder" haplotypes that were introduced from one population into the other. An initial set of candidate founder haplotypes are identified by searching for:

1. Identical haplotypes found in both geographical areas; and
2. Haplotypes that are either:
   a. Inferred internal nodes in the haplotype tree but that are not actually present in the sample, but that do have derivatives in the haplotype tree that are found in both locations; or
   b. A haplotype found in only one area, but whose immediate evolutionary derivatives include at least one haplotype found in the second area.

One problem with these criteria is deviation from the infinite-alleles model. Such deviations can cause the same haplotype to arise in parallel in different geographical locations. Richards et al. (2000) provided additional criteria for narrowing the list of candidate founders by using haplotype frequency data in conjunction with the topology of the haplotype tree. For example, coalescent theory indicates that deviations from the infinite-alleles model due to recurrent mutation are most likely when the ancestral haplotype is common in both areas and recurrent mutations produce the same derived, tip haplotype state in parallel in the two regions. Hence, excluding tip haplotypes derived from common ancestral haplotypes helps eliminate such false matches. Additional and more quantitative criteria for further exclusions are given in Richards et al. (2000).

Once a list of founder types has been identified, haplotype clades derived from the founder haplotypes can be used to quantify the impact of the migration event in terms of the percent of the gene pool in the geographical area that is in the founder clade. Moreover, genetic diversity can be measured within founder clades in the new geographical area to estimate the founding time in the new geographical area in mutational units, which can then be converted into an absolute age if the rate of mutational accumulation has been estimated (usually through the use of a calibrated, molecular clock). Figure 5 shows age ranges for founder clades that represent movement from the Near East into Europe, and collectively account for 76% of the European mtDNA gene pool (Richards, 2003). As can be seen, this method detected genetic input into Europe over a broad time period, ranging from the Neolithic about 9,000 years before the present (YBP) to the early Upper Palaeolithic, about 45,000 YPB. Moreover, many of these migration events are clustered within the late Upper Paleolithic (14,500 YPB). The results in Figure 5 are corrected for backmigration from Europe to the Near East. When all gene flow was regarded as being from the Near East to Europe, more gene flow was detected, and the main cluster of migration moved from the late Upper Paleolithic to the Neolithic. This shows the importance of not forcing prior models of one-way gene flow onto the data. Such assumptions tend to produce self-reinforcing inferences that are not robust to more general models of genetic interchange.

Nested-clade analysis

The founder analysis of Richards et al. (2000) extends the use of haplotype trees to migratory events influencing

Founder Clade Name: Percent Current European Gene Pool
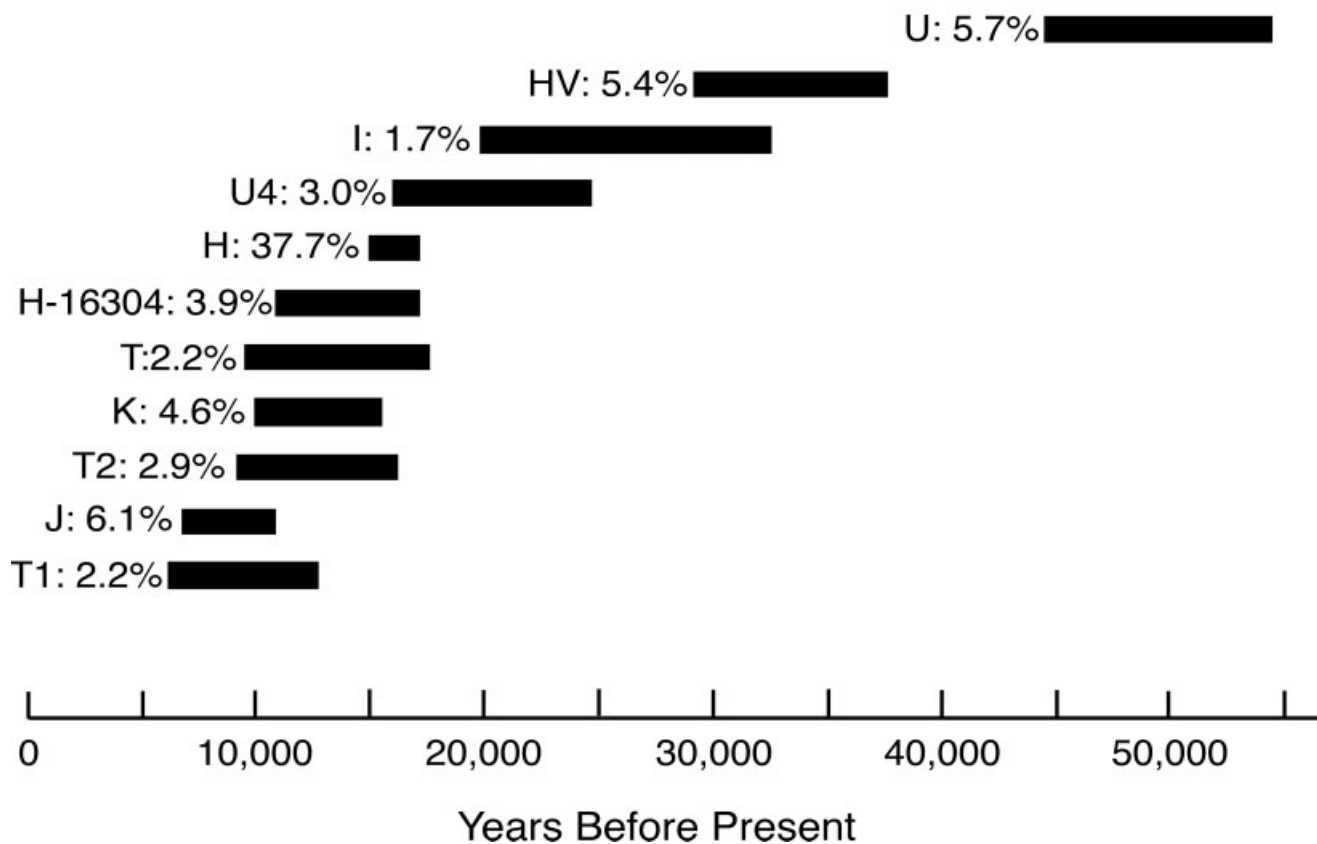
Fig. 5. Age ranges for major founder clades that indicate genetic input from Near East into Europe. Bars by clade names indicate 95% confidence range for age of founder clade in Europe. Percentages by bars indicate percentages of current European mtDNA gene pool contained within that haplotype clade. Founder clade designations are given in Richards et al. (2000). Modified from Richards (2003).

current human population structure to about 50,000 YBP. Moreover, the founder-analysis approach overlaps temporally with the rare tip-haplotype approach (which is a special case of founder analysis). In principle, haplotype trees contain information up to the TMRCA, but mostly in the more recent half of the TMRCA. Even half the TMRCA for many DNA regions is much older than 50,000 YBP, so founder analysis does not use all of the historical information that is potentially contained within haplotype trees. Nested-clade analysis is designed to extract information from the deeper portions of the haplotype tree, although its temporal range of utility overlaps with that of founder analysis.

Nested-clade analysis uses the haplotype tree to define a series of hierarchically nested clades (branches within branches). Such nested hierarchies are commonly used in comparative evolutionary analyses of species or higher

taxa, but can also be applied to the haplotype variation found within a species in DNA regions with no or low recombination. Given a haplotype tree, just a few rules are sufficient to convert it into a nested series of clades (Templeton et al., 1987; Templeton and Sing, 1993). For example, Figure 6 gives the haplotype tree estimated via statistical parsimony for the human *RRM2P4* pseudogene on the X chromosome (Garrigan et al., 2005). The first step of the nesting procedure is to nest the haplotypes into one-step clades. This is accomplished by starting at the tips of the tree (the tip haplotypes in Figure 6 are haplotypes *A*, *C*, *E*, and *H*), and moving one mutational step into the interior tree, uniting all haplotypes that are interconnected by such a step into a single clade. After the one-step clades associated with tip haplotypes have been defined, they are pruned off the tree, and the remaining interior haplotypes are nested in the same fashion, start-

Fig. 6. Haplotype tree and nested-clade design for human *RRM2P4* pseudogene. **A:** Haplotypes are indicated by capital letters, and solid lines indicate single mutational change. "0" indicates an interior node that represents a haplotype state necessary to interconnect sampled haplotypes, but that was not found in sample. Solid boxes indicate one-step clades generated by moving one mutational step in from tips; dashed boxes indicate one-step clades obtained after the solid boxed set is excluded, followed by moving in one mutational step from tips of pruned tree. **B:** Tree of one-step clades and resulting two-step clades when same nesting rules used in A are applied to one-step clades instead of haplotypes.
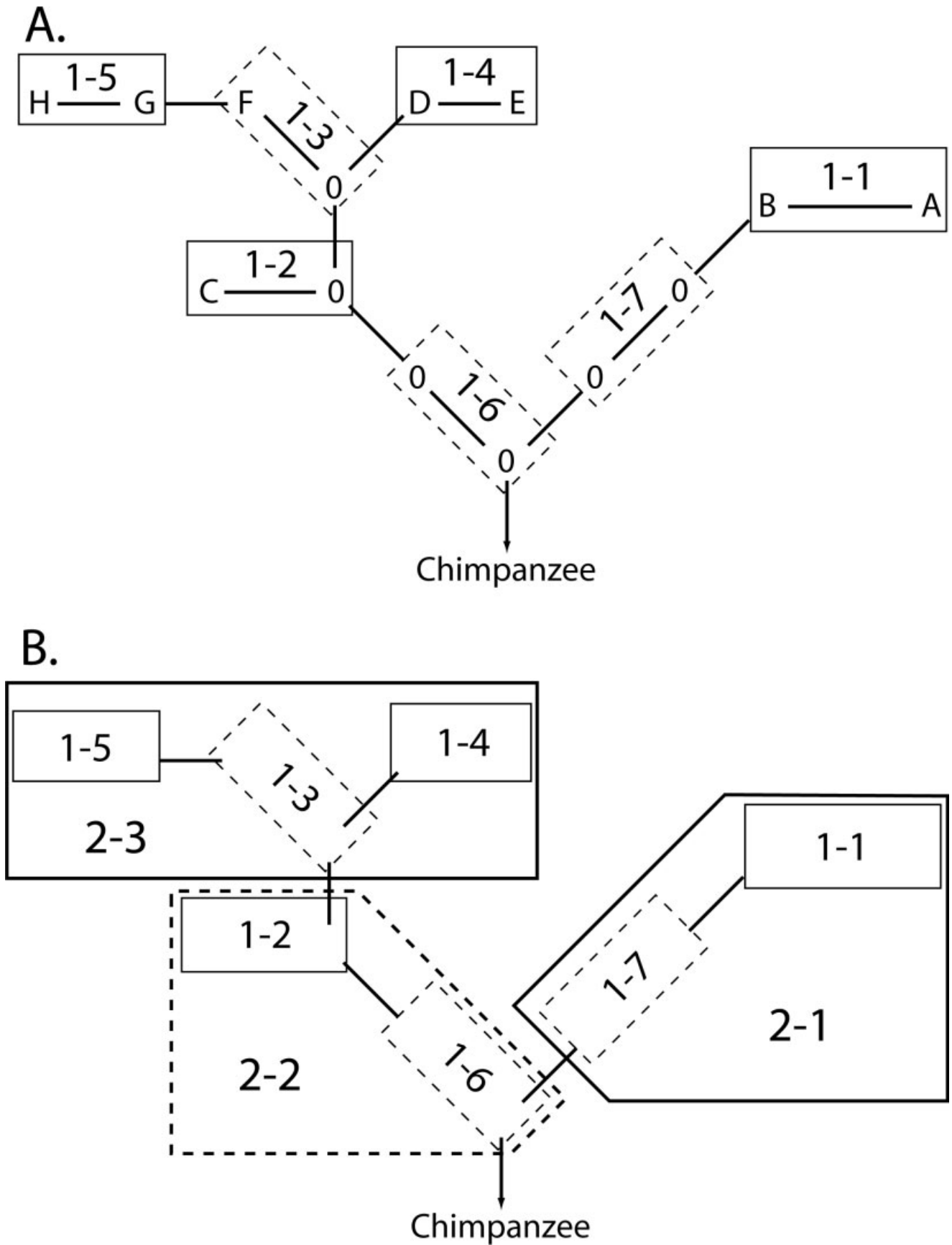
Fig. 6.

TABLE 1. Nested-clade analysis of RRM2P4 *pseudogene, using nested design given in Figure 6 and geographical distribution data given in Garrigan et al. (2005)*

Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
D[1]	5,224[2]	5,265						
E	0	4,647						
I − T[3]	5,224	617	1-4	5,241	5,831			
F			1-3	0	2,603			
G	2,649[S4]	5,954						
H	6,133	7,276						
I − T	−3,484[S]	−1,322	1-5	6,762	6,937			
1-2-11-12 range expansion[5]			I − T	−5,833	−3,658	2-3	6,145[S]	6,485
C			1-2			2-2	0	6,148
A	1,862	2,492	1-1			2-1	2,949	13,496[L]
B	0	6,603				I − T	−5,905[L]	−863
I − T	−1,862	4,111				1-2-3-5-6-7 restricted gene flow with some long-distance dispersal		

[1] Interior clades are indicated by shading.
[2] Distances are given in kilometers.
[3] I − T indicates difference of pooled interior minus pooled tip distances.
[4] Superscript "S" means significantly small; superscript "L" means significantly large (5% level).
[5] Numbers refer to sequence in inference key, followed by biological conclusion.

ing with the tips of the pruned tree. Pruning and nesting are repeated as necessary, until all haplotypes and internal nodes in the tree have been placed into one-step clades, as shown in Figure 6A. Additional nesting rules are sometimes used to deal with more complex tree topologies and with uncertainties in the estimated haplotype tree (Templeton and Sing, 1993). The next level of nesting is achieved by applying the same nesting rules to the one-step clades. This results in a set of two-step clades, as shown in Figure 6B for the *RRM2P4* pseudogene. The two-step clades are next nested into three-step clades, and so on until all clades in the tree would be nested into a single category, at which point the nesting stops. For the *RRM2P4* region, there are two tip two-step clades (2-1 and 2-3), both united to the same interior twp-step clade (2-2) through a single mutational step. Accordingly, the three-step clade that emerges from the nesting rules includes the entire tree, so the nesting is stopped at the level of two-step clades nested within the total tree.

Nested clades contain temporal information. If rooted to an interior haplotype or node, tip haplotypes or clades must be younger than the interior haplotype or clade to which they are connected. This is the case in Figure 6, where the arrow pointing to the chimpanzee indicates the root of the tree. Even if the tree were unrooted, coalescent theory predicts that tips are highly likely to be younger than the interiors to which they are connected (Castelloe and Templeton, 1994). Moreover, a clade can be no older and is generally younger than the clade within which it is nested. Therefore, both rooted and unrooted trees contain temporal information in their nested clade hierarchies.

A nested-clade analysis also requires the spatial distribution of haplotypes and clades of haplotypes to be quantified. The geographical data are quantified by two distances (Templeton et al., 1995). The first is the clade distance, D_c, which measures how widespread the clade is spatially. When measuring spatial spread with geographical distances, the clade distance is determined by calculating the average latitude and longitude for all observations of the clade in the sample, weighted by the local frequencies of the clade at each location. This estimates the geographical center for the clade. Next, the great circle distance (the shortest distance on the surface of a sphere

between two points on the surface) from a location containing one or more members of the clade to the geographical center is calculated, and these distances are averaged over all locations containing the clade of interest, once again weighted by the frequencies of the clade in the local samples. The second measure of geographical distribution of a haplotype or clade is the nested-clade distance, D_n, which quantifies how far away a haplotype or clade is located from those haplotypes or clades to which it is most closely related evolutionarily, i.e., the clades with which it is nested into a higher-level clade. For geographical distance, the first step in calculating the nested-clade distance is to find the geographical center for all individuals not only bearing members of the clade of interest, but also bearing any other clades that are nested with the clade of interest at the next higher level of nesting. The nested-clade distance is then calculated as the average distance that an individual bearing a haplotype from the clade of interest lies from the geographical center of the nesting clade. Once again, all averages are weighted by local frequencies. The clade distances and nested-clade distances are shown in Table 1 for the *RRM2P4* pseudogene, using the nesting design given in Figure 6.

These distance measures are used to test the null hypothesis that the haplotypes or clades nested within a high-level nesting clade show no geographical associations, given their overall frequencies. This null hypothesis is tested by randomly permuting the observations (the number of times a clade is observed at a particular sampling location) within a nesting clade across all the sampling locations in which one or more members of the nesting clade are found. The permutation procedure preserves the overall clade frequencies and sample sizes per locality (Templeton et al., 1995). After each random permutation, the clade and nested-clade distances are recalculated. By doing this a thousand or more times, the distribution of these distances is simulated under the null hypothesis that all nested clades, given a fixed overall frequency, are randomly distributed throughout the geographical range of the nesting clade. The observed clade and nested-clade distances are then contrasted to this null distribution, to infer which distances are statistically significantly large and which are significantly small. The distances at the *RRM2P4* pseudogene that are significantly large at the

5% level are indicated by a superscript "L" in Table 1, and those significantly small by a superscript "S." The absolute value of the distance alone does not determine significance; rather, significance is a function of both the spatial distribution of the clade and its frequency. For example, haplotype B in Table 1 has a clade distance of 0 km, which is not significantly small. In contrast, haplotype G has a clade distance of 2,649 km, which is significantly small at the 0.1% level. The reason for this apparent discrepancy is that haplotype B occurred only once in the total sample. Because the permutation procedure preserves the marginal frequencies, the single observation of haplotype B must always be placed at a single location when being permuted across the locations that haplotypes A and B collectively share, which always results in a clade distance of zero. Hence, given its frequency, haplotype B must have $D_c = 0$, so the clade distance of zero is without statistical significance (this points out the danger once again of making inferences from how a haplotype tree overlays upon space without any assessment of sample sizes). In contrast, haplotype G occurs six times in the overall sample, with five observations from two locations in Africa, and one observation from Europe. Haplotypes G and H (members of the nesting clades 1–5 in Fig. 6) are collectively found in Africa, Europe, Asia, North America, and South America. The permutation procedure reveals that is extremely unlikely that six observations would show such a tight geographical clustering over such a broad geographical range if both haplotypes G and H were indeed homogeneously distributed geographically. Such calculations are preformed by the computer program GEODIS (Posada et al., 2002).

Additional distance measures are generated by subtracting the average of the clade distances for all the tips pooled together from the corresponding average for the older interiors. This is also done for nested-clade distances. Pooling increases the sample size, and thereby augments statistical power. The average interior-tip difference captures the temporal contrast of old vs. young within a nesting clade. These interior-tip differences are also shown in Table 1 for the *RRM2P4* locus.

Statistical significance is not the same as biological significance. Statistical significance tells us that the measures we are calculating are based on a sufficient number of observations that we can be confident that geographical associations exist with the haplotype tree. To arrive at biological significance, we must examine how various types of recurrent gene flow or historical events can create specific patterns of geographical association. Moreover, some statistically significant geographical associations are biologically uninterpretable. For example, suppose a geographically widespread species is characterized by gene flow with isolation by distance. However, if one only sampled local populations from two separate geographical clusters that were very distant from one another and did not sample geographically intermediate populations, the result would be two genetically differentiated geographical clusters. Such a pattern could be confused with fragmentation, and the pattern associated with isolation by distance would only become apparent when geographically intermediate populations were sampled. Nested-clade analysis therefore addresses the issue of biological interpretation though an explicit inference key that uses predictions from coalescent theory, and that checks for sampling artifacts.

Consider, for example, the patterns expected with gene flow constrained by isolation by distance. When a mutation first occurs, the resulting new haplotype is found only in its area of origin. With each passing generation, a haplotype lineage that persists has a greater and greater chance of spreading to additional locations via restricted gene flow. Hence, the clade distances should increase with time under a model of restricted gene flow. Moreover, newer haplotypes are found within the geographical range of the ancestral haplotype from which they were derived (taking into account sampling error), and since geographical centers move slowly under isolation by distance, the clade and nested distances should yield similar patterns of statistical significance.

Consider now the patterns expected by a range expansion involving the movement of a whole population. When range expansion occurs, those haplotypes found in the ancestral population that was the source of the range expansion will become widespread geographically (large clade distances). This will sometimes include relatively young haplotypes or clades that are globally rare and often restricted just to the ancestral area (recall the discussion of rare, tip haplotypes). However, some of those young, rare haplotypes in the ancestral source population can be carried along with the population range expansion, resulting in clade distances that are large for their frequency. An example of this is shown in Table 1 for the *RRM2P4* locus. Within nesting clades 1–5, the older haplotype (G) is restricted mostly to Africa and has a significantly small clade distance of 2,649 km, whereas the younger tip haplotype (H) is found in Asia and the Americas with a clade distance of 6,133 km, thereby having a much wider geographical distribution than its immediate ancestral haplotype. This is exactly the opposite of the pattern associated with isolation by distance, and indicates a range expansion. Moreover, because the ancestral haplotype is found almost exclusively in Africa, this represents an expansion out of Africa.

Other patterns can lead to the inference of range expansion, and there are other expected patterns for additional evolutionary forces and events. No single test/statistic discriminates between recurrent gene flow, past fragmentation, and past range expansion; rather, it is a pattern formed from several statistics that allows discrimination. Also, many different patterns can lead to the same biological conclusion because a single evolutionary event or process can have multiple genetic impacts. Moreover, as pointed out above, a statistically significant pattern can still be biologically ambiguous because of inadequate geographical sampling. Finally, nested-clade analysis searches out multiple, overlying patterns within the same data set. In light of these complexities in biological and sampling interpretation (which reflect reality), an inference key was provided as an appendix to Templeton et al. (1995), with the latest version being available at http://darwin.uvigo.es/ along with the program GEODIS for implementing nested-clade analyses. The inference key was extensively validated by applying nested-clade analysis to actual data sets with 150 a priori expectations (Templeton, 2004b). The inference key makes few errors, with the most common error being failure to detect an expected event. Only rarely did a nested-clade analysis result in a false positive.

Although the analysis of Templeton (2004b) showed that the inference key works well with actual data sets, Knowles and Maddison (2002) claimed that the inference key yields a high false-positive rate with simulated data. There is widespread misunderstanding of exactly what these simulations show, because the details of the simulations were not published in Knowles and Maddison (2002),

and were only posted on their website many months after publication of their paper. Knowles and Maddison (2002) simulated only one situation: a case in which every local population is completely isolated from all other local populations. This situation was called "microvicariance" by Templeton et al. (1995, p. 773), and was explicitly *excluded* from the inference key. Instead, a different test was published to deal with microvicariant fragmentation (Hutchison and Templeton, 1999). The inference key had been designed to yield an inconclusive inference under microvicariance, so the report by Knowles and Maddison (2002) that the inference key inferred the wrong historical process between 75–80% of the time is still troubling. However, a detailed examination of their simulation undercuts this conclusion.

First, Knowles and Maddison (2002) only ran 10 simulated replicates of their single scenario, thereby insuring that their results had no statistical significance. Most simulation studies involve hundreds to tens of thousands of replicates to ensure statistical validity, so the simulations of Knowles and Maddison (2002) were deficient by several orders of magnitude below the normal standards of the simulation literature.

Second, each isolate in their simulations had an inbreeding effective size of 10,000, and the time between fragmentation events that resulted in microvicariance was 5,000 generations within a total simulated time of 10,000 generations. Given that the expected coalescence time within each of their simulated local populations was 40,000 generations (see Basic Coalescent Theory, above), the parameter choices of Knowles and Maddison (2002) ensured the retention of much ancestral polymorphism across isolates. Inferring temporally shallow (relative to coalescent time) fragmentation events is difficult for any technique. Knowles and Maddison (2002) advocated a computer-simulation approach to phylogeographic inference (which will be discussed in more detail below), and the original version of their paper (provided by Knowles) focused on how well their computer-simulation test statistic performed on the simulated data. However, they evaluated their test statistic only under the simulated case of microvicariance and panmixia, and found that their statistic correctly discriminated in favor of microvicariance over panmixia. The incorrect inferences in the nested-clade analysis concerned discriminating between fragmentation and isolation by distance. Hence, a true comparison would involve discrimination between fragmentation and isolation by distance for both techniques. When this change was made in the manuscript of Knowles and Maddison (2002), they decided to drop the entire section on their statistic, and instead inserted a single sentence that their test also had "poor performance" with these simulated data sets. There is *no* statistic that can deal decisively with the difficult situation that they simulated, including their own.

Third, Knowles and Maddison (2002) made extremely unrealistic sampling assumptions. They assumed that only 10 individuals were sampled from each isolate of 10,000, but they assumed exhaustive sampling of all isolates. None of the actual data sets analyzed by nested-clade analysis corresponded to such a peculiar sampling situation (Templeton, 2004b). Templeton (2004b) redid all of the analyses reported in Knowles and Maddison (2002), but with the more realistic assumption that not all local populations were sampled. When this single change was made in the sampling assumptions, the 10 simulated data sets in Knowles and Maddison (2002) no longer yielded false-positive rates between 75–80%, but instead, 82% of the inferences were inconclusive (Templeton, 2004b). Hence, under sampling conditions that typify real data sets, the inference key performs as it should when dealing with microvicariance. Hence, the simulation studies of Knowles and Maddison (2002) actually showed the excellent performance of nested-clade analysis when dealing with a scenario outside the domain of the inference key.

Felsenstein (2004, p. 484) also criticized nested-clade analysis, claiming that it did "not attempt to take into account the uncertainty of the estimate of the tree." This criticism has no basis in fact. Nested-clade analysis was first developed for looking at genotype-phenotype associations (Templeton et al., 1987), and uncertainty in the estimate of the haplotype tree was explicitly incorporated into nested-clade analysis *before* it was ever applied to phylogeography (Templeton and Sing, 1993). Indeed, the very first nested-clade phylogeographic analysis (executed before the technique was completely formalized) was an analysis of human mtDNA that *explicitly* incorporated the uncertainty in the haplotype tree into the nested design (Templeton, 1993). *Every* nested-clade analysis performed by the author has incorporated tree uncertainty (when it exists). In particular, *all* data sets, without exception, that were analyzed for human phylogeographic inference were assessed for uncertainty in the haplotype tree. *Only* the inferences that were robust to that tree uncertainty were given (Templeton, 2002, 2005). Hence, the critique of Felsenstein (2004) is patently false.

Recently, Eswaran et al. (2005) cited Felsenstein (2004) as their sole support for a claim that nested-clade analysis "is regarded with skepticism by population geneticists." This assertion can be readily tested by examining the citation history of the basic paper outlining this method (Templeton et al., 1995). As of September 7, 2005, this paper was cited in the journal literature 331 times, with 207 of these citations being since 2002. These citations are mostly by authors who used this technique to analyze their data. Another way of accessing the accuracy of the statement by Eswaran et al. (2005) is to look at recent issues of the journal *Molecular Ecology*, the premier journal that publishes articles on intraspecific phylogeography. In the most recent issue at the time of this writing (volume 14, number 10, 2005), six articles were published on intraspecific phylogeography using haplotype data, and four of them used nested-clade analysis. This is typical for issues of this journal published over the last several years. Contrary to the claim of Eswaran et al. (2005), nested-clade analysis is one of the standard and most widely used analytical tools in the field of intraspecific phylogeography.

Despite its strengths, nested-clade analysis should not be regarded as the only tool to be used in intraspecific phylogeography. As illustrated in Table 1, inferences from nested-clade analysis can extend back to a time when there were just a handful of variable DNA lineages present, but it loses power as one approaches the present. Rare tip haplotype analysis and founder analysis are more appropriate for shallower times. Collectively, these three methodologies allow inferences from the recent, historical past back to ancient portions of a haplotype tree. This broad time range of inference is beyond the capabilities of any one of these analyses by itself.

Multilocus nested-clade analysis

The analyses described above collectively cover a broad time range, but they all share a serious limitation: all the

analyses described above are based on a single locus or DNA region (because of its lack of recombination, the entire mtDNA genome behaves as a single locus in evolution). Single-locus analyses have many severe restrictions on the inferences that can be made from them.

First, recall that the only branches in the coalescent process that are observable are those marked by a mutation. Mutations are required at the right place and time for population-level events or processes to have an observable effect on genetic variation and its spatial distribution. This means that any one DNA region will in general mark only a subset of the events and processes that have affected the evolutionary history of the species. Hence, any single haplotype tree provides at best an incomplete view of the species' evolutionary history, and the analysis of data sets with prior expectations reveals that the most common error of nested-clade analysis is the failure to detect an event (Templeton, 2004b). Human evolutionary history is not the same as the evolutionary history of the haplotype variation found in mtDNA or any other DNA region. A fuller view of human evolutionary history requires that many DNA regions be studied.

Second, the frequency and distribution of haplotypes at a particular locus may have been skewed by natural selection. If the focus of one's study is on the historical events and gene-flow patterns that have influenced human evolution, such loci may be misleading. Selection tends to operate in an idiosyncratic fashion at each locus, so patterns induced by selection tend to be locus-specific. However, if an inference is based on a single locus or DNA region, it is difficult to separate the idiosyncratic features induced by selection from the more universal signals associated with a shared evolutionary history and demographic processes.

Third, the coalescent process itself is random, so even a neutral DNA region can sometimes produce unusual patterns just by chance. As with selection, this means that the inferences drawn from a single locus can sometimes be misleading, but typically in an idiosyncratic fashion.

Fourth, coalescent theory predicts, and observations confirm (Fig. 3), that there is much variation in the time to coalescence from one DNA region to the next. The phylogenetic information in a haplotype tree is generally exhausted sometime between half of the TMRCA and the TMRCA (Templeton, 2002), which varies substantially from locus to locus (Fig. 3). On the other end of the time scale, a haplotype tree can only capture information marked by a mutation, and these mutational markers become less and less likely on coming closer to the present, but are more likely for those DNA regions with higher mutation rates. Hence, every DNA region has its own unique and limited temporal window into the past.

Fifth, nested-clade analysis requires multiple tests of the null hypothesis of no geographical associations, but biological inferences come from the patterns of several tests. This makes it difficult to do a multiple-test correction.

A multilocus analysis with integrated and cross-validated inference addresses all five of these limitations. With multiple loci, it is much more likely that an event or process had been marked by an appropriately timed and placed mutation at some loci, so there will be fewer false negatives. Also, multiple loci can collectively cover a much broader time range than just one locus, so the temporal breadth of the analysis is greatly augmented. Requiring inferences to be cross-validated by two or more loci reduces the rate of false positives that can arise either by chance in the coalescent process or by selection operating upon a particular locus. Cross-validation is also a standard, and often more powerful, substitute for multiple-test correction. Consequently, multiple-locus analysis provides a fuller and more reliable inference structure than is possible with any one single-locus or DNA region.

Templeton (2004a) provided a formal maximum likelihood framework for cross-validation of inferences when nested-clade analyses are applied to multiple loci. An inference is concordant across loci when more than one locus infers the same type of event or process (e.g., a population range expansion) in the same locations (e.g., a range expansion out-of-Africa into Eurasia) at the same time (e.g., an out-of-Africa range expansion 100,000 years ago). Judging the first two types of concordance is straightforward as they are categorical variables, but temporal concordance requires a quantitative framework. The age of mutations or haplotypes that mark events or processes is commonly estimated using phylogenetic approaches and the assumption of the molecular clock, but often these estimates do not reflect the considerable evolutionary stochasticity of the coalescent process itself. To incorporate this source of error, Templeton (1993) used earlier theoretical work (Kimura, 1970; Tajima, 1983) to show that the age of a node in a haplotype tree can be regarded as a random variable with a gamma probability distribution function with mean given by T, the standard phylogenetic estimator of age of the node, and variance given by:

$$\sigma^2 = \frac{T^2}{(1+k)} \tag{8}$$

where k is the number of mutations that have accumulated in the descendants of the node whose age is estimated to be T. Equation 8 incorporates two sources of error into the variance associated with T. First, the numerator of Equation 8 is T^2, reflecting the evolutionary stochasticity of the coalescent process itself, in which the variance is proportional to the square of the mean (Hudson, 1990; Donnelly and Tavare, 1995). The other factor that influences the variance is k, the number of mutations that are used to age the event. Because k is generally very small for recent events, phylogenetic dating procedures are often unreliable for recent events (Rannala and Bertorelle, 2001). Accordingly, this approach will only be applied to events in human evolutionary history that are about 100,000 years old or older.

Equation 8 is applicable only to species that have been evolving as a single evolutionary lineage, with no long-term fragmentation events or isolates that have contributed to the present-day population. This assumption is justified for humans over our recent evolutionary past (Templeton, 2002). Given that the estimated ages can be treated as gamma-distributed, a maximum likelihood framework can be created for testing many hypotheses, including temporal concordance across loci. When the hypothesis of concordance is not rejected, this same framework can be used for estimating ages and confidence intervals based on multiple loci (Templeton, 2004a).

Templeton (2005) performed nested-clade phylogeographic analyses upon the 25 loci or DNA regions indicated in Figure 3, followed by cross-validation of inferences with the maximum likelihood testing framework. The inferences older than 100,000 years, using a 6 MYA calibration for the divergence of chimpanzees and humans, fell into two types. First, the 25 loci collectively yielded 15 inferences of range expansion involving African and Eurasian populations (Fig. 7). The log likelihood ratio test rejects
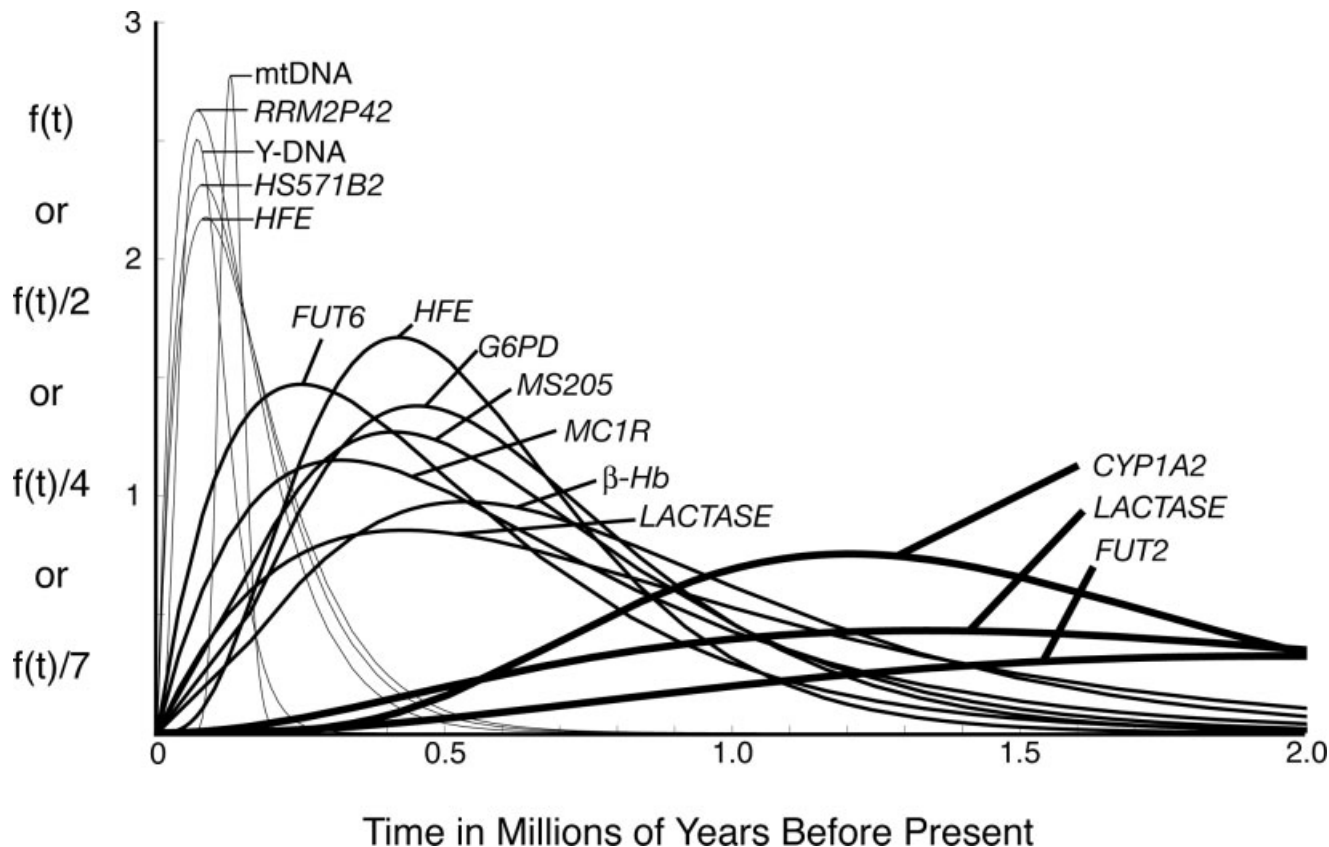
Fig. 7. Gamma distributions for ages of range expansion events involving Africa and Eurasia, all of which are out-of-Africa events when geographical origin is unambiguous. X-axis gives age in millions of years before present; y-axis gives gamma probability distribution, f(t), fitted to data from particular locus or DNA region. Because probability mass is so concentrated close to y-axis for several genes, gamma distribution was divided by seven for mtDNA, by four for Y-DNA, and by two for *HFE*, *HS571B2*, and *RRM2P4* to yield better visual presentation. Age distributions fall into three clusters, shown by thin lines, medium lines, and thick lines, respectively.

the null hypothesis that all 15 events are temporally concordant, with a probability value of 3.89×10^{-15}. An inspection of Figure 7 reveals that the time distributions for the 15 events cluster into three distinct groupings. Accordingly, the null hypotheses of temporal concordance within each of these three groupings were tested, and in all cases the null hypothesis of temporal concordance could not be rejected ($P = 0.95$ for the most recent expansion out-of-Africa, $P = 0.51$ for the middle expansion, and $P = 0.62$ for the oldest expansion). Hence, multilocus nested-clade analysis indicates that at least three expansions occurred out-of-Africa into Eurasia, at 1.9 MYA (with a 95% confidence interval of 0.99–3.10 MYA), 0.65 MYA (95% confidence interval of 0.39–0.97 MYA), and 0.13 MYA (95% confidence interval of 0.096–0.169 MYA). Each of these expansion events is cross-validated by multiple loci or DNA regions (Fig. 7). These dates overlay well upon the fossil and archaeological record, with the first expansion corresponding to the original expansion of *Homo erectus* out of Africa into Eurasia (Aguirre and Carbonell, 2001; Bar-Yosef and Belfer-Cohen, 2001; Antón et al., 2002; Vekua et al., 2002) and the development of a culture capable of keeping impaired individuals alive for many years (Lordkipanidze et al., 2005). The second expansion corresponds to the spread of Acheulean culture into much of Eurasia after an earlier African origin (Asfaw et al., 1992; Hou et al., 2000) and the initiation of a substantial increase in cranial capacity (Ruff et al., 1997; Relethford, 2001b;

Rightmire, 2004). The most recent expansion out of Africa corresponds to the spread of several anatomically modern traits into Eurasia after an earlier African origin (Stringer, 2002; White et al., 2003).

There were also 19 inferences of gene flow, mostly restricted by isolation by distance, between Africa and Eurasia (Fig. 8). Because gene flow is a recurrent process rather than an event, there is no expectation that all the time estimates should be homogeneous across loci. In this case, cross-validation can be measured by degree of overlap of the gamma distributions across loci. Only one locus, *MX1*, which was an outlier in the original and more limited multilocus nested-clade analysis (Templeton 2002), has the bulk of its probability mass in the Pliocene (not shown in Fig. 8), whereas all other gamma distributions show extensive overlap across loci throughout the Pleistocene. In contrast to the 15 inferences of expansion events (Fig. 7), the 18 inferences of Pleistocene gene flow show no clustering (Fig. 8). This continuum implies that gene flow restricted by isolation by distance was a recurrent evolutionary force throughout the Pleistocene, with no lengthy interruptions. The 18 cross-validated loci indicating Pleistocene gene flow jointly indicate that such African/Eurasian gene flow occurred as far back as 1.46 MYA with 95% confidence (Templeton, 2005). Thus, recurrent gene flow between African and Eurasian populations goes back to the Lower Pleistocene and was established at or shortly after the initial spread of *Homo erectus* out of Africa.

Time in Millions of Years Before Present
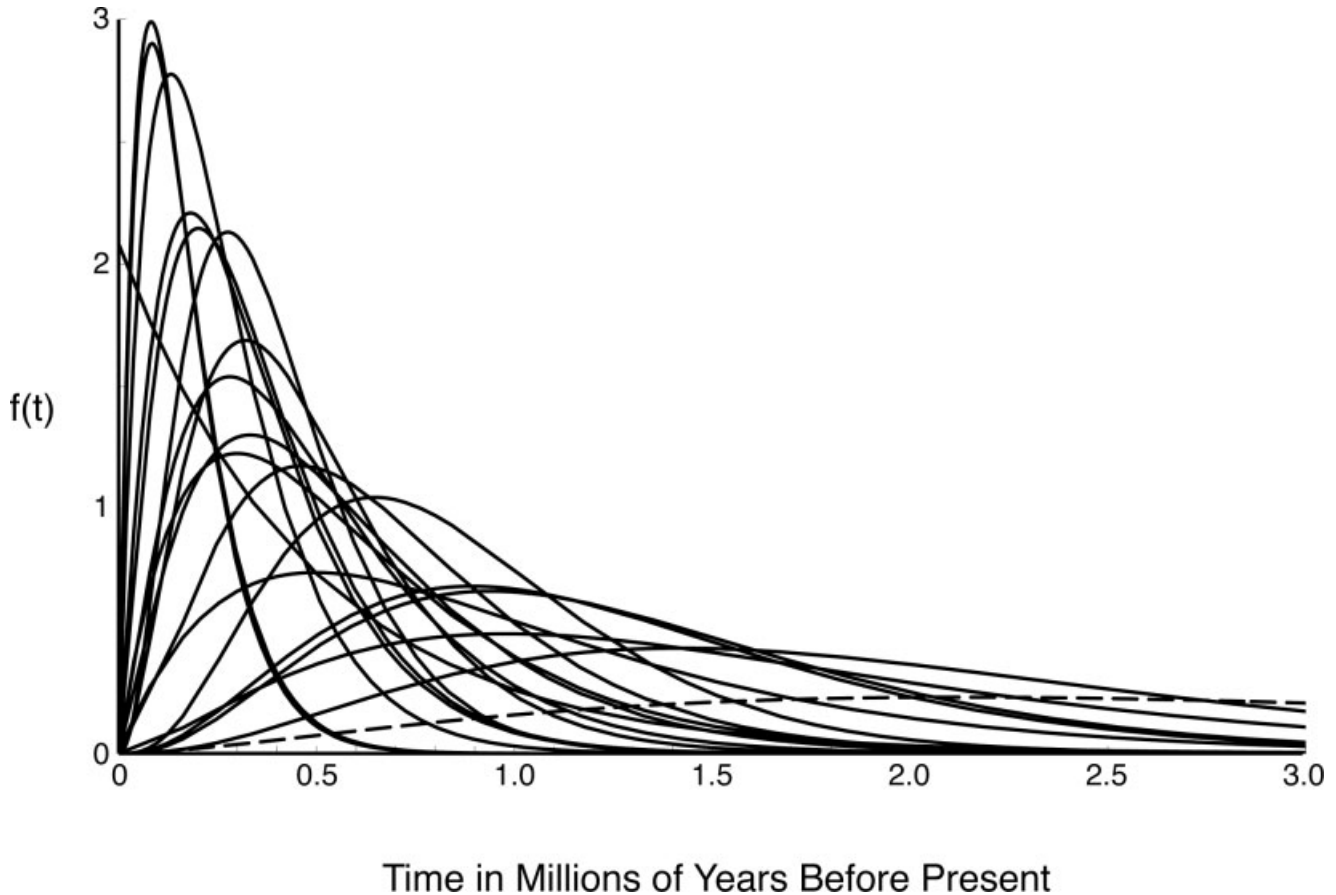
Fig. 8. Distributions for ages of youngest clade contributing to significant inference of restricted gene flow, primarily with isolation by distance. X-axis gives age in millions of years before present; y-axis gives gamma probability distribution, f(t). Genes or DNA regions yielding these distributions are, as ordered by their peak values of f(t) going from left to right: Xq13.1, *MSN/ALAS2*, *HFE*, *FIX*, *HFE*, *G6PD*, β*Hb*, *ECP*, *RRM2P4*, *EDN*, *PDHA1*, *CYP1A2*, *FUT2*, *FUT6*, *FUT6*, *FUT2*, *CYP1A2*, *CCR5*, and *MX1*. Curve for *MX1* is shown as dashed line, to emphasize its outlier status (Templeton, 2002). Several other inferences of restricted gene flow that were too recent to date phylogenetically are not shown, and are not used in analyses.

A number of additional hypotheses can be tested using the distributions shown in Figures 7 and 8. For example, there has been much controversy in the literature as to whether out-of-Africa expansions were accompanied by the total genetic extinction of contemporary Eurasian populations and their replacement by the expanding African population. Because all 25 loci were sampled only from living humans, all inference is limited to past human populations that left descendents today. Therefore, if total replacement had occurred in the past, any earlier signal involving Eurasian populations would have been erased. The Acheulean expansion is no older than 1.05 MYA with 99% confidence, using the cross-validated confidence interval. The null hypothesis that none of the inferences involving Eurasian populations (Figs. 7, 8) are older than 1.05 MYA is rejected with $P = 0.035$ when *MX1* is excluded (Templeton, 2005). Hence, the expanding African populations with the Acheulean culture did not totally replace the Eurasian populations they encountered, but rather interbred with at least some of them. Similarly, the most recent out-of-Africa expansion is no older than 0.177 MYA with 99% confidence. Excluding *MX1*, the null hypothesis that no other event or process involving Eurasian populations is older than 0.177 is rejected with a probability of less than 10^{-17}. Hence, the recent out-of-Africa replacement hypothesis is overwhelmingly rejected.

Figure 9 summarizes the cross-validated, statistically significant conclusions from the nested-clade analysis of 25 genes or DNA regions. As seen in Figure 9, the breadth and precision of inference from a multilocus analysis far exceed that possible with any single locus.

The strong rejection of the recent out-of-Africa replacement hypothesis is particularly noteworthy, as this hypothesis has dominated much of the discussion of recent human evolution over the last couple of decades (Stringer and Andrews, 1988; Vigilant et al., 1991; Stringer, 2002). Moreover, a recent analysis claimed that the genetic data "clearly" support the out-of-Africa replacement hypothesis (Ray et al., 2005). The basis of this discrepancy stems from fundamental issues in statistical philosophy, as debated by Knowles and Maddison (2002) and Templeton (2004b). Knowles and Maddison (2002) advocated a computer-simulation approach to phylogeographic inference, and Table 2 summarizes the differences between the multilocus nested-clade and the simulation approach. Both procedures have legitimate uses in phylogeographic inference (Templeton, 2004b), with nested-clade analysis being appropriate when a prior model of evolutionary history is not known in detail, and the simulation approach being appropriate when one wants to examine a small set of possible, detailed alternatives. The inference universe of the simulation approach is completely limited to this finite
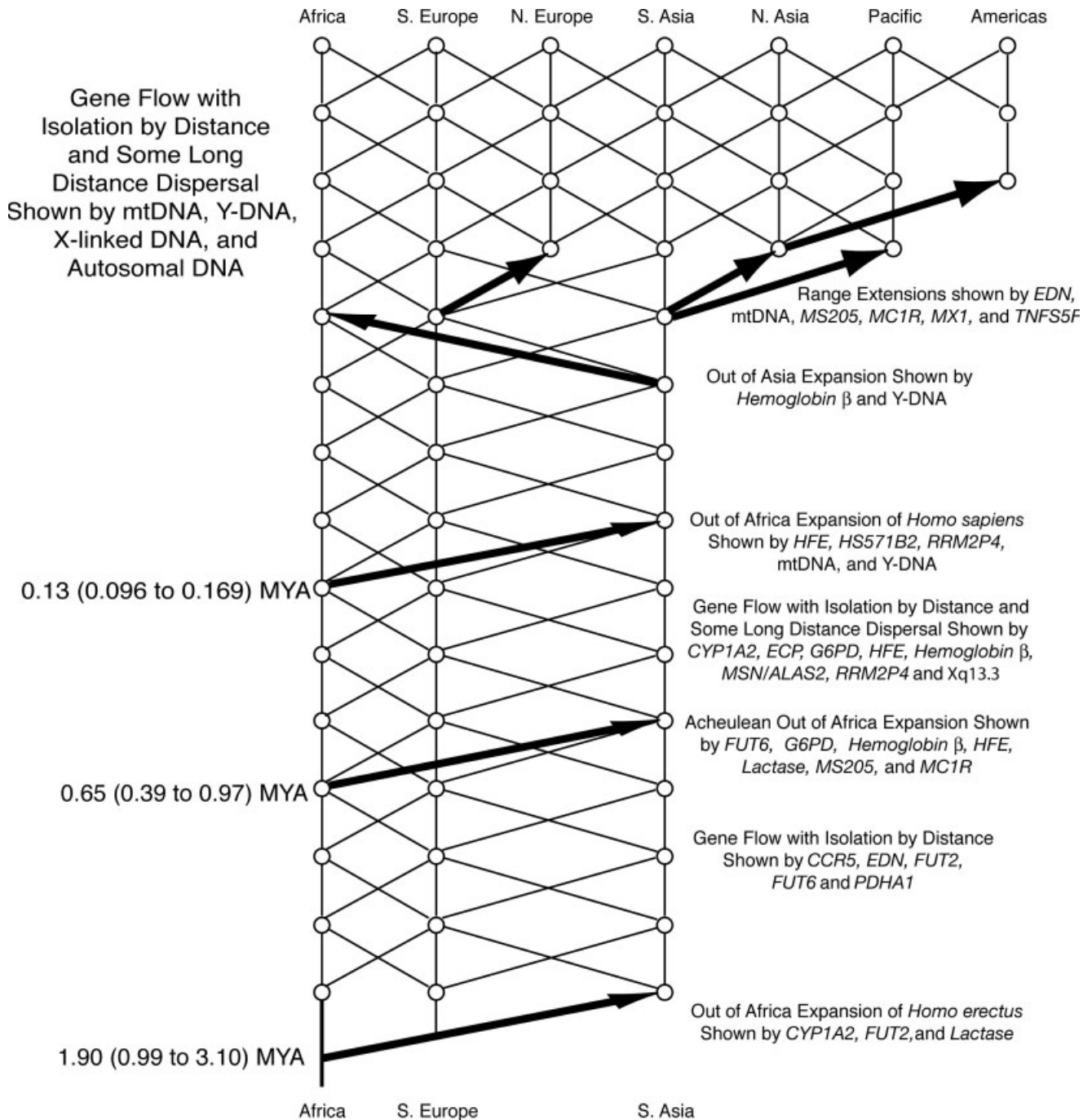
Fig. 9. Model of recent human evolution, as inferred from nested-clade analysis. All inferences are cross-validated by two or more genes, and are statistically significant at least at 5% level. Major expansions of human populations are indicated by arrows, and time periods for out-of-Africa are 95% confidence limits. Genetic descent is indicated by vertical lines, and gene flow by diagonal lines. Events or processes less than 100,000 years old are not dated.

set. The importance of this set was shown earlier by the statistic given in the original version of the manuscript by Knowles and Maddison (2002) (Knowles, personal communication) that discriminated well in favor of fragmentation when the inference universe was limited to fragmentation and panmixia, but had "poor performance" when the inference universe was expanded to include isolation by distance. Hence, what appears to be excellent discrimination among alternatives can be transformed into no discrimination simply by expanding the universe of alterna-

tive models. Unless extensive prior information exists that only a finite set of detailed alternatives is at all plausible, the inferences based on computer simulation have no general validity.

The number of models and debates over them in the recent literature indicate that such detailed prior information does not exist for recent human evolution. In this regard, Ray et al. (2005) did not consider a model with an Acheulean expansion or a model with out-of-Africa expansion coupled with some interbreeding, even though both

TABLE 2. Contrast between multilocus nested-clade approach and computer-simulation approach to intraspecific phylogeographic inference[1]

Property	Multilocus NCA	Computer simulation
Choosing phylogeographic model	No prior model is specified. Final inferred model emerges from data analysis.	Finite set of prior models is specified in great detail. Final model is limited to this set.
Biological inference	Biological inferences are drawn from statistically significant geographical associations, using explicit a priori criteria drawn from coalescent theory.	Computer simulations are used to determine relative goodness of fit of data to finite set of prior biological models, favoring model with highest relative fit.
Procedure for dealing with inference error	Individual inferences are cross-validated across loci for type, location, and time.	Goodness-of-fit statistic distribution is determined through replicate simulations.
Nature of hypotheses tested	Hypotheses are phrased as null hypotheses within formal maximum-likelihood framework that can be rejected with quantifiable probabilities.	Goodness-of-fit criteria lack quantifiable probabilities for rejecting alternatives or models not in original a priori inference set.

[1] NCA, nested-clade approach.

were indicated by prior genetic analyses (Relethford, 2001b; Templeton, 2002), and they recognized these models as being potentially more realistic. Hence, the model shown in Figure 9 lies completely outside the inference universe of Ray et al. (2005), so none of their results are relevant to the relative merits of the out-of-Africa replacement model vs. the model shown in Figure 9. In contrast, inferences are data driven in the multilocus nested-clade analysis and not limited by prior beliefs. For example, the Achuelean expansion was not part of any of the major models of recent human evolution that have been debated for the last 20 years, yet this completely unexpected inference emerged from the nested-clade analysis and overlies well upon the fossil and archaeological data.

The computer-simulation approach also requires detailed information *within* each of the plausible alternatives. Ray et al. (2005) claimed that they used "an approach based on realistic simulations." For example, their simulations of the out-of-Africa replacement hypothesis assume that the total human population size between 900,000–120,000 years ago was 20,000. They obtained this "realistic" number from estimates of the inbreeding effective size of humans (discussed below), but "realistic" simulations should be based on census sizes, not effective sizes. The inbreeding effective size of Pleistocene human populations is compatible with a census size of 300,000 (Eller, 2002), a figure more consistent with a nongenetic estimate of around half a million individuals during the Pleistocene (Weiss, 1984). So what is the "realistic" number for the total human population in Africa and Eurasia: 20,000 or 500,000? A large number of other demographic parameters have to be specified in order to perform the simulations, and even less information exists about these parameters. Given this lack of knowledge, it is impossible to define what is "realistic." Moreover, the goodness of fit of the alternative models can be measured in several ways, and it is not always obvious which is the most appropriate. For example, Ray et al. (2005) measured the fit of their prior models to the data through the pairwise genetic distance statistic R_{ST}. Interestingly, two species of African ungulates, sampled at the same locations, yield statistically indistinguishable values of a closely related statistic, F_{ST}, but had completely different phylogeographic histories under nested-clade analysis (Templeton, 1998). Hence, the goodness of fit of alternative models could be extremely sensitive to the statistic used to measure that fit. Do these details matter? A strong indication that such details do indeed matter is provided by contrasting the work of Ray et al. (2005) with that of Eswaran et al. (2005). Both sets of authors claimed to test the out-of-Africa replacement hypothesis vs. a multiregional model with restricted gene flow with the computer-simulation approach. Both claimed that the simulations clearly discriminate between these two hypotheses. The trouble is that Ray et al. (2005) claimed that they "unambiguously distinguish" in favor of the out-of-Africa replacement model over the multiregional model with gene flow, and Eswaran et al. (2005) claimed that their simulations show that "genomics refutes an exclusively African origin of humans" and that "living human populations carry within them a substantial genetic inheritance that had its origins in non-African archaics." So whose simulations do we believe? The two sets of simulations differed in many of the detailed demographic parameters that are required to execute a simulation, differed in the statistics used to measure goodness of fit, and differed in the data they considered. Any one of these factors or any combination could be behind their opposite conclusions. What is clear is that the computer simulations considered in toto do not discriminate between the out-of-Africa replacement model and the multiregional model with restricted gene flow, and they certainly do not discriminate between either of these models with the model shown in Figure 9, which lies completely outside both of their inference spaces. The contradictory results of Ray et al. (2005) and Eswaran et al. (2005) collectively have little value in elucidating recent human evolution.

Both Ray et al. (2005) and Eswaran et al. (2005) also claimed to be testing the out-of-Africa replacement hypothesis, but neither paper contains a test of this hypothesis. Ray et al. (2005) showed that their out-of-Africa *simulation* explains four times more variance in population differentiation than the best of their *simulated* multiregional models, but what does this mean? Is this result significant at the 5% level, the 1% level, etc.? Despite their claims to quantify the "likelihood of this model," no likelihoods are given, no *P*-values are given, and no actual test of the out-of-Africa replacement model is given. All that is given is the difference in goodness of fit along with a subjective, nonquantitative assessment of the significance of these differences. The same is true for Eswaran et al. (2005). In contrast, the multilocus nested-clade analysis tests the

out-of-Africa replacement model as a null hypothesis with an explicit likelihood ratio test, yielding a chi-square statistic of 118.18 with 17 degrees of freedom and a P-value of less than 10^{-17} (Templeton, 2005). Despite the verbal claims, the computer-simulation approaches have *not tested* the out-of-Africa replacement hypothesis in any meaningful statistical sense, and they are *incapable* of defining the model that best describes recent human evolutionary history because they only look at a small number of prior possibilities, while excluding many realistic alternatives from their inference space. Multilocus nested-clade analysis can both test hypotheses through the well-defined and accepted framework of maximum likelihood, and allow the data, rather than prior belief, to define the model that best describes recent human evolution (Fig. 9).

HAPLOTYPE TREES WITH ANCIENT DNA

The inferences given above are all based on samples of current DNA, and as such, they are limited to past populations that have left genetic descendants in present-day humanity. This feature allowed the replacement hypotheses to be tested, and both a recent out-of-Africa replacement and an Acheulean replacement were rejected. In both cases, this is a rejection of the hypothesis of *total* Eurasian replacement, and allows the possibility that some Eurasian populations did indeed go to genetic extinction. One way of addressing the question of whether or not a particular population went extinct is to sample their DNA directly from fossils. This approach was applied to the possibility of replacement of European Neanderthals by more anatomically modern forms between 30,000–40,000 years ago.

Working with ancient DNA is difficult. Because mtDNA is much more abundant than nuclear DNA, almost all DNA so far obtained from Neanderthal fossils is mtDNA. Ancient DNA is subject to damage over time, and resulting lesions can create mutational artifacts (Caldararo and Gabow, 2000; Hansen et al., 2001). One test for artifacts makes use of the considerable age range found in the Neanderthal fossils used as sources for DNA. If the apparent divergence is real, then the oldest Neanderthal samples should tend to be closest to current human mtDNA because they are temporally closer to the common ancestral sequence for Neanderthal and modern human mtDNA. In contrast, if DNA damage has made a large contribution to the apparent divergence, then the oldest Neanderthal sequences should be the furthest from those of modern humans. The latter pattern is true (Gutierrez et al., 2002). Because the samples are small, one could argue that just by chance the oldest Neanderthal sequences just happened to come from an abnormally highly divergent lineage of Neanderthal mtDNA, but these results indicate that DNA damage cannot be discounted as a significant source of error in these studies. In addition, ancient DNA extracts induce mutational artifacts in a nonrandom fashion, such that the same artifacts are independently created in controlled experiments (Pusch and Bachmann, 2004). Many of the sites at which these artifacts repeatedly occur are the same sites observed in Neanderthal mtDNA divergence (Pusch and Bachmann, 2004). These results indicate that great caution should be exercised in interpreting ancient DNA sequence data.

When Neanderthal mtDNA is added on to the contemporary human mtDNA haplotype tree, using chimpanzees as an outgroup, the oldest branching event is a split between a clade containing all the Neanderthal mtDNA and a clade containing all modern human mtDNA. This pattern was interpreted as meaning that there was no or extremely little interbreeding between Neanderthals and their more anatomically modern contemporaries, and that Neanderthals perhaps represent a distinct species from *Homo sapiens* (Krings et al., 2000; Caramelli et al., 2003; Knight, 2003; Currat and Excoffier, 2004; Serre et al., 2004; Lalueza-Fox et al., 2005). However, there are difficulties with these conclusions.

First, the degree of divergence between modern human and Neanderthal mtDNA is much smaller than the corresponding divergence between subspecies of chimpanzees (Curnoe and Thorne, 2003), making the elevation of Neanderthals into a separate species on the basis of mtDNA highly questionable. Second, and more importantly, the tree topology indicating separate Neanderthal and modern human clades is based on only a few sites, because only small fragments of fossil mtDNA have been studied. It is therefore critical to make a statistical assessment of the estimated tree vs. alternatives that place Neanderthal and modern human mtDNA into a single clade. Gutierrez et al. (2002) performed such a statistical assessment, and discovered that there was no significant discrimination between these two hypotheses of tree topology. Moreover, Gutierrez et al. (2002) showed that 27% of the human-Neanderthal pairwise genetic differences were lower than the largest modern human-modern human genetic difference, indicating that Neanderthal sequences are not so different from those of extant humans. The lack of any statistically significant evidence that Neanderthal mtDNA was a separate clade indicates that it is premature to elevate Neanderthals to a separate species on the basis of fossil DNA.

Regardless of the topological position of Neanderthal mtDNA in the human mtDNA haplotype tree, Neanderthal mtDNA is nevertheless distinct and therefore potentially informative about interbreeding between Neanderthals and more modern-looking human populations. The failure to find Neanderthal-like sequences in contemporary and fossil modern humans is often interpreted to mean that there was no interbreeding with Neanderthals. This conclusion is also premature.

First, mtDNA is incapable biologically of completely reflecting a population's evolutionary history and of rejecting the hypothesis of admixture. MtDNA is sensitive to only female-mediated gene flow, and can totally miss even extensive interbreeding mediated through males. For example, Figure 9 shows a cross-validated expansion out-of-Asia that occurred after the most recent out-of-Africa event. This expansion event is marked by autosomal DNA and Y-DNA (Templeton, 2002). This expansion event was not detected by mtDNA, even though the largest samples, the best geographical coverage, and the greatest genetic resolution existed for mtDNA. This implies that this expansion event was primarily mediated by males coming out of Asia and was invisible to mtDNA. The Lemba, as discussed earlier, provide a more recent example. In this case, Y-DNA provides evidence for extensive interbreeding between Jewish and Bantu populations, but mtDNA indicates that there was no interbreeding at all, a pattern consistent with their oral history (Soodyall, 1993; Thomas et al., 2000; Wilson and Goldstein, 2000). Thus, mtDNA alone is biologically incapable of detecting all admixture events, even those such as the Lemba in which the degree of admixture was 100% (given that their origin was due to Jewish men mating with Bantu women).

Second, the evolutionary history of a single gene or DNA region should never be equated to the evolutionary

history of a population. One needs multiple loci to obtain an accurate reconstruction of evolutionary history (Wall, 2000) and to protect against false inferences due to evolutionary stochasticity and natural selection skewing the results of a particular gene (Templeton, 2002, 2004a,b). In this regard, human mtDNA is known to have been subject to natural selection (Templeton, 1996; Curnoe and Thorne, 2003; Elson et al., 2004), and this provides alternative interpretations to the fossil mtDNA patterns. For example, suppose that there was much female-mediated gene flow between Neanderthals and modern humans, but that in addition there was selection against the Neanderthal mtDNA haplotypes in a hybrid or predominantly modern human nuclear background, a phenomenon known as cytonuclear incompatibility. In such a situation, even high levels of interbreeding could go undetected by mtDNA.

Third, the small sample sizes preclude the ability to dismiss significant amounts of gene flow between Neanderthals and moderns (Nordborg, 1998; Wall, 2000; Relethford, 2001a; Currat and Excoffier, 2004; Pearson, 2004). Currat and Excoffier (2004) produced the smallest estimate of the amount of female-mediated gene flow compatible with the mtDNA, assuming neutrality, with an upper limit of 0.1%. However, within population genetics, gene flow is regarded as significant if the product of the population size times the migration rate is equal to or greater than one (Crow and Kimura, 1970). Thus, if Neanderthals had a population size equal to or greater than 1,000, there could be biologically significant gene flow. The simulations used by Currat and Excoffier (2004) to obtain the limit of 0.1% started with a Neanderthal population size of 35,000. Hence, even this conservative evaluation indicates that a biologically significant, female-mediated gene flow of neutral mtDNA cannot be rejected.

The ancient mtDNA results are therefore suggestive of replacement of Neanderthals, but are not conclusive. This inconclusive state will persist until ancient DNA technology can deal adequately with genetic artifacts and can be used to study multiple genes rather than just mtDNA, and until larger sample sizes are available. Progress on these issues is being made (Ricaut et al., 2005; Willerslev and Cooper, 2005), so there is much potential for ancient DNA studies contributing to our understanding of human evolution.

USING HAPLOTYPE TREES FOR DEMOGRAPHIC ANALYSES

As shown above in Basic Coalescent Theory, many measurable attributes, such as heterozygosity levels, can be related to underlying demographic parameters, such as inbreeding effective size. Hence, haplotypes can also be used to make inferences about some demographic attributes of past human populations. Most of these applications do not use the haplotype tree per se, but rather use nucleotide differences such as expected heterozygosity or pairwise differences among haplotypes. Accordingly, these techniques will not be reviewed in detail here, but they are in Templeton (2005). Some brief comments will be made, however, because the inferences from these approaches interface with the inferences made from haplotype trees.

Two main conclusions have emerged from initial analyses of paleodemography based on mtDNA: that there was a sudden increase in population size sometime between 30,000–130,000 years ago (Harpending and Rogers, 2000),

and that the inbreeding effective size of the ancestral population before this expansion was around 10,000 (Hawks et al., 2000). However, the recent literature has clouded the validity of both of these conclusions.

First, many of the demographic analyses assume the infinite-sites model, an assumption much more restrictive than the infinite-alleles model used for haplotype trees. Initially it was argued that deviations from the infinite-sites model did not seriously bias demographic estimates (Rogers et al., 1996), but more recent work indicates that the bias can be more severe (Yang, 1997; Schneider and Excoffier, 1999).

Second, many of the coalescent estimators are based on simulating various demographic models and measuring the fit to the data. Thus, the inference universe is limited to the scenarios that were simulated. Even an excellent fit to the data does not insure that the simulated model is the right one, because often several models can fit the data well. For example, Adams and Hudson (2004) pointed out that one of the African populations they studied fit both models of constant population size *and* a variety of growth models, including fivefold growth beginning no earlier than 36,000 years ago. The real weakness of the simulation approach is that it is never possible to simulate all possible scenarios, so many alternative hypotheses are never evaluated at all. For example, none of the simulations performed in the papers referred to here take into account the model shown in Figure 9. Thus, the demographic scenarios with the most compelling genetic evidence have never even been considered in these simulations.

Third, the statistics used for demographic inference are also sensitive to natural selection (Tajima, 1989a,b; Harpending and Rogers, 2000). When inference is limited to coding synonymous nucleotide polymorphisms and noncoding DNA regions (the mutational classes least likely to be subject to selection), the analyses for population growth imply little or no change in population size (Wooding and Rogers, 2002). Another method to circumvent the confoundment of selection with demography is to use cross-validation across multiple loci, just as was done above with nested-clade analysis. The mtDNA inference of a population size expansion between 30,000–130,000 years ago is cross-validated by some nuclear loci (Marth et al., 2003, 2004), but others do not (Harpending and Rogers, 2000). An analysis of 10 noncoding DNA regions (to minimize selection) found no evidence for significant population size expansion in Africa, and only nominal significance (without correcting for multiple testing) in Eurasian samples (Pluzhnikov et al., 2002). Thus, the evidence for a small ancestral human population size followed by massive growth is mixed, and a coherent picture has yet to emerge (Pluzhnikov et al., 2002; Wooding and Rogers, 2002). No single expansion or bottleneck can explain all the genetic data.

The estimate of an inbreeding effective size of 10,000 has also become clouded in the recent literature, despite the tendency of some to regard this figure as a given constraint on human evolution (Pearson, 2004). First, many of the studies have a strong sampling bias for Eurasians, but studies having a large sample of Africans tend to have much larger estimates of ancestral human population size (Tishkoff and Verrelli, 2003), even to the extent of showing no significant or moderate population size growth from the past to the present within Africa (Pluzhnikov et al., 2002; Adams and Hudson, 2004). Second, the figure 10,000 is an estimate of inbreeding effective size, not census size. There

is no expectation for the inbreeding effective size to equal the census size. For example, Eller (2002) and Eller et al. (2004) showed that a human census size of approximately 300,000 through the Pleistocene is compatible with an inbreeding effective size of 10,000. Thus, no coherent picture has emerged from genetic studies concerning the size of Pleistocene hominin populations.

USING HAPLOTYPE TREES FROM NONHUMAN SPECIES

Humans have not evolved independently of other species, so in some cases inferences made from the haplotype trees of other species can shed light on human evolution. What is needed is another species whose evolution is strongly coupled to that of human evolution. One type of evolutionary coupling is through an obligate parasite or pathogen/host relationship. For example, lice are obligate parasites of mammals or birds that complete their entire life cycle on the body of the host and cannot survive more than a few days off the host (Reed et al., 2004). As a consequence, the evolutionary patterns of lice should be strongly coupled with the evolutionary patterns of their hosts, and hence provide an independent marker of host evolution. A major caveat emerges from more general studies that indicate that coevolutionary patterns can vary over space and time, are influenced by gene flow and dispersal patterns of both host and parasites (including host shifts for the parasites), and depend on other species in the ecological community (Pellmyr, 2003; Perlman and Jaenike, 2003; Shingleton and Stern, 2003; Zangerl and Berenbaum, 2003; Forde et al., 2004; Segraves and Pellmyr, 2004). The idea of a simple, parallel phylogenetic pattern is unrealistic is many cases. As a consequence, parasite phylogenetic studies must be interpreted with great caution, and must always be regarded at best as *indirect* evidence of host evolution.

Useful evolutionary inferences are possible even when the phylogeny of parasites is obviously discordant with the phylogeny of the host. For example, Glenn and Brooks (1986) performed a phylogenetic analysis on the presence or absence of particular parasites on a variety of hominid and other primate hosts, and discovered that humans are more similar to baboons (*Papio* spp.) than to chimpanzees or gorillas. This conclusion is obviously incorrect phylogenetically, and shows the danger of equating parasite evolution to host evolution. The parasites found on a particular host are due to a combination of vertical transmission (parallel evolutionary descent) and horizontal transfer (jumping from one host to another). Horizontal transfer often reflects the types of habitats a species uses and lives in and its diet, particularly for intestinal parasites (Ashford, 2000). Hence, the results of Glenn and Brooks (1986) do not reflect hominid phylogeny, but do indicate that early humans lived in habitats and had a diet more similar to that of baboons than to that of our closest evolutionary relatives, the chimpanzees and gorillas.

The strongest coevolutionary patterns within the human lineage are expected for those parasites that are human specialists. Human specialists whose geographical origin can be inferred are predominantly from sub-Sahara Africa, consistent with the inference shown in Figure 9 of an African origin for the human lineage (Ashford, 2000). However, a considerable group of human specialists is primarily Eurasian, indicating that the human lineage has been outside of Africa for a long time in a temporally continuous fashion (Ashford, 2000), a conclusion also consistent with

Figure 9. Interestingly, five human specialists depend on our eating of beef or pork, and all of these appear to be of Eurasian origin, suggesting that humans became specialist cow and pig eaters first in Eurasia (Ashford, 2000). One group of human specialists, three species of Taeniid tapeworms, were initially thought to fit into this pattern. However, a more recent study that estimated an mtDNA haplotype tree of 35 species in the genus *Taenia* indicated that these three human specialists evolved from two independent host shifts to humans from tapeworm species that used carnivores (hyaenids, canids, or felids) and bovids (but not from the genus *Bos*) as hosts (Hoberg et al., 2001). The molecular dating of one of these transfers to humans was probably older than a million years ago. Even though there is considerable error in this figure, it is clear that this transfer to human hosts occurred well before the domestication of cattle or pigs. Hence, Hoberg et al. (2001) concluded that humans around a million years ago either hunted or scavenged bovids such as antelope and thereby acquired these tapeworm parasites, which were then subsequently transferred from humans to domestic cattle and pigs.

One interesting phylogenetic pattern that emerged from the mtDNA haplotype tree of the Taeniid tapeworms is that one of the host shifts to humans resulted in two species of human specialist tapeworms that split from one another about a million years ago, with one form being principally confined to Asia (Hoberg et al., 2001). Reed et al. (2004) reported a split pattern for lice. They studied the evolution of head/body lice (*Pediculus humanus*) that exist in two morphologically similar but ecologically distinct forms, head lice that are confined to the scalp and body lice. The molecular date of the split of the human louse from its sister species that uses chimpanzees as a host is 5.6 MYA, a figure consistent with the hypotheses of cospeciation of the lice with their hominoid hosts. Two highly divergent clades were found within *P. humanus* that split around 1.18 MYA, with one clade being found worldwide and consisting of both head and body lice, and the other clade consisting of only head lice and confined to the American continents. Similarly, Pavesi (2004) reported two clades of polyomavirus JC (JCV), a double-stranded DNA virus that is ubiquitous in human populations, with the Asian/Amerind strains being highly divergent.

Reed et al. (2004) explained this unexpected pattern by hypothesizing that the *Homo* lineage split around a million years ago into a Eurasian and an African species, and that this host speciation event caused codivergence of the lice (and the other parasite pairs mentioned above), with subsequent expansion of humans out of Africa resulting in direct physical contact with the Eurasian species that allowed the two divergent parasite lineages to come back into contact. The idea of separate Eurasian and African species of *Homo* around a million years ago is strongly falsified by the direct studies of human genes that indicate recurrent interchange of genes between African and Eurasian populations going back at least 1.46 MYA, with 95% confidence (Fig. 7). The pattern of divergent pairs of parasites and JCV does *not* require that the Eurasian populations of *Homo* were a separate species from African populations around a million years ago. Just as the curves shown in Figures 7 and 8 strongly falsify replacement, they do not preclude the existence of one or more populations of *Homo* in Eurasia that were indeed isolated, as pointed out previously in the discussion of Neanderthals. Indeed, the recent discovery of *Homo floresiensis* (Brown et al., 2004) shows that such isolated populations could

have well existed. An Asian isolate of humans, which may or may not have been a separate species, could explain these patterns of parasite evolution. As pointed out by Reed et al. (2004), it is necessary that the lineage of humans from which we are descended had to come ultimately into physical contact with this isolate, and perhaps reproductive contact as well, in order for the divergent parasites and JCV that coevolved with the isolate to transfer to the existing human lineage.

Pavesi (2004) suggested yet another possibility. His analysis had the most extensive geographical sampling, and with his enhanced geographical resolution he concluded that both lineages of JCV transferred to humans in Africa at two different times, and both subsequently expanded out of Africa in two different major migratory waves that brought the two parasitic lineages back into contact. This hypothesis is compatible with data on modern human specialist pathogens, such as HIV, that exist in more than one highly divergent clade because of multiple, independent host transfers to humans in Africa (Sharp et al., 2001). Hence, multiple independent host shifts to humans is a real epidemiological possibility. This hypothesis of independent host shifts followed by separate out-of-Africa expansions is also compatible with the reconstruction of human evolution shown in Figure 9 that indicates at least three major expansions of human populations out of Africa, with one being around 700,000 years ago. Given that the lineages of parasites and JCV must have diverged some time before their expansion out of Africa, the Acheulean expansion would explain the dates estimated by Hoberg et al. (2001) and Reed et al. (2004), without the need to invoke another human species or even population isolate.

Another important parasite of humans is *Plasmodium falciparum*, the malarial parasite. There has been speculation that this parasite could have moved out of Africa along with the most recent out-of-Africa expansion event of its human hosts, or alternatively it spread much more recently due to changes associated with the emergence of agriculture (Volkman et al., 2001; Joy et al., 2003). To address these issues, mtDNA sequence variation was recently surveyed in 100 worldwide isolates of the malarial parasite (Joy et al., 2003), and Templeton (2004a) performed a nested-clade phylogeographic analysis of these data. Isolation by distance dominates the most recent evolution of *P. faciparum* in this analysis, but there is also a significant range expansion, most likely out of Africa, that dates to 35,000 or 42,000 years ago, depending on the calibration date used. The same statistical framework discussed above regarding nested-clade analysis with multiple loci can be used to test the concordance of this out-of-Africa range expansion of malaria with the most recent out-of-Africa range expansion observed in humans (Fig. 7). The log-likelihood ratio test of homogeneity of the malarial out-of-Africa range expansion with the pooled human data for the most recent out-of-Africa range expansion yields a chi-square value of 1.97 (*P*-value of 0.16) for the 42,000-year date, and 2.54 (*P*-value of 0.11) for the 35,000-year date. Hence, the data of Joy et al. (2003) are compatible with the hypothesis that the malarial parasite spread out-of-Africa with their human hosts in the most recent out-of-Africa range expansion shown in Figure 9. However, the 95% confidence interval for the malarial expansion out-of-Africa is 113,000–5,900 years ago for the older calibration, and 93,700–4,900 years ago for the younger calibration. Thus, the malarial range expansion is also compatible with the hypothesis of a recent expansion due to the spread of agriculture. Although no definitive conclusions

were possible with this example, it does illustrate that nested-clade analysis of multiple loci provides an inference- and hypothesis-testing framework for both intraspecific evolution and for coevolution.

USING HAPLOTYPE TREES FOR INFERRING ADAPTIVE HUMAN EVOLUTION

Coalescence theory and phylogenetics provide a variety of tests to infer the presence and type of natural selection operating at the molecular level, particularly from haplotype trees of protein-coding genes (Bamshad and Wooding, 2003; Clark et al., 2003; Akey et al., 2004; Storz et al., 2004; Vallender and Lahn, 2004). Indeed, the great abundance of sequence data, not only in humans but in closely related species (outgroup data are frequently required for many of these tests of selection), have allowed massive screening throughout the human genome to identify those genes and DNA regions that were specifically subjected to positive, directional selection in the lineage leading to modern humans, and thus were involved in the adaptive transformation of the human species. This literature is discussed in more detail in Templeton (2005), so only a brief summary is presented here.

The protein-coding genes that were under directional selection only in the human lineage fall into just a few functional categories. The single most common class is genes involved in host-pathogen interactions, and the evidence for positive selection for many of these genes is confined to Eurasia. This implies that as the human lineage spread into Eurasia, humans encountered many new pathogens to which they had to adapt.

Another common class of genes showing directional selection in the human lineage is a set of genes coding for dietary enzymes. This indicates that humans greatly shifted their diet from that of chimpanzees, and this conclusion is compatible with studies of human intestinal parasites that indicate more similarity with baboons than with chimpanzees, as discussed above.

Many genes involved in sensory perception, central nervous system functioning, and brain anatomy also show positive selection in humans. This conclusion is compatible with the fossil record that reveals extensive changes in brain size and anatomy over the past two million years, and with the archaeological record that reveals extensive behavioral changes during this same time period.

CONCLUSIONS

Ever since the pioneering study of Cann et al. (1987), haplotype trees have played an important and central role in inferences about human evolution. Since 1987, our abilities have increased dramatically to estimate haplotype trees from a wide class of genes, to incorporate ancient DNA into the trees, and to analyze the trees for inferences on past population structure, demography, historical events, and patterns of natural selection. There is still much controversy about exactly what it is that these haplotype trees tell us about human evolution, but there is an emerging consensus on at least some points. First, most workers in this area now acknowledge that no one gene or DNA region can tell the whole story of human evolution. As a result, there is a growing trend to study multiple genes, or at least to integrate studies of one gene with studies of other genes. As shown in Figure 9, a multilocus approach yields a far richer picture of human evolution than is possible with any single gene.

A second point of growing consensus is the need for basing inferences from haplotype trees on rigorous statistical analyses. The early literature was dominated by drawing haplotype trees and then making visual inferences. As the literature has matured, there has been a growing realization that the data represented in haplotype trees must be analyzed statistically, and that hypotheses must be tested. One hypothesis about recent human evolution was the out-of-Africa replacement hypothesis, in which anatomically modern humans arose first in Africa, then expanded out-of-Africa as a new type (or even species) of humans, and drove the older "type" of humans found on the Eurasian continent to total genetic extinction. The early work on mtDNA haplotype trees was often presented as proof of this hypothesis, but there was no effort to test the replacement hypothesis vs. alternatives with the mtDNA (Templeton, 1994). With multilocus data sets, the hypothesis of total replacement can be tested, and it is strongly rejected ($P < 10^{-17}$). Thus, the hypothesis of total replacement can no longer be regarded as tenable.

A third factor to emerge from this growing literature is that genes and haplotype trees are not always the definitive cure-all for testing hypothesis about recent human evolution. The coalescent process itself has a high degree of stochasticity, and our ability to make inferences is limited by the number of informative mutations and their distribution in time and space, all factors over which we have no or limited control. The multilocus approach can ameliorate these deficiencies, but not completely eliminate them. This is illustrated by the current muddled state of inferences about past human inbreeding effective sizes and growth rates, or by the failure of the malaria mtDNA haplotype tree analysis to be able to discriminate between two hypotheses of range expansion that differ in their timing by an order of magnitude.

Genetic analyses of haplotype trees do not override fossil and archaeological data; rather, genetic data should be integrated with fossil and archaeological data. In this regard, the fossil and archaeological data overlay extremely well upon the haplotype tree-based inferences summarized in Figure 9. Three major expansions of human population out of Africa are inferred from the haplotype trees, and each is corroborated by fossil and archaeological data (Templeton, 2005). The out-of-Africa replacement hypothesis is strongly rejected by the haplotype tree data, and this is also supported by the fossil and current human data indicating that many of the anatomically modern traits that first appear in Africa are still polymorphic in living human populations, and that many other traits showed regional continuity throughout this time period. Thus, different modern traits show different evolutionary patterns, which is incompatible with one "type" of human replacing another "type," but which is compatible with different Mendelian traits spreading in their own unique fashion due to the interaction of selection, drift, and gene flow among *interbreeding* populations. Genetics, paleontology, and archaeology are synergistic disciplines for studying human evolution, and many exciting inferences about the evolutionary history of our species will undoubtedly emerge from this synergism.

ACKNOWLEDGMENTS

LITERATURE CITED

Adams AM, Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics 168: 1699–1712.

Aguirre E, Carbonell E. 2001. Early human expansions into Eurasia: the Atapuerca evidence. Quatern Int 75:11–18.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. Plos Biol 2:1591–1599.

Antón SC, Leonard WR, Robertson ML. 2002. An ecomorphological model of the initial hominid dispersal from Africa. J Hum Evol 43:773–785.

Asfaw B, Beyne Y, Suwa G, Walter RC, White TD, WoldeGabriel G, Yemane T. 1992. The earliest Acheulean from Konso-Gardula. Nature 360:732–735.

Ashford RW. 2000. Parasites as indicators of human biology and evolution. J Med Microbiol 49:771–772.

Avise JC. 2000. Phylogeography. Cambridge, MA: Harvard University Press.

Bacon AL, Dunlop MG, Farrington SM. 2001. Hypermutability at a poly(A/T) tract in the human germline. Nucleic Acids Res 29:4405–4413.

Bamshad MJ, Wooding SP. 2003. Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111.

Bar-Yosef O, Belfer-Cohen A. 2001. From Africa to Eurasia—early dispersals. Quatern Int 75:19–28.

Brown P, Sutikna T, Morwood MJ, Soejono RP, Jatmiko, Wayhu Saptomo E, Awe Due R. 2004. A new small-bodied hominin from the late Pleistocene of Flores, Indonesia. Nature 431:1055–1061.

Buzin CH, Feng JN, Yan J, Scaringe W, Liu Q, den Dunnen J, Mendell JR, Sommer SS. 2005. Mutation rates in the dystrophin gene: a hotspot of mutation at a CpG dinucleotide. Hum Mutat 25:177–188.

Caldararo N, Gabow S. 2000. Mitochondrial DNA analysis and the place of Neandertals in *Homo*. Ancient Biomol 3:135–158.

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

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, Bertorelle G. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. Proc Natl Acad Sci USA 100:6593–6597.

Castelloe J, Templeton AR. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol Phylogenet Evol 3:102–113.

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng XG, White TJ, Sninsky JJ, Adams MD, Cargill M. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960–1963.

Crandall KA. 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. Syst Biol 43:222–235.

Crandall KA, Templeton AR. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134:959–969.

Crandall KA, Templeton AR. 1999. Statistical approaches to detecting recombination. In: Crandall KA, editor. The evolution of HIV. Baltimore: Johns Hopkins University Press. p 153–176.

Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M. 2004. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet 36:700–706.

Crow JF, Kimura M. 1970. An introduction to population genetic theory. New York: Harpr & Row.

Curnoe D, Thorne A. 2003. Number of ancestral human species: a molecular perspective. Homo J Comp Hum Biol 53:201–224.

Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. Plos Biol 2:421.

Donnelly P, Tavare S. 1995. Coalescents and genealogical structure under neutrality. Annu Rev Genet 29:401–421.

Eller E. 2002. Population extinction and recolonization in human demographic history. Math Biosci 177:1–10.

Eller E, Hawks J, Relethford JH. 2004. Local extinction and recolonization, species effective population size, and modern human origins. Hum Biol 76:689–709.

Elson JL, Turnbull DM, Howell N. 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 74:229–238.

Eswaran V, Harpending H, Rogers AR. 2005. Genomics refutes an exclusively African origin of humans. J Hum Evol 49:1–18.

Ewens WJ. 1990. Population genetics theory—the past and the future. In: Lessard S, editor. Mathematical and statistical developments of evolutionary theory. Dordrecht: Kluwer Academic Publishers. p 177–227.

Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates, Inc.

Forde SE, Thompson JN, Bohannan BJM. 2004. Adaptation varies through space and time in a coevolving host-parasitoid interaction. Nature 431:841–844.

Fryxell KJ, Moon W-J. 2005. CpG mutation rates in the human genome are highly dependent on local GC content. Mol Biol Evol 22:650–658.

Garrigan D, Mobasher Z, Severson T, Wilder JA, Hammer MF. 2005. Evidence for archaic Asian ancestry on the human X chromosome. Mol Biol Evol 22:189–192.

Glenn DR, Brooks DR. 1986. Parasitological evidence pertaining to the phylogeny of the hominid parasites. Biol J Linn Soc 27:331–354.

Gutierrez G, Sanchez D, Marin A. 2002. A reanalysis of the ancient mitochondrial DNA sequences recovered from Neandertal bones. Mol Biol Evol 19:1359–1366.

Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P. 2001. Statistical evidence for miscoding lesions in ancient DNA templates. Mol Biol Evol 18:262–265.

Harpending H, Rogers A. 2000. Genetic perspectives on human origins and differentiation. Annu Rev Genom Hum Genet 1:361–73.

Hawks J, Hunley K, Lee S, Wolpoff M. 2000. Population bottlenecks and Pleistocene human evolution. Mol Biol Evol 17:2–22.

Hein J, Schierup MH, Wiuf C. 2005. Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford: Oxford University Press.

Hoberg EP, Alkire NL, Queiroz Ad, Jones A. 2001. Out of Africa: origins of the Taenia tapeworms in humans. Proc R Soc Lond [Biol] 268:781–787.

Hou YM, Potts R, Yuan BY, Guo ZT, Deino A, Wang W, Clark J, Xie GM, Huang WW. 2000. Mid-Pleistocene Acheulean-like stone technology of the Bose Basin, South China. Science 287:1622–1626.

Hudson RR. 1990. Gene genealogies and the coalescent process. Oxf Surv Evol Biol 7:1–44.

Hutchison DW, Templeton AR. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53:1898–1914.

Innan H, Zhang K, Marjoram P, Tavare S, Rosenberg NA. 2005. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. Genetics 169:1763–1777.

Joy DA, Feng XR, Mu JB, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su XZ. 2003. Early origin and recent expansion of Plasmodium falciparum. Science 300:318–321.

Kimura M. 1970. The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population. Genet Res 15:131–133.

Knight A. 2003. The phylogenetic relationship of Neandertal and modern human mitochondrial DNAs based on informative nucleotide sites. J Hum Evol 44:627–632.

Knowles LL, Maddison WP. 2002. Statistical phylogeography. Mol Ecol 11:2623–2635.

Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, Haesler Av, GrossSchmitz K, Possnert G, Paunovic M, Pääbo S. 2000. A view of Neandertal genetic diveresity. Nat Genet 26:144–146.

Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martinez-Maza C, Bastir M, Fortea J, Rasilla Mdl, Bertranpetit J, Rosas A. 2005. Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula. Mol Biol Evol 22:1077–1081.

Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. TIG 18:337–340.

Lordkipanidze D, Vekua A, Ferring R, Rightmire GP, Agusti J, Kiladze G, Mouskhelishvili A, Nioradze M, de Leon MSP, Tappen M, Zollikofer CPE. 2005. Anthropology: the earliest toothless hominin skull. Nature 434:717.

Marth G, Schuler G, Yeh R, Davenport R, Agarwala R, Church D, Wheelan S, Baker J, Ward M, Kholodov M, Phan L, Czabarka E, Murvai J, Cutler D, Wooding S, Rogers A, Chakravarti A, Harpending HC, Kwok P-Y, Sherry ST. 2003. Sequence variations in the public human genome data reflect a bottlenecked population history. Proc Natl Acad Sci USA 100:376–381.

Marth GT, Czabarka E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166:351–372.

McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584.

Nordborg M. 1998. On the probability of Neanderthal ancestry. Am J Hum Genet 63:1237–1240.

Palsboll PJ, Berube M, Aguilar A, Notarbartolo-Di-Sciara G, Nielsen R. 2004. Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (Balaenoptera physalus) populations. Evolution 58:670–675.

Pavesi A. 2004. Detecting traces of prehistoric human migrations by geographic synthetic maps of polyomavirus JC. J Mol Evol 58:304–313.

Pearson OM. 2004. Has the combination of genetic and fossil evidence solved the riddle of modern human origins? Evol Anthropol 13:145–159.

Pellmyr O. 2003. Yuccas, yucca moths, and coevolution: a review. Ann Mo Bot Gard 90:35–55.

Perlman SJ, Jaenike J. 2003. Infection success in novel hosts: an experimental and phylogenetic study of Drosophila-parasitic nematodes. Evolution 57:544–557.

Pluzhnikov A, Di Rienzo A, Hudson RR. 2002. Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics 161:1209–1218.

Posada D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol 19:708–717.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808.

Posada D, Crandall KA. 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA 98:13757–13762.

Posada D, Crandall KA. 2002. The effect of recombination on the accuracy of phylogeny estimation. J Mol Evol 54:396–402.

Posada D, Crandall KA, Holmes EC. 2002. Recombination in evolutionary genomics. Annu Rev Genet 36:75–97.

Pusch CM, Bachmann L. 2004. Spiking of contemporary human template DNA with ancient DNA extracts induces mutations under PCR and generates nonauthentic mitochondrial sequences. Mol Biol Evol 21:957–964.

Rannala B, Bertorelle G. 2001. Using linked markers to infer the age of a mutation. Hum Mutat 18:87–100.

Ray N, Currat M, Berthier P, Excoffier L. 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. Genome Res 15:1161–1167.

Reed DL, Smith VS, Hammond SL, Rogers AR, Clayton DH. 2004. Genetic analysis of lice supports direct contact between modern and archaic humans. Plos Biol 2:340.

Relethford JH. 2001a. Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. Am J Phys Anthropol 115:95–98.

Relethford JH. 2001b. Genetics and the search for modern human origins. New York: John Wiley & Sons.

Ricaut FX, Fedoseeva A, Keyser-Tracqui C, Crubezy E, Ludes B. 2005. Ancient DNA analysis of human Neolithic remains found in northeastern Siberia. Am J Phys Anthropol 126:458–462.

Richards M. 2003. The Neolithic invasion of Europe. Annu Rev Anthropol 32:135–162.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Rienzo AD, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H-J. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276.

Rightmire GP. 2004. Brain size and encephalization in early to mid-Pleistocene *Homo*. Am J Phys Anthropol 124:109–123.

Rogers AR, Praley AE, Bamshad MJ, Watkins WS, Jorde LB. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. Mol Biol Evol 13:895–902.

Ruff CB, Trinkaus E, Holliday TW. 1997. Body mass and encephalization in Pleistocene *Homo*. Nat 387:173–176.

Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. Genetics 152:1079–1089.

Segraves KA, Pellmyr O. 2004. Testing the out-of-Florida hypothesis on the origin of cheating in the yucca-yucca moth mutualism. Evolution 58:2266–2279.

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G, Pääbo S. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. Plos Biol 2:313–317.

Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. 2001. The origins of acquired immune deficiency syndrome viruses: where and when? Philos Trans R Soc Lond [Biol] 356.

Shingleton AW, Stern DL. 2003. Molecular phylogenetic evidence for multiple gains or losses of ant mutualism within the aphid genus *Chaitophorus*. Mol Phylogenet Evol 26:26–35.

Soodyall H. 1993.Mitochondrial DNA polymorphisms in southern African populations. Ph.D. thesis, University of the Witwatersrand, Johannesburg.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989.

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

Storz JF, Payseur BA, Nachman MW. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol Biol Evol 21:1800–1811.

Stringer C. 2002. Modern human origins: progress and prospects. Philos Trans R Soc Lond [Biol] 357:563–579.

Stringer CB, Andrews P. 1988. The origin of modern humans. Science 239:1263–1268.

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M. 1995. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57:1463–1475.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Tajima F. 1989a. The effect of change in population size on DNA polymorphism. Genetics 123:597–601.

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

Templeton AR. 1993. The "Eve" hypothesis: a genetic critique and reanalysis. Am Anthropol 95:51–72.

Templeton AR. 1994. "Eve:" hypothesis compatibility versus hypothesis testing. Am Anthropol 96:141–147.

Templeton AR. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics 144:1263–1270.

Templeton AR. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol Ecol 7:381–397.

Templeton AR. 2002. Out of Africa again and again. Nature 416:45–51.

Templeton AR. 2004a. A maximum likelihood framework for cross validation of phylogeographic hypotheses. In: Wasser SP, editor. Evolutionary theory and processes: modern horizons. Dordrecht: Kluwer Academic Publishers. p 209–230.

Templeton AR. 2004b. Statistical phylogeography: methods of evaluating and minimizing inference errors. Mol Ecol 13:789–809.

Templeton AR. 2005. Population biology and population genetics of Pleistocene hominins. In: Henke W, Rothe H, Tattersall I, editors. Handbook of palaeoanthropology. Berlin: Springer-Verlag. In press.

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

Templeton AR, Boerwinkle E, Sing CF. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. Genetics 117:343–351.

Templeton AR, Sing CF, Kessling A, Humphries S. 1988. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics 120:1145–1154.

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633.

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

Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF. 2000a. Recombinational and mutational hotspots within the human *Lipoprotein Lipase* gene. Am J Hum Genet 66:69–83.

Templeton AR, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF. 2000b. Cladistic structure within the human *Lipoprotein Lipase* gene and its implications for phenotypic association studies. Genetics 156:1259–1275.

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, Roux ML, Bradman N, Goldstein DB. 2000. Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the "Black Jews of Southern Africa." Am J Hum Genet 66:674–686.

Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genom Hum Genet 4:293–340.

Tomatsu S, Orii KO, Bi Y, Gutierrez MA, Nishioka T, Yamaguchi S, Kondo N, Orii T, Noguchi A, Sly WS. 2004. General implications for CpG hot spot mutations: methylation patterns of the human iduronate-2-sulfatase gene locus. Hum Mutat 23:590–598.

Tomso DJ, Bell DA. 2003. Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. J Mol Biol 327:303–308.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993a. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC. 1993b. mtDNA

variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591–608.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. 2005. Widespread recombination in published animal mtDNA sequences. Mol Biol Evol 22:925–933.

Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. Hum Mol Genet 13:245–R254.

Vekua A, Lordkipanidze D, Rightmire GP, Agusti J, Ferring R, Maisuradze G, Mouskhelishvili A, Nioradze M, de Leon MP, Tappen M, Tvalchrelidze M, Zollikofer C. 2002. A new skull of early *Homo* from Dmanisi, Georgia. Science 297:85–89.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507.

Volkman SK, Barry AE, Lyons EJ, Nielsen KM, Thomas SM, Choi M, Thakore SS, Day KP, Wirth DF, Hartl DL. 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. Science 293:482–484.

Wall JD. 2000. Detecting ancient admixture in humans using sequence polymorphism data. Genetics 154:1271–1279.

Weiss KM. 1984. On the numbers of members of the genus *Homo* who have ever lived, and some evolutionary implications. Hum Biol 56:637–649.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature 423:742–747.

Willerslev E, Cooper A. 2005. Ancient DNA. Proc R Soc Lond [Biol] 272:3–16.

Wilson JF, Goldstein DB. 2000. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. Am J Hum Genet 67:926–935.

Wooding S, Rogers A. 2002. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics 161:1641–1650.

Yang Z. 1997. On the estimation of ancestral population sizes of modern humans. Genet Res 69:111–116.

Zangerl AR, Berenbaum MR. 2003. Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. Evolution 57:806–815.

Zhao Z, Boerwinkle E. 2002. Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. Genome Res 12:1679–1686.