The Myth of Eve: Molecular Biology and Human Origins

Francisco J. Ayala

It has been proposed that modern humans descended from a single woman, the "mitochondrial Eve" who lived in Africa 100,000 to 200,000 years ago. The human immune system *DRB1* genes are extremely polymorphic, with gene lineages that coalesce into an ancestor who lived around 60 million years ago, a time before the divergence of the apes from the Old World monkeys. The theory of gene coalescence suggests that, throughout the last 60 million years, human ancestral populations had an effective size of 100,000 individuals or greater. Molecular evolution data favor the African origin of modern humans, but the weight of the evidence is against a population bottleneck before their emergence. The mitochondrial Eve hypothesis emanates from a confusion between gene genealogies and individual genealogies.

The DNA of organisms embodies virtually unlimited information about their evolutionary history. The information is encoded in the linear sequence of the four nucleotide components of DNA (adenine, cytosine, guanine, and thymine) in the same fashion as semantic information is encoded in the sequence of letters of an English text. Molecular evolution investigations of DNA sequences have been largely directed toward reconstructing the phylogeny (evolutionary history) of species relations. But many other issues can be explored. I have used genes of a particular kind, those involved in the immune response, to investigate the size of human populations from the time of the origin of our species and earlier in order to ascertain whether the existence of the mythical Eve has been confirmed by science, as some reports claim (*1*).

Evolution is a time-dependent gradual process. At the genetic level, evolution occurs by accumulated substitutions of one nucleotide by another in the DNA of the organism. Nucleotide mutations arise with constant probabilities, but most are lost by chance shortly after their origin. The fate of the rest depends on their effects on the organism. Many mutations are injurious and are readily eliminated by natural selection. Other mutations are adaptively neutral or nearly so; that is, the replacement of one nucleotide by another is of little or no consequence to the organism's welfare. These mutations occasionally spread and become fixed in the species at rates that are constant, so that the number of differences between two species is roughly proportional to the time since their divergence from a common ancestor. Other mutations are favored by natural selection because they benefit the organism. Some of these mutations spread among the individuals and accumulate in the specie, and this process is also time-dependent. The regularity of the process by which nucleotide substitutions occur makes it possible to reconstruct the historical relations among species and also to assign dates to relevant events. That is, there is a molecular clock of evolution, which is not a metronomic clock like ordinary timepieces that measure time intervals precisely, but rather a stochastic clock that is dependent, like radioactive decay, on events that occur with constant probabilities.

The evolutionary information encoded in the DNA has the notorious attribute of being effectively inexhaustible. The nucleotide sequence of a single gene or small DNA fragment is often sufficient to solve a particular evolutionary question, but organisms have many genes (~100,000 in primates and other mammals). The evolutionary information obtained by investigating one gene can be supplemented with the investigation of additional genes until the information becomes sufficient to settle a particular issue. The practical limits are resources, financial costs, and time.

The Reconstruction of Genealogy

The methodology used to reconstruct evolutionary genealogies can be illustrated with a simple example (*2*). Figure 1 shows a schematic comparison among four *DRB1* gene sequences, two from humans and two from chimpanzees, each of 270 nucleotides; the data for the six pairwise comparisons are given in Table 1. The most similar pair are the human gene *Hs*1103* and the chimpanzee gene *Pt*0309*, which have been arranged in Fig. 2 to indicate that they share a recent ancestor with each other. The other two genes, *Hs*0302* and *Pt*0302*, are also more similar to one another than to *Hs*1103* or *Pt*0309*, and they are also paired in Fig. 2. The two doublets are then linked to indicate that all four genes derive from a common ancestral gene. Branch lengths can be made proportional to the number of nucleotide changes that occurred in each branch, as inferred from Table 1.

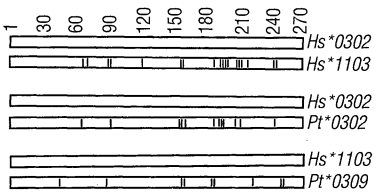

Fig. 1. Comparison of two human (*Hs*) and two chimpanzee (*Pt*) DNA sequences. The segments compared are 270 nucleotides that make up exon 2 of the *DRB1* gene. The differences between sequences are represented by vertical bars.

One noteworthy property of these four genes is that the two human sequences are more different from one another than each is from one chimpanzee gene. Hence, the lineages of the two human genes diverged from each other more than 6 million years ago (Ma), the approximate time when the lineages of humans and chimpanzees diverged. The ancient origin of these and other *DRB1* gene lineages is a property that makes them particularly suitable for ascertaining the history of ancient human populations.

Genealogy of *DRB1* Genes

DRB1 is one of ~100 genes that make up the human leukocyte antigen (HLA) complex, which extends over a DNA segment 4 million nucleotides long located on chromosome 6. The HLA genes specify molecules with a critical role in tissue compatibility and in defense against pathogens and parasites. These genes are arranged in two distinct groups, class I and class II, separated by several dozen genes that have functions mostly unrelated to the immune response (Fig. 3).

The HLA complex is homologous to the major histocompatibility gene complex (MHC) of mammals and other vertebrates (*3*). MHC molecules on the surfaces of certain cells bind protein fragments (antigens) and present them to lymphocytes called T

Table 1. Numbers of nucleotide differences among four *DRB1* genes, two from humans (*Hs*) and two from chimpanzees (*Pt*).

Gene	*Hs*1103*	*Pt*0309*	*Pt*0302*
*Hs*0302*	18	18	12
*Hs*1103*		9	20
*Pt*0309*			20

The author is the Donald Bren Professor of Biological Sciences at the University of California, Irvine, and chairman of the AAAS Board of Directors. This article is adapted from a lecture delivered at the 161st annual meeting of the American Association for the Advancement of Science in Atlanta on 18 February 1995.

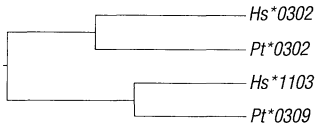

Fig. 2. Genealogical tree of the four *DRB1* genes in Fig. 1, representing their patterns of evolutionary divergence. Branch lengths are proportional to the nucleotide differences between the genes (Table 1).

cells. When antigen-presenting cells contact T cells that bear receptors matching a particular combination of protein fragment and MHC molecule, the T cells are stimulated to proliferate and to initiate the specific arm of the immune response, including the secretion of specific antibodies.

The recognition of protein fragments is mediated by a specialized groove on the surface of the MHC molecule, called the peptide-binding region, that consists of some 50 amino acids (*4*). The composition of these amino acids varies from one MHC molecule to another, and this variation is responsible for the tremendous polymorphism characteristic of the MHC molecules and their encoding genes. In people, as well as in other mammals, scores of gene variants (alleles) may exist at any one of several MHC loci, and some of the allelic pairs may differ at >100 nucleotide sites (*5*). The MHC polymorphisms are ancient, with gene lineages that can be traced back for millions of years, in primates (*6–8*) as well as in rodents (*9*).

Figure 4 represents the genealogy of 119 *DRB1* genes, of which 59 are from humans, 40 are from apes, and 20 are from Old World monkeys (Table 2). In this genealogy, the length of each branch is proportional to the number of nucleotide substitutions that have occurred in that lineage. Thus, we see at the bottom of the figure that of the two macaque genes, *Mm*0301* and *Mm*0302*, the first one has changed more than the second.

Age of the Human *DRB1* Lineages

The *DRB1* human gene lineages are very ancient, as is apparent from the distant associations between some human genes relative to those from other species. For example, the nine human genes at the top of Fig. 4 are closely related to one another, but they are more distantly related to other human genes in the genealogy than they are to the six genes immediately below them, including one drill gene and four macaque genes. The evolutionary divergence of the hominoids (humans and apes) from the cercopithecoids (Old World monkeys) occurred around 35 Ma, at the boundary between the Eocene and Oligocene epochs. The relations shown in Fig. 4 demonstrate that several human gene lineages already existed at that time.

The age of the human *DRB1* lineages can be appraised by calibrating the *DRB1* molecular clock, which yields a rate of 1.06×10^{-9} nucleotide substitutions per site per year (*10*). Figure 5 displays a genealogy of the 59 human *DRB1* genes, constructed by a method that uses the average rate of evolution to determine the length of branches. The time scale at the bottom highlights three points that approximately correspond to important events in human evolution: the divergence of the orangutan lineage at 15 Ma, the divergence of humans and African apes (chimpanzees and gorillas) at 6 Ma, and the emergence of *Homo erectus* at 1.7 Ma.

These points of reference are useful for determining the number of gene lineages in existence at a given time, which is done by counting the number of lineages intersected by a vertical line drawn from the time point. For example, 32 human lineages were already in existence around 6 Ma, which implies, correspondingly, that the other 27 lineages arose after the hominid lineage diverged from the African apes. The genealogy of all human genes coalesces at ~60 Ma, which is to say that the human *DRB1* genes started diverging that long ago.

Human Ancestral Populations

If 32 *DRB1* gene lineages have persisted since 6 Ma, it follows that no fewer than 16 individuals could have lived at any given time over that long span. The minimum number of individuals must have been much larger, because the probability is effectively zero that all 16 individuals in a population would be heterozygotes (that is, carrying two different genes), each for two genes different from all others. The theory of gene coalescence and other mathematical theories formulated in the last decade provide models and analytical tools for extracting information about human populations from the *DRB1* genealogy.

The coalescence theory examines the genealogical relations between genes (*11*). According to this theory, all genes (alleles) present in an extant population must have descended from a single gene, to which they coalesce. The theory was first formulated for neutral or nearly neutral genes that do not modify the welfare of the organism. In a random mating population at equilibrium, the mean coalescence time of neutral genes is given by $T = 4N[1 - (1/i)]$ generations, where T is the number of generations to coalescence, N is the effective size of the population, i is the number of sampled genes, and the variance is large (*12*). For any two genes ($i = 2$), the mean coalescence time reduces to $T = 2N$ generations; for a large number of genes, the mean coalescence time is $T \approx 4N$. Thus, in a population with $N = 1$ million individuals, genes are expected to converge to their one ancestor 4 million generations earlier.

The coalescence equation can be used in the opposite direction, so that we can estimate population size if the coalescence time is known. The human *DRB1* genes coalesce to their last common ancestor at ~60 Ma (Fig. 5), in the middle Paleocene, 10 to 20 million years (My) before the divergence of the New World and Old World monkeys. If we assume that the average generation time over this long span is 15 years (which is surely too high), the coalescence occurs 4 million generations before the present. If

Table 2. Species and number of *DRB1* genes.

Symbol	Scientific name (common name)	Genes
	Apes	
Hs	*Homo sapiens* (human)	59
Pp	*Pan paniscus* (bonobo or pygmy chimpanzee)	4
Pt	*Pan troglodytes* (chimpanzee)	24
Gg	*Gorilla gorilla* (gorilla)	10
Or	*Pongo pygmaeus* (orangutan)	2
	Old World monkeys	
Mm	*Macaca mulatta* (rhesus macaque)	16
Mn	Macaca nemestrina (pigtail macaque)	1
Ml	*Mandrillus leucophaeus* (drill)	2
Ph	*Papio hamadryas* (hamadryas baboon)	1

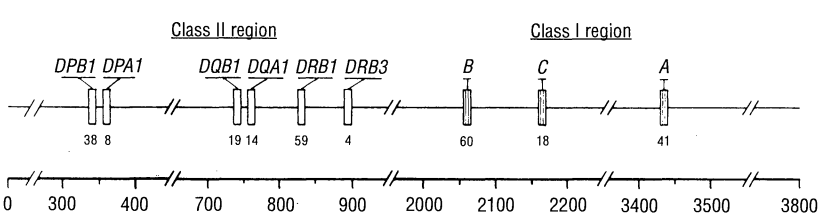

Fig. 3. Location of some polymorphic genes within the HLA complex in human chromosome 6. There are two sets of genes, class I and class II, separated by a region with unrelated genes. The number of variant genes (alleles) known at a locus appears under the box that indicates the location of the gene. The scale below gives the number of nucleotides (×1000) along the segment.

the *DRB1* polymorphism were neutral, the coalescence date would require a long-term population of 1 million individuals.

We need to qualify this conclusion with two observations. First, although N is an estimate of the mean of the population over time, the estimate refers to the harmonic mean, which has the important property of being affected disproportionately more by smaller than by larger numbers. The estimate of N is therefore consistent with population sizes much greater than 1 million for many generations, but it is not compatible with much smaller population sizes for very many generations. Second, because of the large variance of the coalescence equation, the estimate of N is readily compatible with values between, say, 500,000 and 2 million individuals (*13*).

The previous calculations are for neutral genes, but the *DRB1* genes are subject to overdominant natural selection, that is, heterozygotes for HLA genes are better off than are homozygotes (*3*, *14*); for instance, heterozygotes display enhanced resistance to *Plasmodium falciparum*, the parasite that causes malignant malaria (*15*). Overdominant selection, like other forms of balancing selection, increases the probability that gene polymorphisms will persist over time, and thus it reduces the number of individuals required for their persistence.

In the case of balancing selection, the coalescence process has the same structure as in neutral gene genealogy, except for a scaling factor, f_s, so that the time to coalescence in generations is estimated by $T = Nf_s$ (*16–18*). The scaling factor depends on the selected mutation rate as well as on the selection coefficient, s, which measures the advantage accruing to the heterozygotes. Estimates of s range from 0.0007 to 0.019 (*7*, *14*, *15*). If we assume values of $s = 0.01$ to 0.03, the population size required for maintaining the *DRB1* polymorphism is ~125,000 to ~300,000 individuals (Table 3) (*19*, *20*).

The time to coalescence has been estimated at 60 My on the basis of the genealogy reconstructed in Fig. 5. We can reach the same estimate by a somewhat different route. According to the theory of coalescence, if we sample any two genes at random, the expected value for their coalescence is $2N$. Figure 6 displays the distribution of genetic distance between pairs of the 59 human *DRB1* alleles. The mean distance value is 0.067, which, on the basis of the rate of the *DRB1* evolutionary clock, yields an average coalescence time of $2N = 31.6$ My. This value is in reasonable agreement with the value of $4N = 60$ My obtained by reconstructing the genealogy.

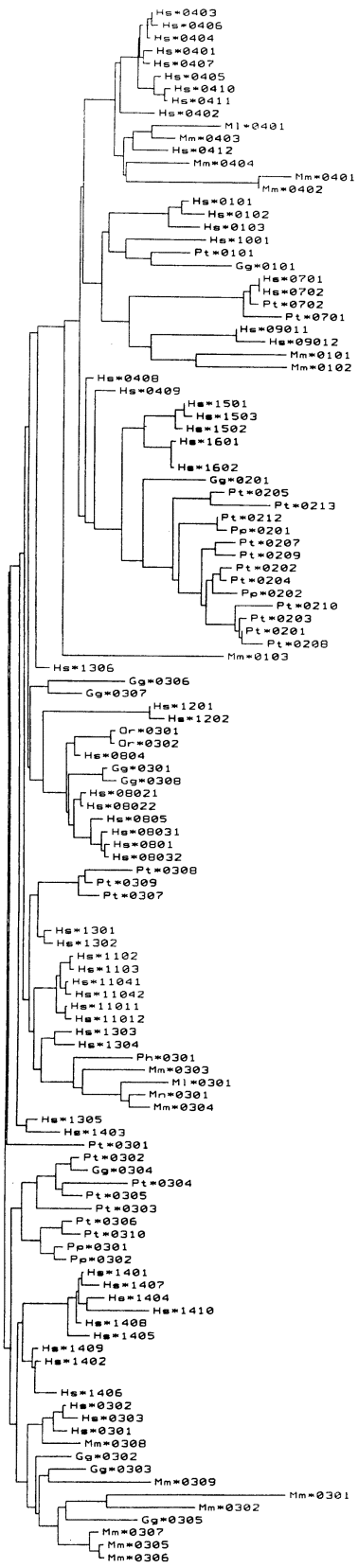

Fig. 4. Genealogical tree of 119 *DRB1* genes (exon 2) from humans (*Hs*) and other primates (Table 2). The tree was made with the use of the neighbor-joining algorithm (*41*) on the basis of genetic distances (*D*) estimated by Kimura's two-parameter method (*42*). The matrix of genetic distances has 7021 entries that represent all pairwise comparisons between the 119 genes. Branch lengths are proportional to the estimated nucleotide substitutions along the branch; the scale bar represents 0.01 *D*. Sequence sources are listed in (*5*).

Experimental Simulated Populations

The coalescence theory draws inferences about past events on the basis of observations about current polymorphisms. These

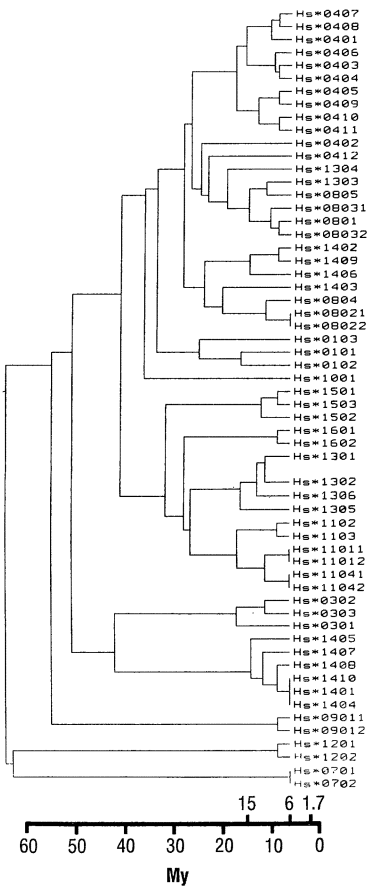

Fig. 5. Genealogical tree of 59 human *DRB1* genes (exon 2), constructed according to the "average linkage" method [(*39*), p. 31; also known as the unweighted pair-group method using arithmetic averages (UPGMA) (*43*)], on the basis of genetic distances (*D*) estimated by Kimura's two-parameter method (*42*) for the 1711 pairwise comparisons between the genes. The average linkage method assumes equal rates of evolution along the branches. The time scale is based on a rate of evolution of 1.06×10^{-9} substitutions per nucleotide site per year, estimated by the minimum method (*8*) on the basis of genetic distances between species that diverged at different times. The species compared for estimating the rate of evolution are *Homo-Pan*, *Homo-Gorilla*, and *Pan-Gorilla* at 6 My; *Homo-Pongo*, *Pan-Pongo*, and *Gorilla-Pongo* at 15 My; *Pan-Macaca* and *Gorilla-Macaca* at 35 My; and *Homo*, *Pan*, and *Gorilla*, each compared with species of New World monkeys at 40 My. The regression equation is $D = 0.0099 + 0.00106$ My; if the minimum-minimum method (*8*) is used instead, the regression equation is $D = 0.0027 + 0.00111$ My. The fit in both cases is $r > 0.85$.

inferences can be tested by computer experiments, in which the time direction is reversed and the process is examined as it proceeds from past to present conditions. The experiments simulate populations consisting of a certain number of individuals that reproduce each generation by random reassortment of genes. Figure 7 presents the results of three sets of experiments. The fate of 60 different genes, initially in identical frequencies, is examined over 100,000 generations, a time approximately equivalent to the 1.7 My from the emergence of *Homo erectus* to the present. The probability of reproduction is $s = 0$ to 0.2 higher for heterozygotes than for homozygotes.

For populations of $N = 1000$ individuals (Fig. 7, top panel), without selection, all 60 genes but one are rapidly eliminated, consistent with the theoretical expectation that fixation for one gene will occur in $4N = 4000$ generations. If the heterozygotes have $s = 0.01$ to 0.02 advantage, which is the magnitude operating in real populations, all genes but two or three are eliminated. Even with unrealistically large s advantages of 0.1 to 0.2, only 8 to 11 genes persist after 100,000 generations. Populations of 10,000 individuals (Fig. 7, middle panel) are likewise incapable of maintaining 60 genes over many generations. Without natural selection, all genes but one are soon eliminated. With $s = 0.01$ to 0.02, only 9 to 11 genes persist at the end of the experiment.

Mutation is ignored in the simulated populations. If the mutation rate to alleles favored by natural selection is 5×10^{-7} mutations per gene per generation (*8*), a new mutation appears every 100 generations in a population of $N = 10{,}000$ individuals. The mutation process increases the number of genes in the populations, but it would not be nearly sufficient to maintain 60 genes if $s = 0.01$ to 0.02. With $N = 10{,}000$ and $s = 0.01$ or 0.02, the process of elimination of the first 40 genes is very fast.

For populations of $N = 100{,}000$ individuals (Fig. 7, bottom panel), with $s = 0.01$ to 0.03, 30 to 50 genes remain at the end of the experiment. If we assume, as above, a selected mutation rate of 5×10^{-7} mutations per gene per generation, a new overdominant mutation will arise every 10 generations. Most newly arisen mutations have a high probability of being quickly lost by chance, even if they are favorably selected, but mutation would increase the number of persisting genes. Depending on the mutation rate, the number of genes might approach or reach 60, the number of *DRB1* genes. In conclusion, these computer-simulated populations yield results consistent with the theoretical calculation that to maintain ~60 *DRB1* genes, human ancestral populations must have consisted of 100,000 or more individuals over their long history.

Table 3. Effective population size, N, as a function of various parameter values. T, time in generations; s, selection advantage of the heterozygotes; u, mutation rate per gene per generation for selected alleles. If the long-term generation time is 15 years, then $T = 4.0 \times 10^6$ corresponds to 60 My. If the mutation rate per nucleotide site per year is $m = 10^{-9}$, half the sites yield selected alleles ($n = 135$), and the generation time is $g = 15$ years, then $u = 2.0 \times 10^{-6}$; the same value of u results if $m = 5 \times 10^{-9}$, $n = 40$, and $g = 10$.

T	u	s	N
4.0×10^6		0	1,000,000
	4.5×10^{-6}	0.01	297,789
		0.02	257,430
		0.03	230,213
	2.0×10^{-6}	0.01	194,240
		0.02	149,072
		0.03	125,343
3.0×10^6		0	750,000
	4.5×10^{-6}	0.01	211,793
		0.02	178,691
		0.03	157,795
	2.0×10^{-6}	0.01	131,196
		0.02	98,992
		0.03	82,505
2.5×10^6		0	625,000
	4.5×10^{-6}	0.01	169,720
		0.02	141,111
		0.03	123,668
	2.0×10^{-6}	0.01	101,920
		0.02	76,118
		0.03	63,105

The Mitochondrial Eve

Most of the genetic information is stored in the chromosomes inside the cell nucleus. The total DNA in a human cell nucleus amounts to 6×10^9 nucleotides or base pairs (bp), half in each set of 23 chromosomes that are inherited from each parent. A relatively small amount of DNA (~16,569 bp) exists in the mitochondria, cell organelles outside the nucleus. The inheritance of the mitochondrial DNA (mtDNA) is distinctive and follows a pattern of maternal inheritance. Sons and daughters inherit mtDNA from their mothers, but only the daughters transmit it to their progeny.

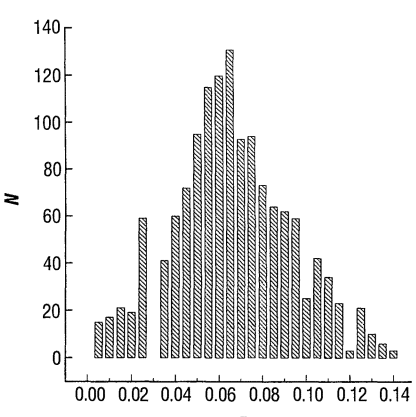

Fig. 6. Distribution of 1711 genetic distances between pairs of 59 human *DRB1* genes (exon 2); the mean is $\bar{D} = 0.0675 \pm 0.0276$.

Analyses of the mtDNA from >100 ethnically diverse individuals have shown that the mtDNA sequences of modern humans coalesce to one ancestral sequence, the "mitochondrial Eve" that existed in Africa ~200,000 years ago (*21*). This Eve, however, is not the one mother from whom all humans descend, but rather a mtDNA molecule (or the woman carrier of that molecule) from which all modern mtDNA molecules descend. The inference that all humans descend from only one or very few women (*1*) is based on a confusion between gene genealogies and individual genealogies. Gene genealogies gradually coalesce toward a single DNA ancestral sequence (such as in Figs. 4 and 5), whereas individual genealogies increase by a factor of 2 each generation: An individual has two parents, four grandparents, and so on. (The

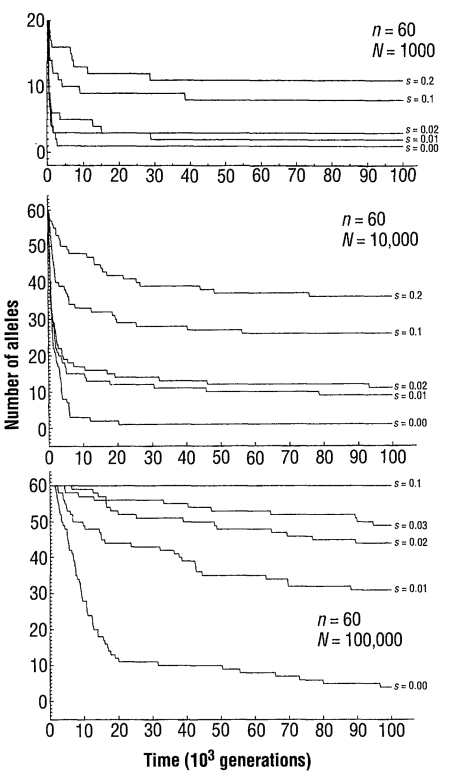

Fig. 7. Loss of genetic polymorphism in computer-simulated populations initiated with 60 genes (alleles) at identical frequencies. The 100,000 generations correspond to ~1.7 My, the time elapsed since the emergence of *Homo erectus*. The effective size of populations is N individuals; the selection advantage of the heterozygotes is s. The persisting genes are censused at 100-generation intervals. The populations "reproduce" by random sampling, at each generation, of N pairs of genes from an infinite pool that has the gene frequencies observed in the previous generation.

theoretical number of ancestors for any one individual becomes enormous after some tens of generations, but "inbreeding" occurs; that is, after some generations, ancestors appear more than once in the genealogy.) Coalescence to one ancestral gene originally present in one individual does not disallow the contemporary existence of many other ancestors from whom we have inherited the other genes.

The conclusion warranted by the mtDNA analysis is that the mitochondrial Eve is the ancestor of modern humans in the maternal line. Any person has a single ancestor in the maternal line in any given generation. Thus, a person inherits the mtDNA from the great-grandmother in the maternal line, but also inherits other genes from the three other great-grandmothers and the four great-grandfathers (about one-eighth of the total DNA from each great-grandparent). The mtDNA that we have inherited from the mitochondrial Eve represents a four-hundred-thousandth part of the DNA present in any modern human. The rest of the DNA, 400,000 times the amount of mtDNA, was inherited from other contemporaries of the mitochondrial Eve.

The theory of gene coalescence makes it possible to estimate the number of ancestors who were contemporaries of the mitochondrial Eve. The mtDNA is inherited as a single copy, from only one parent, and the mtDNA polymorphism is largely neutral. According to the theory, the coalescence into a single ancestral molecule is expected to be $T = 2N_f$ generations ago, where N_f is the number of mothers per generation. This inference assumes constant population size and other conditions that are unlikely in reality, so that the conclusions reached are only rough approximations. If we assume 20 years per generation, the 200,000 years of the mtDNA coalescence yield $T = 10{,}000$ generations; therefore, $N_f = 5000$ mothers and $N = 10{,}000$ individuals, which is almost certainly an underestimate (*22*).

Other estimates place the coalescence of the mtDNA at 143,000 and 298,000 years ago (*23*). One recent analysis, which takes into account variable substitution rates for different sites of the mtDNA genome and is supported by computer simulations, yields an estimate of 622,000 to 889,000 years ago (*22*) and corresponds to $N = 31{,}100$ to 44,450 individuals. As noted earlier, estimates of mean coalescence time as a function of N, and vice versa, have large variances. When the sample of genes is large, the standard deviation of the mean for mitochondrial DNA is $>N/2$ (*13*). The 95% confidence interval coalescence correspondingly extends at the upper end to >88,900 generations and an equal number of individuals. Thus, despite considerable uncertainty, the mtDNA results yield a mean population size that ranges between 10,000 and >50,000 individuals throughout the Pleistocene. This finding is consistent with the estimate based on the *DRB1* polymorphism of a population size on the order of 100,000 individuals over the last 60 My.

The *ZFY* Adam

A genetic gender counterpart of the mtDNA is the Y chromosome, which is transmitted from fathers to sons. Regions of the Y chromosome do not have homologs in the X chromosome and thus are transmitted through the paternal line, just as the mtDNA is inherited through the maternal line. A 729-bp fragment of the Y-chromosome *ZFY* gene (thought to be involved in the maturation of testes or sperm) has been sequenced in 38 men, of diverse geographic origins and representative of major ethnic groups, and no variation was observed (*24*). Comparison with the homologous region of the great apes yields a rate of substitution of 1.35×10^{-9} nucleotides per site per year. Assuming that substitutions in this DNA segment are neutral, coalescence theory yields 270,000 years as the expected date for the last common ancestor of the *ZFY* gene of modern humans, with 95% confidence limits from 0 to 800,000 years (*24*). The carrier of the ancestral *ZFY* gene predicted by these calculations is the ancestor of all modern humans in the paternal line. As was the case for the mitochondrial Eve, this "*ZFY* Adam" is the individual from which all humans have inherited the *ZFY* gene, but he is not our only ancestor in his generation. We have inherited the other thousands of genes from many other contemporaries of this Adam.

The expected coalescence of a DNA polymorphism that is transmitted in a single copy and paternally inherited is $T = 2N_m$, where N_m is the number of males. If we assume 20 years per generation, the coalescence to the ancestral *ZFY* gene yields an effective population size of 6750 fathers, or 13,500 individuals, with a 95% confidence upper limit of $N = 40{,}000$ individuals. If we account for the standard deviation of the mean coalescence, the 95% upper limit for N would increase to 80,000 individuals. The *ZFY* results are thus consistent with the mtDNA-derived and *DRB1*-derived estimates that human ancestral populations have been ~100,000 individuals for millions of years.

Population Bottlenecks

Neither the mtDNA results nor the *ZFY* results lead to the conclusion that narrow population bottlenecks consisting of one or very few couples have occurred in human ancestral history. We want, nevertheless, to explore the likelihood that a bottleneck did occur. When the number of individuals in a population oscillates from generation to generation, small numbers have a disproportionately large effect on the value of N, because N is the harmonic mean of the population size over time. Therefore, very narrow or long-lasting bottlenecks are inconsistent with the *DRB1* results. However, the question remains whether an occasional population bottleneck may have occurred, and if so, how small it could be. It has been suggested that a population bottleneck occurred at the transition from archaic to modern *H. sapiens*, some 100,000 to 200,000 years ago (*1, 21, 24*).

The consequences of a population bottleneck depend not only on the size of the bottleneck, N_b, but also on the number of bottleneck generations, t_b. A useful measure for evaluating the effects of a bottleneck is the ratio N_b/t_b, which if <10 leads to a drastic reduction in genetic polymorphism (*18*). Thus, a bottleneck of 100 individuals would substantially reduce genetic variation if it lasted 10 or more generations, but balancing selection facilitates the persistence of polymorphisms through a bottleneck (*25*). The persistence of HLA polymorphisms over millions of years requires that the size of human ancestral populations be at least $Ns = 10$ at all times (*7, 17*). If $s = 0.01$, the minimum population size possible at any time would therefore be $N_b = 1000$. The minimum number must have been in fact much larger, because human population bottlenecks cannot last just a few generations. Many generations are required for a human population to grow from 1000 to its long-term mean, which we have estimated to be ~100,000 individuals. The rate of growth of human populations throughout the Pleistocene has been estimated to be ~0.02% per generation (*26*).

Computer-simulated populations are useful for exploring the minimum bottleneck size that would allow the persistence of the *DRB1* polymorphism. The results in Fig. 8 are based on 200 separate computer runs. If there is no selection ($s = 0$) and we ignore the time required for a population to grow back to its long-term population size, and if we assume that 70 gene lineages were present before the bottleneck, the smallest bottleneck that would allow the persistence of 60 of these 70 gene lineages is 510 to 550 individuals (Fig. 8, top panel). When we account for the time required for the population to recover to its average size, the minimum population size at the bottleneck becomes substantially larger. Assuming a rate of population increase of $R = 1\%$ per generation (which is 50 times the average growth rate of human populations throughout the Pleistocene) (*26*) and population

growth to 100,000 individuals, 4490 to 4590 individuals are the minimum bottleneck for passing 60 alleles (Fig. 8, top panel).

Overdominant selection reduces only slightly the minimum number of individuals required at the bottleneck. When $s = 0.01$, the minimum bottleneck size for passing 60 of the 70 alleles is 500 to 540 individuals if we ignore the required growth back to long-term numbers, but 4310 to 4380 when we take population growth to 100,000 individuals into account (Fig. 8, middle panel). Even with the unrealistically high $s = 0.1$, the corresponding values become 490 to 530 and 3410 to 3510 (Fig. 8, bottom panel). It may be concluded that, to account for the *DRB1* polymorphism, the minimum possible number of individuals at a bottleneck is at least 4000; this number is consistent with the lower estimates derived from the mtDNA and the ZFY gene.

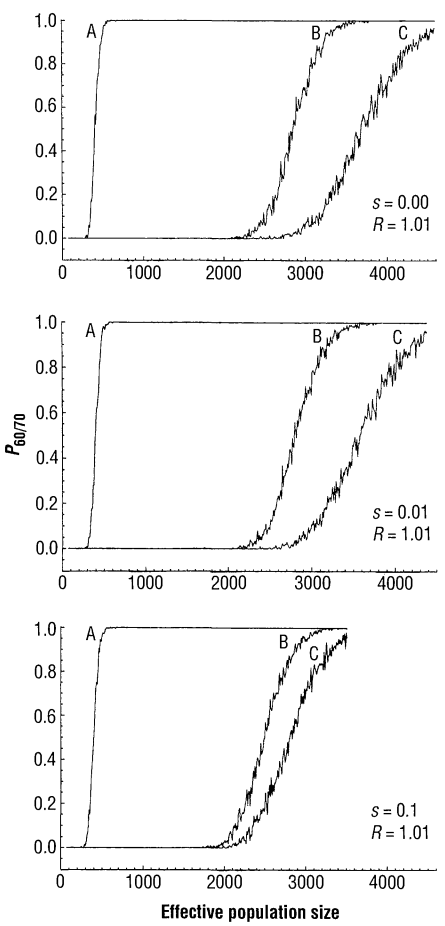

Fig. 8. Minimum size of bottleneck populations. The ordinate displays the average probability $P_{60/70}$ (based on 200 computer-simulated populations) that 60 of the initial 70 genes will persist. The selective advantage of the heterozygotes is s, and the rate of population growth per generation is $R = 1.01$. The bottleneck lasts 10 generations. A gives the probability of survival at the end of the bottleneck; B and C take into account the growth of the population to N = 10,000 and 100,000 individuals, respectively.

Census Populations

The parameter N that I have used for estimating the size of human populations is a theoretical construct that corresponds approximately to the number of synchronously reproducing individuals. The census number is likely to be about four or five times greater. In humans and in other primates, a number of individuals, possibly one-third of the total, do not reproduce at all. Of the females who reproduce, only about one-third are actively reproducing at any given time; the others are juveniles or are beyond reproductive age. With the use of these rough approximations, we can conclude that N is about two-ninths of the census population. A long-term effective population size of N = 100,000 corresponds, therefore, to a census population of 400,000 to 500,000 individuals. A bottleneck population of N = 4000 similarly corresponds to 15,000 to 20,000 census individuals. Population subdivision would affect the value of N estimated from the *DRB1* genealogy. If human ancestral populations consisted of several synchronous but relatively isolated populations, fewer individuals than were estimated above would be required to account for the *DRB1* polymorphisms, although the order of magnitude would not change (*18*).

The Origin of Modern Humans

The hominid lineage diverged from the chimpanzee lineage at ~6 Ma, and it evolved exclusively in the African continent until the emergence of *Homo erectus* somewhat before 1.7 Ma. The first known hominid, *Ardipithecus ramidus*, lived around 4.4 Ma, but it is uncertain whether it was bipedal or in the direct line of descent to modern humans (*27*). The recently described *Australopithecus anamensis*, dated at 3.9 to 4.2 Ma, was bipedal and has been placed in the line of descent to *A. afarensis*, *H. habilis*, *H. erectus*, and *H. sapiens* (*28*).

Shortly after its emergence in Africa, *H. erectus* spread to other continents. Fossil remains of *H. erectus* have been found in Africa, Indonesia (Java), China, the Middle East, and Europe (*29–33*). *Homo erectus* fossils from Java have been dated at 1.81 ± 0.04 and 1.66 ± 0.04 Ma (*30*), and from Georgia between 1.6 and 1.8 Ma (*31*). Anatomically distinctive *H. erectus* fossils deposited before 780,000 years ago have been found in Spain (*32*).

The transition from *H. erectus* to *H. sapiens* occurred around 400,000 years ago, but there is uncertainty as to whether some fossils are *H. erectus* or "archaic" forms of *H. sapiens* (*33*). Moreover, *H. erectus* persisted further in Asia, until 250,000 years ago in China and perhaps until 100,000 years ago in Java (*33*). The subspecies *H. sapiens neanderthalensis* appeared in Europe around 200,000 years ago and persisted until 30,000 years ago. The Neanderthals have been thought to be ancestral to anatomically modern humans, but now we know that modern humans appeared at least 100,000 years ago, much before the disappearance of Neanderthal fossils. It is puzzling that in caves in the Middle East, fossils of anatomically modern humans precede as well as follow Neanderthal fossils. Some modern humans are dated at 120,000 to 100,000 years ago, whereas Neanderthals are dated at 60,000 and 70,000 years, followed by modern humans dated at 40,000 years (*34*). It is unclear whether the two forms repeatedly replaced one another by migration from other regions, or whether they coexisted, or indeed whether interbreeding may have occurred.

There is considerable controversy about the origin of modern humans (*22, 33–36*). Some anthropologists argue that the transition from *H. erectus* to archaic *H. sapiens*, and later to anatomically modern humans, occurred consonantly in various parts of the Old World. Proponents of this "multiregional model" emphasize fossil regional continuity in the transition from *H. erectus* to archaic and then modern *H. sapiens*. However, they postulate that genetic exchange occurred from time to time between populations, so that the species evolved as a single gene pool, even though geographic differentiation occurred and persisted, just as geographically differentiated populations exist in other animal species (*29, 35*). This explanation is dependent on the postulate of persistent migrations and interbreeding between populations from different continents, of which no direct evidence exists. Moreover, it is difficult to reconcile the multiregional model with the contemporary existence of different species or forms in different regions.

Other scientists argue instead that modern humans first arose in Africa or in the Middle East somewhat earlier than 100,000 years ago and spread from there throughout the world, replacing the preexisting populations of *H. erectus* or archaic *H. sapiens* (*21, 23, 37–39*). Some proponents of this African replacement model argue further that the transition from archaic to modern *H. sapiens* was associated with a very narrow bottleneck, and that this bottleneck consisted of a small number of individuals who are the ancestors of all modern humans. The postulate of a narrow population bottleneck derives from a misunderstanding and need not be entertained. The most serious difficulty with the replacement model is that it leaves unexplained the apparent morphological continuity observed in some regions, most notably in Australasia (*29, 35*).

The reconstruction of the mtDNA genealogical tree places its root in Africa (*21, 23*), consistent with the African origin favored by proponents of the replacement

model (*37–39*). This evidence is, however, far from conclusive. The mtDNA root pinpoints only the ancestor in the maternal line, but the much larger nuclear DNA may have genealogical roots in other parts of the world.

Additional molecular evidence favoring African ancestry derives, however, from the analysis of 30 DNA polymorphisms from 14 worldwide populations (*38*). The genealogical tree derived from average genetic distances separates ancestral African from derived non-African populations. The deepest split in the genealogy is dated at ~156,000 years ago, which thus estimates the time when modern humans spread from Africa throughout the world. These results are particularly noteworthy because the analysis sought to determine the history of human populations rather than the ancestry of individual genes [see also (*39*)]. Even so, there seems to be no definitive reason to exclude the possibility that different genes may have different populational origins. The average distances would then reflect the relative genomic contribution of various ancestral populations. The results would thus be compatible with a model in which a modern African replacement was concomitant with some regional continuity (*40*).

In conclusion, the weight of the molecular evidence favors a recent African origin for modern humans. Ethnic differentiation between modern human populations would therefore be evolutionarily recent, a result of divergent evolution between geographically separated populations during the last 50,000 to 100,000 years. However, the replacement of archaic *H. sapiens* by anatomically modern humans may not have been complete everywhere. Some interbreeding between the colonizing modern humans and local populations would account for the apparent morphological continuity in some regions, particularly Australasia (*29, 35, 39*). The current blossoming of molecular evolutionary anthropology surely will soon provide more definitive and precise answers.

REFERENCES AND NOTES

1. The claim that evolutionists have shown that all modern humans descend from one or few women has repeatedly been made in the media, as well as in popular scientific publications [J. M. Lowenstein, *Pac. Discovery* **39**, 38 (1986)] and by scientists [W. M. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3605 (1980)].
2. A good introduction to various methods used in the molecular reconstruction of phylogeny is D. L. Swofford and G. J. Olsen, in *Molecular Systematics*, D. M. Hillis and C. Moritz, Eds. (Sinauer Associates, Sunderland, MA, 1990), pp. 411–501.
3. J. Klein, *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986); J. Kaufman, H. Völk, H.-J. Wallny, *Immunol. Rev.* **143**, 63 (1995); F.-W. Schwaiger and J. T. Epplen, *ibid.*, p. 199.
4. P. J. Bjorkman *et al.*, *Nature* **329**, 506 (1987); *ibid.*, p. 512; J. H. Brown *et al.*, *ibid.* **364**, 33 (1993).
5. R. E. Bontrop, *Immunogenetics* **39**, 81 (1994); R. E. Bontrop *et al.*, *Immunol. Rev.* **143**, 33 (1995); J. Klein and F. Figueroa, *Crit. Rev. Immunol.* **6**, 295 (1986); S. G. E. Marsh and J. G. Bodmer, *Immunogenetics* **33**, 321 (1991); *ibid.* **37**, 79 (1993); H. McDevitt, *Immunol. Rev.* **143**, 113 (1995); C. O'hUigin *et al.*, *Immunogenetics* **38**, 165 (1993); World Health Organization, *Bull. WHO* **70**, 801 (1992).
6. T. Bergström and U. Gyllenstein, *Immunol. Rev.* **143**, 13 (1995); W. Fan *et al.*, *Hum. Immunol.* **26**, 107 (1989); U. B. Gyllenstein and H. A. Erlich, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9986 (1989); U. B. Gyllenstein *et al.*, *ibid.* **87**, 1835 (1990); D. A. Lawlor *et al.*, *Nature* **335**, 268 (1988); W. E. Mayer *et al.*, *EMBO J.* **7**, 2765 (1988).
7. F. J. Ayala *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6787 (1994).
8. F. J. Ayala and A. A. Escalante, *Mol. Phylogenet. Evol.*, in press.
9. B. Arden and J. Klein, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2342 (1982); F. Figueroa *et al.*, *Nature* **335**, 265 (1988); T. J. McConnell *et al.*, *ibid.* **332**, 651 (1988).
10. This rate is obtained by the "minimum" method described in (*8*). See also Y. Satta *et al.*, in *Molecular Evolution of the Major Histocompatibility Complex*, J. Klein and D. Klein, Eds. (Springer, Heidelberg, 1991), pp. 51–62.
11. R. C. Griffiths, *Theor. Popul. Biol.* **17**, 37 (1980); R. R. Hudson, *Oxf. Surv. Evol. Biol.* **7**, 1 (1990).
12. J. F. C. Kingman, *Stochast. Processes Appl.* **13**, 235 (1982); *J. Appl. Probab.* **19**, 27 (1982); F. Tajima, *Genetics* **105**, 437 (1983); N. Takahata and M. Nei, *ibid.* **110**, 325 (1985); S. Tavaré, *Theor. Popul. Biol.* **26**, 119 (1984).
13. In the case of nuclear genes, the standard deviation for the mean is $>2N$ [see M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987), p. 395, eq. 13.74].
14. P. W. Hedrick, *Am. Nat.* **143**, 945 (1994); A. L. Hughes and M. Nei, *Nature* **335**, 167 (1988); *ibid.* **355**, 402 (1992); J. Klein and C. O'hUigin, *Philos. Trans. R. Soc. London Ser. B* **346**, 351 (1994); W. D. Potts and P. R. Slev, *Immunol. Rev.* **143**, 181 (1995); Y. Satta *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7480 (1993).
15. A. V. S. Hill *et al.*, *Nature* **352**, 595 (1991); A. V. S. Hill *et al.*, *ibid.* **355**, 403 (1992); A. V. S. Hill *et al.*, *ibid.* **360**, 434 (1992); L. H. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2415 (1994).
16. If u is the selected mutation rate per gene per generation and s is the selection coefficient, then
$$f_s \approx \left(\frac{\sqrt{2Ns}}{2Nu}\right)\left[\ln\left(\frac{Ns}{8\pi N^2u^2}\right)\right]^{-3/2}$$
[see (*7, 8, 17, 18*)].
17. N. Takahata, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2419 (1990); ——— and M. Nei, *Genetics* **124**, 967 (1990).
18. N. Takahata, *Mol. Biol. Evol.* **10**, 2 (1993).
19. Among the values in Table 3, more nearly appropriate for *DRB1* are $T = 4.0 \times 10^6$ (which corresponds to 60 My at 15 years per generation), $u = 2.0 \times 10^{-6}$ or 4.5×10^{-6}, and $s = 0.01$ or 0.02. The range corresponding to these values is $N = 149{,}072$ to 297,789 (Table 3).
20. Estimates on the order of $N = 100{,}000$ for ancestral human populations over the last 60 My are also obtained by analyzing the HLA *DQB1* polymorphisms (*8*).
21. R. L. Cann *et al.*, *Nature* **325**, 31 (1987); M. Stoneking *et al.*, *Genetics* **124**, 717 (1990); L. Vigilant *et al.*, *Science* **253**, 1503 (1991).
22. C. Wills, *Evolution* **49**, 593 (1995).
23. S. Horai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 532 (1995); M. Ruvolo *et al.*, *Mol. Biol. Evol.* **10**, 1115 (1993).
24. R. L. Dorit, H. Akashi, W. Gilbert, *Science* **268**, 1183 (1995). M. F. Hammer [*Nature* **378**, 376 (1995)] has analyzed a polymorphic fragment of the Y chromosome in 16 men of diverse geographic origin, and estimated the age of the ancestral sequence at 188,000 years ago. L. S. Whitfield, J. E. Sulston, and P. N. Goodfellow (*ibid.*, p. 379), based on analysis of a different Y chromosome fragment from five men, have estimated the coalescence at 37,000 to 49,000 years ago. The discrepancy between these two estimates and Dorit *et al.*'s estimate underscores the need for more extensive and accurate data. These new estimates are consistent with the conclusions reached here concerning the size of ancestral human populations.
25. Because alleles behave as neutral whenever $Ns < 1$, if the selection is weak (for example, $s = 0.01$), N must be correspondingly large (at least 100) for selection to play a role.
26. J. N. Spuhler, in *Genetics of Cellular, Individual, Family, and Population Variability*, C. F. Sing and C. L. Hanis, Eds. (Oxford Univ. Press, New York, 1993), pp. 262–297.
27. T. D. White, G. Suwa, B. Asfaw, *Nature* **371**, 306 (1994); *ibid.* **375**, 88 (1995); G. WoldeGabriel *et al.*, *ibid.* **371**, 330 (1994).
28. M. G. Leakey *et al.*, *ibid.* **376**, 565 (1995).
29. G. A. Clark, in *The Middle Paleolithic: Adaptation, Behavior and Variability*, H. Dibble and P. Mellars, Eds. (Univ. of Pennsylvania Museum, Philadelphia, 1992), pp. 183–205; *Recherche (Paris)* **263**, 316 (March 1994); ——— and J. M. Lindly, *Am. Anthropol.* **9**, 962 (1989).
30. C. C. Swisher III *et al.*, *Science* **263**, 1118 (1994).
31. L. Gabunia and A. Vekua, *Nature* **373**, 509 (1995).
32. E. Carbonell *et al.*, *Science* **269**, 826 (1995); J. M. Parés and A. Pérez-González, *ibid.*, p. 830.
33. S. Jones, R. Martin, D. Pilbeam, Eds., *The Cambridge Encyclopedia of Human Evolution* (Cambridge Univ. Press, Cambridge, 1992), pp. 246–257.
34. C. B. Stringer, *Sci. Am.* **263**, 98 (December 1990).
35. M. H. Wolpoff *et al.*, *Science* **241**, 772 (1988); A. G. Thorne and M. H. Wolpoff, *Sci. Am.* **266**, 76 (April 1992); D. M. Waddle, *Nature* **368**, 452 (1994); G. Bräuer, in *Continuity or Replacement? Controversies in* Homo sapiens *Evolution*, G. Bräuer and F. H. Smith, Eds. (Balkema, Rotterdam, Netherlands, 1992), pp. 83–98.
36. The controversy about the origin of modern humans has at times been heated and prone to overstatement. Recently, a science journalist quoted with implicit approval a distinguished scientist saying that "there is no genetic data that is *consistent* with the multiregional hypothesis" [A. Gibbons, *Science* **267**, 1272 (1995); my italics]. Apparently, neither the scientist nor the journalist reflected that Mendel's laws, the double helix, and the immense majority of genetic data are consistent with the multiregional hypothesis, as well as with the African replacement or any other model of human origins.
37. A. R. Rogers and L. B. Jorde, *Hum. Biol.* **67**, 1 (1995); C. B. Stringer, *Philos. Trans. R. Soc. London Ser. B* **337**, 217 (1992); ——— and P. Andrews, *Science* **239**, 1263 (1988).
38. D. B. Goldstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6723 (1995). The polymorphisms are for autosomal microsatellite loci.
39. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
40. Regional continuity is supported by some molecular evidence [W.-H. Li *et al.*, in *Genetics of Cellular, Individual, Family, and Population Variability*, C. F. Sing and C. L. Hanis, Eds. (Oxford Univ. Press, New York, 1993), pp. 253–261; S. S. Deeb *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7262 (1994); see (*7*)].
41. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
42. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
43. P. H. A. Sneath and R. A. Sokal, *Numerical Taxonomy* (Freeman, San Francisco, 1973).

44. I am grateful to A. Escalante for help with the computer simulations and data analysis, and to him and R. R. Hudson for helpful discussions. I am particularly indebted to J. Klein, who stimulated my interest in HLA polymorphisms; many ideas herein presented derive from work done in collaboration with him, although he should not be held responsible for what I now say. Supported by NIH grant GM 42397.