Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Human Evolution 49 (2005) 1–18

Genomics refutes an exclusively African origin of humans

Vinayak Eswaran [a], Henry Harpending [b,*], Alan R. Rogers [c]

[a] *Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India 208016*
[b] *Department of Anthropology, University of Utah, 270S 1400E, Salt Lake City, UT 84112, USA*
[c] *Department of Anthropology, University of Utah, 270S 1400E, Salt Lake City, UT 84112, USA*

Received 30 June 2004; accepted 8 February 2005

Abstract

Ten years ago, evidence from genetics gave strong support to the "recent African origin" view of the evolution of modern humans, which posits that *Homo sapiens* arose as a new species in Africa and subsequently spread, leading to the extinction of other archaic human species. Subsequent data from the nuclear genome not only fail to support this model, they do not support any simple model of human demographic history. In this paper, we study a process in which the modern human phenotype originates in Africa and then advances across the world by local demic diffusion, hybridization, and natural selection. While the multiregional model of human origins posits a number of independent single locus selective sweeps, and the "out of Africa" model posits a sweep of a new species, we study the intermediate case of a phenotypic sweep. Numerical simulations of this process replicate many of the seemingly contradictory features of the genetic data, and suggest that as much as 80% of nuclear loci have assimilated genetic material from non-African archaic humans.

Keywords: Modern human origins; Multiregional hypothesis; Out of Africa hypothesis; Phenotype sweep

Introduction

Since the discovery of apparent signals of strong late Pleistocene population expansions (Rogers and Harpending, 1992; Harpending et al., 1993) in human mitochondrial DNA (mtDNA), a number of studies have sought similar signs in other genetic polymorphisms. Among the data so analyzed have been nuclear sequences, short tandem repeat polymorphisms (STRs), and single nucleotide polymorphisms (SNPs). While mtDNA shows signals of recent expansions in almost every human population, it has by now become clear that the nuclear data do not present an unambiguous picture regarding population expansion associated

* Corresponding author. Tel.: +1 801 581 3776; fax: +1 801 581 6252.
E-mail addresses: eswar@iitk.ac.in (V. Eswaran), harpend@xmission.com (H. Harpending), rogers@anthro.utah.edu (A.R. Rogers).

0047-2484/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jhevol.2005.02.006

with the spread of anatomically modern humans. For example, various analyses of STR data using different statistics have given contradictory signals of expansions, their timing, and the sub-populations involved (Di Rienzo et al., 1998; Reich and Goldstein, 1998; Kimmel et al., 1998; Zhivotovsky et al., 2000). The first detailed evidence from the nuclear genome also showed no evidence at all of expansion (Harris and Hey, 1999).

To explain low interpopulation diversity in humans, it has been suggested that humans passed through a bottleneck (Haigh and Maynard Smith, 1972). It has also been proposed that there was a bottleneck associated with the emergence of modern humans in Africa and their spread throughout the world (Jones and Rouhani, 1986). For example, SNP haplotype block data show a signature of bottlenecks at vastly differing times in the prehistory of African and non-African populations (Reich et al., 2001; Gabriel et al., 2002), even if their cause as yet remains unclear. These bottlenecks would need to have been of extraordinary severity and/or duration to explain some of the data, e.g., for Europeans, Reich et al. (2001) suggested a pre-expansion bottleneck size of 50 individuals for 20 generations (or any size and duration of the same ratio), while Marth et al. (2003) obtained their best fit with size-duration ratios between 1000 to 2500 individuals for, respectively, 240 to 550 generations. Yet, single locus studies (e.g., Harding et al., 1997, 2000; Zhao et al., 2000; Yu et al., 2001, 2002) often find at most mild bottlenecks, or none, in non-Africans, resulting in an overall picture that is puzzling. A recent study by Marth et al. (2003) used 500,000 SNPs to conclude that the dominant population history of humans was a Pleistocene population collapse followed by a mild post-Pleistocene recovery. The importance of these varied signals of bottlenecks and expansions is the subject of this paper.

The significance of genetic signatures of late Pleistocene population expansions is that they directly address the contrasting theories of modern human origins that have been the subject of much debate since the 1980s. The recent African origin model (Cann et al., 1987; Stringer, 1992) proposes that anatomically modern humans arose in Africa around 130,000 years ago as a new species, which subsequently spread across the world, replacing all non-African archaic humans. In contrast, the multiregional evolution model proposes that modern humans emerged across the world from regional archaic human populations that were always linked by gene flow (Wolpoff et al., 1984).

Genetic evidence to date has been interpreted as giving far greater support to the recent African origin (RAO) model than to the multiregional evolution (MRE) one. Signs of population expansions in human mtDNA have seemed to confirm the RAO model [albeit in a modified "weak" form that suggests multiple bottlenecks and then expansions; Harpending et al. (1993)], as the African expansion seems to clearly pre-date the Asian and European ones. These signals have been seen as indicating the rise of modern humans in Africa and their subsequent expansion into the other continents.

It is now routinely assumed that such signals of expansions support the most controversial of the claims of the strict RAO model—that all non-African "archaics" were replaced and all living people are descended exclusively from African modern humans (Manderscheid and Rogers, 1996). For this to be true, the entire extant human genome has to be derived from recent Africans. Therefore, if the signals of expansions in mtDNA, for example, indicate a population history rather than merely the history of a single genetic locus, the same signals should be discernable at all genetic loci.

Others have proposed models intermediate between the strict RAO and MRE models (Smith, 1985; Relethford, 2001; Templeton, 2002). Relethford called his version "mostly out of Africa" because in it there is actual movement of populations from Africa. These newly arrived Africans mostly replace the local archaics, but there is some degree of admixture. On the other hand, in our model, there is no long distance movement of populations at all; change is driven entirely by local gene exchange among demes and natural selection. This model has the advantage of parsimony and simplicity, and it will be important to disentangle the effects of selection and long range migration from the archaeological and fossil records. For example, if there were long range population

movements with local hybridization, then signatures of that hybridization should persist and be discernible in the fossil record and in populations today. In contrast, our model posits that there are hybrids essentially only at the wavefront. Since this front is moving at something more than 3 km per generation, it would take about 30 generations, or 750 years, to travel 100 km. We would then expect to find hybrid-looking fossils only within this small temporal window.

Templeton's (2002) paper was an ambitious attempt to trace ancient gene flow from molecular markers. He examined the geographic distribution of subclades of several markers with an intuitively appealing logic that allowed him to date major movements to and from Africa. There are, however, several problems. His algorithm is regarded with skepticism by population geneticists (Felsenstein, 2003). Moreover, even if his algorithm were to identify real movements between ancestral populations, there is no information about where those populations were at the time. For example, a signature of movement between African and Asian ancestors several hundred thousand years ago might have been a movement between Africa and Asia under MRE, but a movement between adjacent river valleys, for example, under RAO since those ancestral populations would have been in Africa at the time. Templeton's findings provided almost no evidence for distinguishing among models of modern human origins.

A strongly negative value (e.g., ≤ -1.5) of the Tajima D statistic (T_D; Tajima, 1989) for a single locus likely indicates a selective sweep at that locus, while many such values obtained at independent loci would suggest a population expansion. Przeworski et al. (2000), analyzing data from 16 independent loci, found nearly evenly distributed positive and negative T_D-values, thus offering no support for putative population expansions. Stephens et al. (2001), analyzing data from 313 genes, found that 90% had negative values, but only a fraction of these were statistically significant.

In the RAO model, all loci should have strongly negative T_D-values, comparable to that shown in non-African mtDNA [$T_D = -2.28$; Ingman et al. (2000)]. Thus, the nuclear data do not consistently signal expansion, and when they do, the signal is of a mild expansion, perhaps reflecting only post-Pleistocene population growth associated with the spread of agriculture.

Alternative explanations for the puzzling features of the genetic data discussed above may be found in a recently proposed theory of modern human origins. Arguing along the lines of Sewall Wright's (1932) "shifting balance" theory, Eswaran (2002) suggested that the African transition to anatomical modernity may not have been a speciation event, but was rather a "character change" involving alleles at multiple loci that cooperated to confer a co-adapted genetic advantage to modern humans. Given small random movements of hunter-gatherer groups (demic diffusion), and under the condition of a low rate of interbreeding between modern and archaic humans, such an advantageous gene combination could spread as a wave of advance, or a "diffusion wave," of anatomical modernity (Eswaran, 2002).

One can visualize this process as that of the region of modern humans expanding at a steady rate into the region of archaic humans, the two regions being separated by a moving "wavefront" where the modern and archaic populations overlap. Only at the wavefront would both human types coexist; therefore, all hybridization and all selection favoring the moderns against the archaics—and thus all expansions in the modern population—would occur there.

According to this theory, the progress of the wave could be accompanied by considerable hybridization at the wavefront. Even so, the assimilation of archaic human genes into the modern populations would be low if the advantageous modern gene combination were complex enough that hybrids, with no selective advantage from their incomplete complement of modern genes, rarely became fully modern. Under such circumstances, the wave would essentially be an expansion of the modern humans at the wavefront. Further, as the small wavefront modern population would at any time be principally derived from previous wavefront moderns, the wavefront modern population would become severely bottlenecked over the thousands of generations that the wave took to travel from Africa to the far corners of Asia and Europe.

However, as all new modern populations would be created principally by the small wavefront modern population, the bottleneck would be followed by a continuous "rolling" expansion in the wake of the wave. As the signs of the wavefront bottleneck and the subsequent expansion would be passed on to the emergent modern populations, this theory offers an explanation for the bottleneck-and-expansion signature seen in so much human genetic data. It also explains why the expansions in Asia and Europe could have occurred tens of thousands of years after the African one (Harpending et al., 1993), for the wave would have traveled at about 3 km per generation, given the empirical evidence of the spread of modern humans.

Under conditions of a limited rate of archaic assimilation, only a few polymorphisms would survive at each locus in the bottlenecked wavefront modern populations. So, it is possible that African alleles often spread with the wavefront across the world. Such loci would then show signs of an expansion of a previously small set of African polymorphisms into a worldwide population. However, given a non-zero rate of assimilation from archaic populations, it is also possible that the wavefront moderns would, at some point along their spread, assimilate archaic human alleles (at loci unassociated with the functional advantages of modernity), which would then "surf" the wavefront and spread along with anatomical modernity. Thus, at these loci, the final modern world populations would have African/"modern" alleles, as well as alleles assimilated from archaic populations. The latter loci would show a considerable time depth and a corresponding lack of signs of expansions. This seems the most plausible explanation for the inconsistent signals of expansions obtained from various STR statistics (Eswaran, 2003), as well as the weak and variable signs of expansion in humans nuclear SNPs and the correlation among loci between Tajima's D and nucleotide diversity (Stephens et al., 2001).

Missing signs of expansions at many genetic loci would be correlated with assimilation at those loci from non-African archaic populations. Such assimilation would obviously also be compatible with evidence of great time depth in present-day non-Africans and of ancient and uniquely non-African polymorphisms (Harding et al., 1997, 2000; Zhao et al., 2000; Yu et al., 2001, 2002). It would also explain why significant geographical structuring (presumably ancient, and with partly archaic roots) is often seen at such loci, but not in others like mtDNA. The theory thus suggests that present day modern humans are not exclusively derived from early modern Africans, but have a significant genetic inheritance from non-African archaics as well.

In this paper, we explore this proposed scenario through simulations of a modified version of the numerical model of Eswaran (2002). We use the model to compute population statistics of the emergent modern populations for two cases simulating (a) the spread of modern humans from a regional source across a one-dimensional world through the replacement of archaic types, and (b) the analogous spread of a modern human type defined by an advantageous combination of some C unlinked genes. We show that while the replacement case reproduces some of the gross features of the genetic data, the model replicates its subtler details only in the assimilation case—thereby arguing that significant assimilation from non-African archaics accompanied the modern human transition. Our model is the simplest implementation of the idea of a coadapted gene complex, a phenotype, and the consequences for the neutral genome of a selective phenotype sweep.

Materials and methods

The Monte Carlo model used here simulates the diffusion of individuals on a one-dimensional world represented by 101 discrete locations. Individuals moved as a random walk with a parent/child variance distance, σ^2, of 320 km^2/generation, corresponding to a generational step-size of around 18 km. Locations are spaced 115 km apart and the array of locations thus represents a "world-length" of 11,500 km, roughly the distance from Ethiopia to China. Each location comprises 100 individuals, and thus the total simulated effective population size is 10,100 individuals, which is in keeping with the usual estimate of the long-term effective population size of humans. The population size at

each location is fixed, so there is actually *never* any expansion or bottleneck in the simulated population.

Each individual is represented by a genome comprising the C unlinked functional (genotypal) loci that make up the modern genotype, as well as numerous other neutral unlinked loci that are used to generate statistics. The genotypal loci are diallelic. Individuals are characterized as "modern," "archaic," or "hybrid" depending on whether they respectively have all "modern," "archaic," or a mixture of "modern" and "archaic" alleles in the C genotypal loci. The neutral loci have no role in determining modernity, and may each have numerous alleles created by the simulated processes of mutation and drift. All neutral loci are unlinked to each other and the genotypal loci. There is no recombination *within* any locus.

Each generation, the sub-population of 100 genomes at each location is re-generated by choosing 100 diploid progeny from random parents. Mating is assortative, with modern individuals mating largely with modern mates, and non-moderns (archaic or hybrid) with non-moderns. While in each mated pair the individual and its mate are first chosen randomly, an initially chosen non-modern mate of a modern individual is accepted with probability m_0 (≤ 1), or else a modern mate is again sought from the subpopulation. A similar procedure is used for initially chosen modern mates of non-modern individuals. The net effect of this procedure is to ensure that moderns and non-moderns mate with a probability that is m_0 times the random mating probability. The quantity m_0 is called the "interbreeding rate."

A procedure simulating natural selection operates every generation, with moderns enjoying a selective advantage of α per generation. As modernity here is defined as being homozygous with "modern" alleles at the C genotypal loci, this advantage of moderns over non-moderns is *co-adapted*, as it accrues only to the entire gene combination and not to any part thereof. This constraint requires that for complex combinations (with large C) to spread, the advantage must exceed the interbreeding rate m_0 (Eswaran, 2002).

The neutral loci of the simulated genome receive mutations every generation, the number of which is determined by the sequence mutation rate and the population size. These mutations are always unique (the infinite alleles model). For every mutation that occurs, a separate data structure stores information sufficient to recreate both the entire collection of created haplotypes, as well as their phylogenetic relationships. From the extant haplotypes, we can compute mismatch distributions, the number of segregating sites per locus, and any other data to generate standard population genetic statistics.

To create initial worldwide neutral diversity before the wave, the population is put through a process of burn-in involving mutation, mating, and diffusion starting from some ancient time at which all loci are assigned a root haplotype. At the onset of the diffusion wave, the subpopulation at the leftmost location ("Africa") is made modern, while the rest of the population is made archaic. This is done by respectively assigning modern and archaic alleles to the C functional genotypal loci in each individual genome. The neutral loci are left unchanged, and hence carry the pre-wave diversity. The selection process then operates, and the modern population spreads across the world-length in approximately 4000 generations. From the neutral diversity before and after the wave, we discern the effect of the wave and the signature it would leave in the emergent modern populations, which is then compared with the empirical evidence.

Model parameters

The basic model parameters are σ^2, C, α, m_0, and u, which respectively characterize the spatial diffusion, the complexity of the modern genotype, its selective advantage, the interbreeding rate between moderns and non-moderns, and the sequence mutation rate at the neutral loci. The formula $V \approx 0.8\sqrt{(2\sigma^2(\alpha - m_0))}$ approximates the wavefront speed per generation of complex genotypes (say, $C \geqslant 4$). The formula is loosely based on Fisher's (1937) equation for a single advantageous gene, with the factor 0.8 as an empirical correction for the finite discrete populations in our model. Here, $\sigma^2 = 320\ \mathrm{km}^2$/generation, which along with $\alpha = 0.03$/generation and

$m_0 = 0$ for the replacement case, or $\alpha = 0.07$/generation and $m_0 = 0.04$ for the assimilation case, gives a wave-speed of approximately 3.5 km per generation, compatible with the modern transition, which took 4000–5000 generations to reach China from northeastern Africa.

Recall that the modern genotype is modeled by C unlinked functional loci that are all required to be homozygous with modern alleles to have the modern selective advantage. Thus, C models the complexity of the modern genotype, and the rate of assimilation of archaic neutral genes into modern populations is increasingly restricted for higher Cs (Eswaran, 2002). For the cases involving assimilation, we use two values of the complexity parameter: $C = 1$, representing a low-complexity genotype, a single recessive gene; and $C = 8$, representing a high-complexity genotype, where a combination of eight unlinked loci determines modernity.

In the case of large C, the model is of speciation and no possible admixture, while in the other extreme case of $C = 1$ the model is of a single advantageous allele diffusing through the population. In the latter case there are few or no genomic consequences. For intermediate values of C, incorporation of archaic genetic material is possible, but less and less possible as C increases. Consider for example a mildly complex trait with $C = 4$. Offspring of the occasional matings between moderns and archaics would be heterozygotes at all 4 loci and would enjoy no reproductive advantage. If two hybrids mate, then 1/4 of their offspring would be homozygous modern at each locus and $(1/4)^4$, or 1/256, of such offspring would be fully modern at the selected loci, but each would carry 1/2 archaic-derived and 1/2 modern-derived alleles at other loci in the nuclear genome. In this way, increasing genomic complexity decreases the rate of admixture at non-selected loci.

Diagnostic variables

We compute several well known statistics describing genetic diversity in the population for each neutral locus: the mean pairwise difference (Π), the number of segregating sites (K), and the Tajima D statistic (T_D). The effective population (N_e) is estimated from the average mean pairwise difference and the prescribed sequence mutation rate. Site frequency spectra, i.e., the number of segregating sites at the locus showing the novel mutation at a particular frequency of appearance in a sample of chromosomes, are also evaluated for each locus. As averages over all loci are convenient, these spectra are normalized, and presented as frequency histograms.

There are two model variables designed to quantify the amount of non-African archaic assimilation that occurs at neutral loci. The first of these is the African parentage, which estimates the average assimilation in each individual simulated genome (Eswaran, 2002). At the initiation of the diffusion wave, this variable is assigned a value of 1 in the African moderns and 0 in the non-African archaics. Thereafter, during the mating procedure, the African parentage of each new progeny created is assigned a value equal to the average of its parents' values. If assimilation occurs, the averaged African parentage of moderns will be below unity. For example, a value of 0.6 would mean that the moderns have, on average, inherited 60% of their neutral genome from African moderns and the remaining 40% from non-African archaics.

The African parentage, however, quantifies only the average assimilation over the entire genome. In a diffusion wave, it is likely that assimilation would vary considerably among loci. It is thus important to keep track of the assimilation at each locus, which may then be correlated with the other locus statistics. This is done by a parameter called the assimilation coefficient.

To compute this parameter, each copy of every allele at the locus is tagged by its position relative to "Africa" at the time of the wave initiation. This information of the initial position is then passed on with the allele (and its descendents) in the mating procedure. Finally, after the wave, the global average of the initial positions of all extant alleles at the locus is computed and normalized with the average position of alleles at the time of wave initiation; this ratio is the assimilation coefficient. Briefly, its value is 0 if no assimilation takes place at the locus (so all extant alleles after the wave would have originated in "Africa"), and is 1 if there was

"perfect" assimilation (where the geographical distribution of the alleles, and direct ancestor-descendents, after the wave was essentially what it was before). The assimilation coefficient is very useful, as we shall see, in interpreting the mixed signals of expansions. Note that it is a random variable with a non-zero standard deviation, so that values of the statistic greater than 1 are possible with very high assimilation of archaic genes.

Simulated prehistories

Two different types of neutral loci are simulated here, the so-called "nuclear," and so-called "mtDNA," which differ in their mode of inheritance (diploid/haploid) and mutation rates, with the assumed nuclear mutation rate being much lower than the mtDNA rate. As these are forward simulations, each neutral locus is initially assigned a "root" haplotype from which all polymorphisms are produced and propagated by mutation, mating, localized diffusion, and, during the wave phase, selection. The simulations study the effect that a diffusion wave from Africa would have on the neutral genes of a subdivided world population with a given history. Two such histories are prescribed as follows:

a) For "nuclear" sequences, we assume a time depth from the root polymorphism at each locus starting 106,000 generations before the present time. For the first 80,000 generations, the simulated population of 10,100 individuals undergoes mutation and drift under conditions of global random mating. This is followed by 20,000 generations of local random mating, where individuals can mate only with others from the same location (of the 101 linearly spaced locations representing the world), while localized diffusion moves individuals between neighboring areas by a random walk process. The first 80,000 generations of the burn-in thus produces a sufficient time depth, comparable with what has been empirically observed in some nuclear data sets, while the next 20,000 generations produce some regional structuring in the population. The last could be thought of as isolation by distance that arose in human populations in the last half-million years since the transition from *Homo erectus* to "archaic" *Homo sapiens* (Stringer, 1992). At the end of this burn-in period of 100,000 generations, the diffusion wave is initiated from the leftmost area ("Africa") and allowed to run its course (which takes about 4000 generations). The present-day is taken to be 6000 generations after the start of the wave (and 106,000 generations from the initial root).
b) For "mtDNA" sequences the previous history is only 26,000 generations deep, also in keeping with conservative estimates of the time depth at this locus (Pesole et al., 1992; Wills, 1995). Of these, the first 20,000 generations after the root are characterized by local random mating, which provides geographical structure to the genetic profiles. The wave is then initiated, taking 4000 generations to cross the world, while the present day is 6000 generations after the start of the wave.

The mutation rate in our simulations for the nuclear sequences is 10^{-4} per generation, while that of the mtDNA sequences is 10^{-3} per generation. In the nuclear cases, 40 independent diploid loci were simulated, compared to 90 haploid neutral loci in the mtDNA cases. Mismatch distributions, site frequency spectrum, and other statistics are computed for each locus, and then averaged when necessary.

Simulations are done for (a) a low-complexity genotype case with $C = 1$, (b) a high-complexity genotype case with $C = 8$, and (c) a replacement case. The interbreeding rate in the first two cases is $m_0 = 0.04$, while the "modern" selective advantage is $\alpha = 0.07$. In the last case, the interbreeding is zero and the selective advantage is 0.02. These parameters are adjusted to obtain roughly the same wave-speeds. The African parentage estimates the fraction of the neutral genome that is inherited from Africans. In the low complexity case, the moderns furthest from Africa have almost none of their neutral genes inherited from Africans, while in the high complexity case they are 50–70% African, depending on the simulation. In the replacement case they are, of course, 100% African.

Results

Mitochondrial DNA mismatch distributions

The mismatch distribution is the histogram of the number of nucleotide differences at a locus between any two chromosomes chosen randomly from a population. In a population that has been essentially at a constant size for a very long time, the distribution is uneven and jagged, but when a strong expansion has recently occurred, it is unimodal, with a peak at approximately $2uT$ differences, where u is the sequence mutation rate per generation and T is the time since the expansion in generations (Rogers and Harpending, 1992). Human populations almost always show such unimodal distributions, indicating expansions (Harpending et al., 1993; Excoffier and Schneider, 1999).

Simulated mismatch distributions averaged over all 90 loci are shown in Fig. 1. The distributions at 26,000 (generations) are in the present day, with $C = 1$ representing the low-complexity genotype result, $C = 8$ the high-complexity genotype case, and "rep" the replacement case, with no interbreeding between moderns and archaics. The distribution at 20,000 generations is the common profile just before the wave initiation for all three cases.

Note that the distribution before the wave is unimodal. This "expansion," however, is spurious and due to the initial condition—the assignment of the "root" haplotype in 10,000 individuals is like a sudden expansion, the signature of which is not erased in the 26,000 generations of the simulation. To deduce this, note that that the peak of this distribution is at around 40 differences, which indicates, for the sequence mutation rate of $u = 10^{-3}$ per generation used, that the expansion is 20,000 generations before the initiation of the wave.

In the distributions at 26,000 generations, the low-complexity genotype $C = 1$ result shows the "initiation" peak at around 50 differences, now 26,000 generations old. The spread through the population of an advantageous allele at a single locus has little consequence for the rest of the genome. In particular, it does *not* produce the signal of an expansion.

The replacement case distribution after the wave is very different. It shows a strong peak at around 8 differences, which translates to a "sudden" expansion at around 4000 generations before the present—an average time of the actual diffusion wave's "rolling" expansion lasting from 6000 to around 2000 generations ago. Thus, a diffusion wave can produce, as was argued by Eswaran (2002), the same signal as a sudden population expansion, and so provide an explanation for the wave in human mtDNA data.

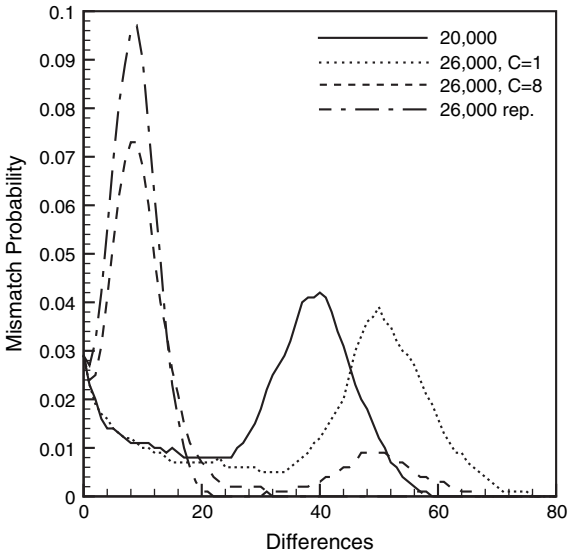

Fig. 1. The mtDNA mismatch distributions for the present-day world for the different cases, averaged over 90 independent loci. The numbers indicate generations after the root initialization: 20,000 generations is just before the onset of the wave, 26,000 generations is the present-day.

Perhaps more remarkable is the high-complexity genotype $C = 8$ case, which shows a peak similar, though somewhat weaker, than the replacement result. This case involves assimilation with the African parentage at the location farthest away from "Africa," being 0.69 in this simulation, indicating a 31% archaic admixture in the modern population there. While we believe—for reasons mentioned below—that archaic mtDNA was actually essentially replaced, it is interesting to see that a high-C assimilating diffusion wave would have a mismatch distribution that is difficult to distinguish from the replacement case. The main difference between the two in Fig. 1 is that the former allows the previous history to show through in the low peak on the right. In actuality, there would be

no "initiation" peak that appears, and it would be even more difficult to discriminate between the assimilation and replacement cases by their mismatch distributions.

Differing continental expansion times

An interesting observation that emerged from the mtDNA mismatch analysis of human diversity is that different continental populations apparently expanded at vastly different times—perhaps tens of thousands of years apart—with the Africans expanding first. Furthermore, these expansions seemed to have been from small groups that had separated very early, around the time of the African expansions (Comas et al., 1997). Although various scenarios have been proposed to explain this pattern (Harpending et al., 1993; Rogers and Jorde, 1995; Ambrose, 1998), none of these has proved compelling.

The diffusion wave mechanism gives an explanation for both the different expansion times and the early separation of the continental groups. Figure 2 shows simulated mismatch distributions at the present day, averaged over all 90 loci, for three locations at 22%, 52%, and 82% along the world length. The figure shows that the distributions peak respectively at about 10, 8, and 5.5 differences (translating to expansions about 5000, 4000, and 2750 generations ago). Figure 3 shows empirical mismatch distributions for the African, Asian, and European populations, which demonstrate a similar disparity of expansion times. These regional differences in the position of the peak are real and statistically significant according to simulations, as described in Harpending et al. (1993). This shows that the diffusion wave mechanism can explain the empirical observations. The slow-moving diffusion wavefront—in the wake of which these expansions would occur—takes several thousand generations to move between continents, explaining the differing continental expansion times.

The principal disparity between the simulated and empirical distributions is that the former shows a much greater frequency of 0 differences between two random individuals. This reflects the small local populations in the simulations, a mere

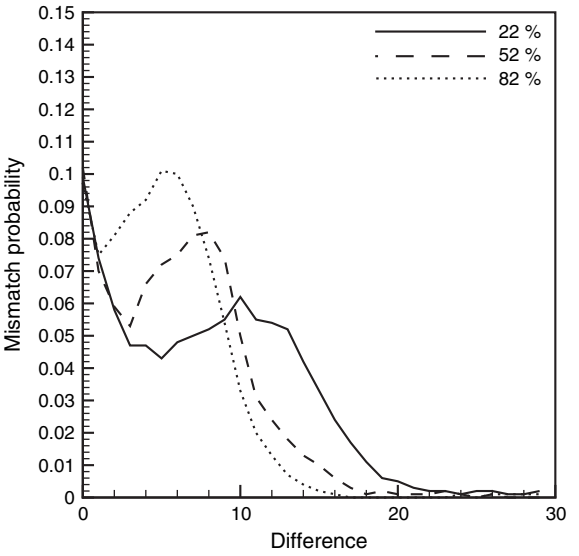

Fig. 2. The mtDNA mismatch distributions for the present-day world for the $C = 8$ case, averaged over 90 independent loci, at various locations along the world-length. The numbers indicate the percentage distance along the world-length.

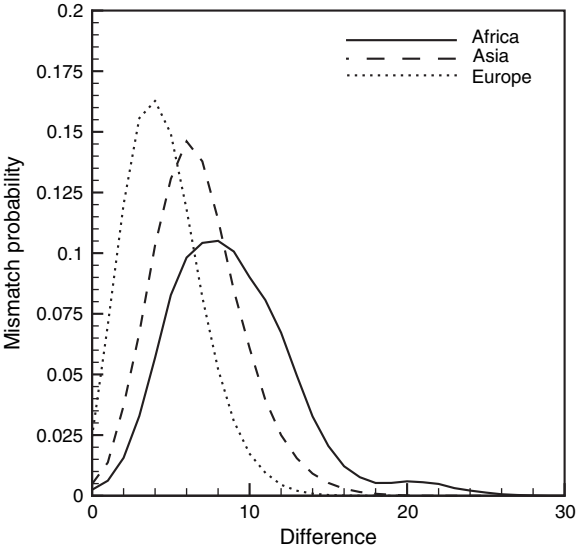

Fig. 3. The empirical mtDNA mismatch distributions for the three continental populations, computed by us from the sequence database maintained by Lynn Jorde, Department of Human Genetics, University of Utah.

100 individuals. Our simulations of a larger world population show stronger peaks and a lower frequency of 0 differences, more similar to the empirical data.

Nuclear mismatch distributions

Nuclear DNA simulations differ from those from mtDNA in having a substantially lower sequence mutation rate (10^{-4} per generation versus 10^{-3}) and a much longer burn-in period, extending 100,000 generations before the wave (see above). Forty unlinked loci are simulated for the same three cases of $C = 1$, $C = 8$, and replacement. The corresponding mismatch distributions from the simulated world populations at the present-day (106,000 generations after the start) and the common distribution for all three cases at the wave initiation (100,000 generations) are shown in Fig. 4. The expected equilibrium distribution for $N_e = 10{,}000$ is also shown.

The distribution at wave initiation shows an excess of moderate to high differences, a consequence of geographical subdivision, compared to the equilibrium expectation. The $C = 1$ distribution after the wave is similar, once again showing that ancient history and population subdivision are not erased by the spread of a single locus or a low-complexity genotype.

The replacement distribution after the wave has an extremely sharp peak at 0 differences compared to the equilibrium expectation. This is the characteristic signal of a population contraction—in a bottleneck, diversity is erased and the probability of two chromosomes being identical by descent is increased. The $C = 8$ assimilation case also shows a similar, if weaker, peak at 0, and indicates a contraction. The two cases otherwise differ mainly in that the assimilation distribution shows a longer tail at large differences, due to assimilated prehistory, than the replacement one. However, both show a contraction. These results are interesting because a recent construction of mismatch distributions from a large databank of SNPs (Marth et al., 2003) also yielded this result, suggesting a population collapse, quite contrary to the prior expectation among researchers of a signal of population expansion.

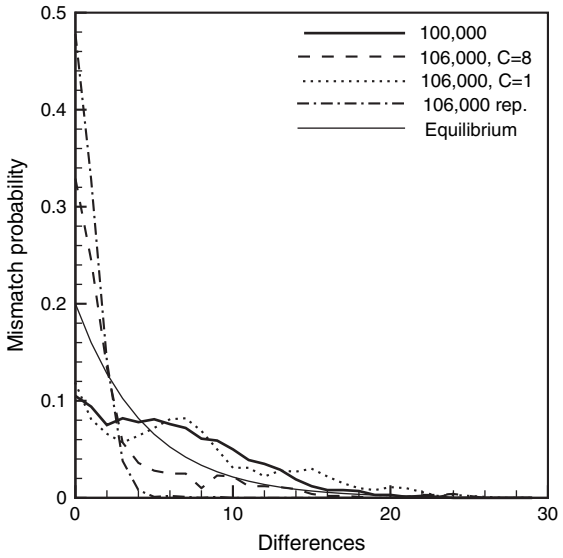

Fig. 4. The nuclear mismatch distributions for the simulated present-day world for the various cases, averaged over 40 independent diploid loci. The numbers indicate generations after the root initialization: 100,000 generations is just before the onset of the wave, 106,000 generations is the present-day.

However, the contrast of nuclear with mtDNA mismatch distributions needs an explanation. The diffusion wave of a complex genotype creates a bottleneck that lasts for thousands of generations in the wavefront, followed by a rolling expansion in its wake. Thus, genetic systems and statistics that have a quick response to demographic changes show an expansion or a bottleneck-and-expansion history. Conversely, systems with a slow response to demographic changes show only a bottleneck in non-African populations.

Mutation rate differences between mtDNA and nuclear sequences are sufficiently great that the higher-mutation-rate mtDNA could show strong expansions, while the nuclear sequences would indicate only the bottleneck. This explains the differences between these simulations—and so, too, seemingly, the empirical data.

Nuclear site frequency spectra

Although nuclear sequences have low mutation rates, they do signal recent expansions in ways other than through mismatch distributions. The

latter, and mean pairwise difference statistics, are relatively insensitive to new mutations that have not yet reached moderate frequencies in the population. However, K, the number of segregating sites at a locus in a sample of the population, is highly sensitive to new mutations that may arise in excess either after a selective sweep or a population expansion. The Tajima D statistic, which is the normalized difference between the population size estimates respectively obtained from the mean pairwise difference and the number of segregating sites, uses the contrasting sensitivity of the two to detect selective sweeps and/or population expansions.

The site frequency spectrum gives a detailed view of the segregating sites at a locus. It assumes that, as is commonly the case, the variation at each segregating site comprises only the ancestral nucleotide and a novel mutation. The site frequency spectrum is a histogram of the number of sites at the locus in which the novel allele respectively occurs with frequency 1, 2,..., $n-1$ in a sample of n chromosomes. The expected spectrum for a population at equilibrium is that the novel allele will appear with frequency i in K/ai sites, where $a \equiv ½ + ⅓ + ¼ + \ldots + 1/(n-1)$. An excess of sites with the novel allele appearing at low frequency suggests either a recent selective sweep at the locus or—if similar patterns are seen in many unlinked loci—a recent population expansion. An excess of sites with the novel allele at moderate to high frequency implies a population contraction, population subdivision, or balancing selection.

Spectra from many unlinked loci can be averaged if each spectrum is normalized by dividing it by K. The averaged normalized spectrum will have the equilibrium expectation of $1/ai$ for frequency $I = 1, 2,\ldots, n-1$. We represent spectra by binning data into five categories such that the expected number of sites in each bin is equal under a model of constant population size. Deviations from the standard model are then readily apparent.

In the simulations done here, frequency histograms were constructed and averaged over the 40 simulated loci from a sample of 1000 randomly drawn from the simulated global population. The histograms are shown in Fig. 5 for the population at the wave initiation (diamonds) and after the wave (gray circles) for the $C = 8$ case. The diamonds show that the profile at wave initiation has a slight excess at intermediate frequencies, indicating population subdivision. However, after the wave, the (gray) average shows a clear excess of low frequency variants, signaling expansions. Yet, the count in the highest frequency (fifth) bin is greater than in its neighboring (fourth) bin, suggesting that the expansion was preceded by a bottleneck. Such a bottleneck-and-expansion scenario also offers the best fit for the empirical site frequency spectra drawn from SNP data of non-Africans (Marth et al., 2004).

Returning to Fig. 5, a remarkable pattern emerges when the allele count after the wave is separately averaged over loci that had archaic assimilation (open circles) and those that did not (black circles). In the figure, the "replacement" loci with no assimilation show a clearly stronger excess of low-frequency counts and a deficit of high-frequency counts—thus indicating an expansion.

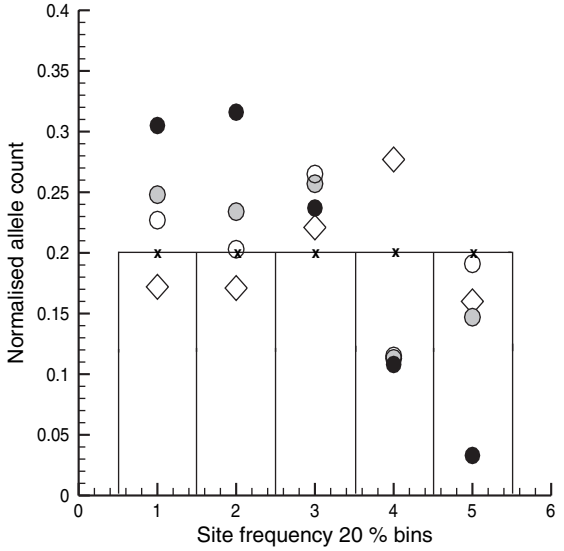

Fig. 5. Site frequency spectra for $C = 8$ from 40 loci, averaged and binned in histograms, with an expected 20% in each bin under equilibrium conditions. The circles are for data after the wave: the black circles are averages over loci that have no assimilation, the open circles are averages over loci with archaic assimilation, and the gray circles are overall averages. The diamonds show the overall averages just before the wave initiation.

In contrast, the assimilation cases show a bottleneck-and-expansion pattern that is also clearly closer to the equilibrium condition in every way than the replacement cases. The replacement loci thus show strong expansions, while the assimilation loci show bottlenecks followed by expansions. As the empirical data (Marth et al., 2004) collectively give the latter signals, these figures suggest that archaic assimilation, not replacement, is consistent with the nuclear data.

However, Fig. 5 shows that even loci with assimilation will often show an excess of low frequency mutations. Such patterns have been found to be ubiquitous in the nuclear genome. It is important to note that such an excess, while signaling expansions, does not preclude the possibility of assimilation. The contrast between the replacement and assimilation loci will become clearer when we examine the simulation results in more detail.

Other statistics

We now consider the other statistics obtained in the nuclear simulations in the light of the assimilation interpretation offered above. Figure 6 shows the scatter plot of global T_D-values versus mean pairwise differences for the 40 simulated loci at 100,000 generations after the start (i.e., just before wave initiation). The dashed lines indicate various levels of confidence for rejecting the null hypothesis of a constant population under no selection at that locus. While the values show wide variation, none of the T_D-values show even 90% significance (which at least some should in a random mating population). This result is the consequence of population subdivision, which is created in the last 20,000 generations of the burn-in (see above). The well known effect of population subdivision is to elevate T_D-values (essentially by elevating Πs). That the somewhat high $N_e =$ 14,875, computed from the average mean pairwise difference, is obtained, even when only 10,000 individuals are explicitly represented in the population, is also due to population subdivision. Note, however, that the simulated data show no signs of expansions in these conditions before the wave.

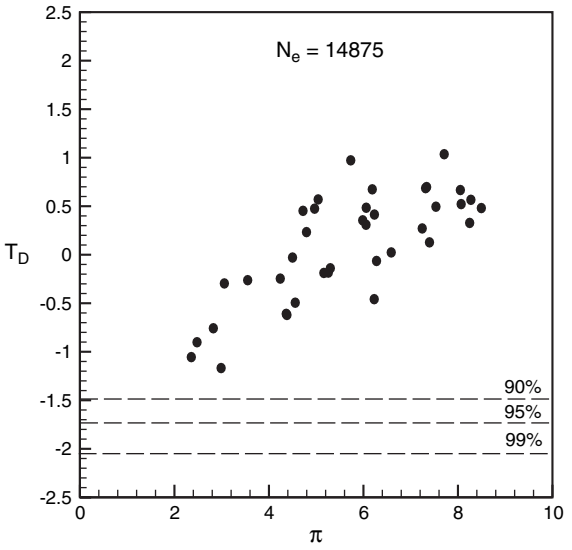

Fig. 6. Scatter plot of the Tajima D statistic (T_D) versus the mean pairwise difference (Π) at 40 independent nuclear loci just before the wave. The dashed horizontal lines indicate the significance levels for the Tajima D statistic for the chosen sample size of 1000 chromosomes. The effective size N_e is computed from the average mean pairwise differences and the mutation rate.

Figure 7 shows the scatter plot *after* the wave for the $C = 1$ case. The pattern remains similar to the pre-wave condition, showing again that a low-complexity genotype wave does little to change neutral gene patterns. Indeed, the effective population increases during the wave because of additional 6000 generations of population subdivision.

A radically different picture is seen after the wave for the replacement case in Fig. 8. Most of the loci show strong negative values, with at least 90% significance, and very low Π values. The effective population after the wave ($N_e = 2005$) is a drastic reduction from its pre-wave value of 15,780. Thus, the replacement case is accompanied by strong signals of expansions and greatly reduced diversity at *all* loci.

Figure 9 shows the scatter plot for the $C = 8$ assimilation case. Here there is a wide variation in both variables, which seem to be strongly correlated, with sharply negative T_D-values coming with low Π-values. Many of the former are significant to at least the 90% level. The effective population

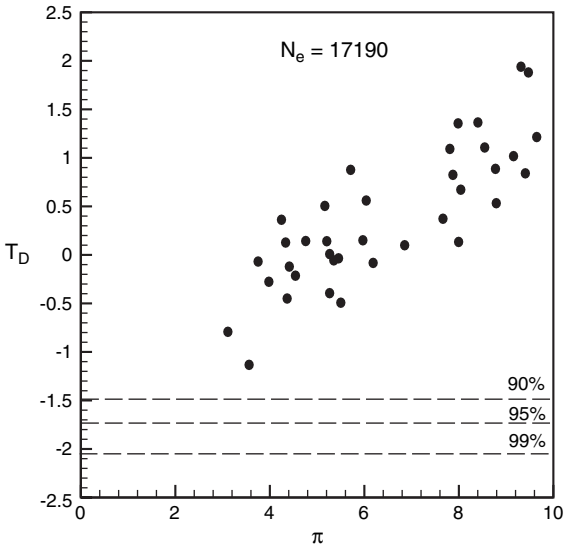

Fig. 7. Scatter plot of the Tajima D statistic (T_D) versus the mean pairwise difference (Π) at nuclear loci *after* the wave for the $C = 1$ case. The pre-wave $N_e = 15{,}780$.

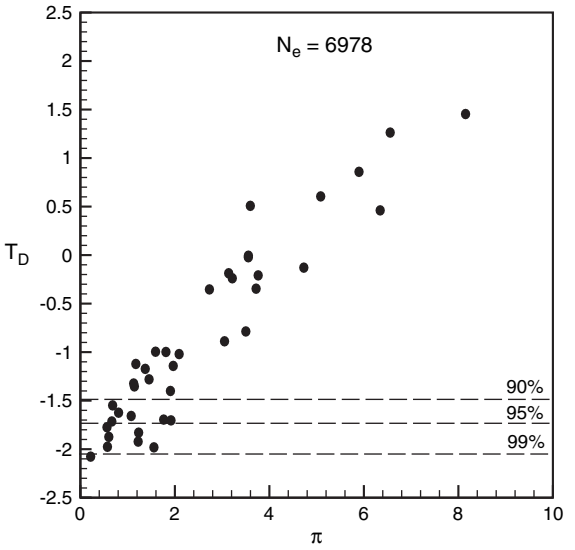

Fig. 9. Tajima D statistic (T_D) versus mean pairwise difference (Π) for the complex-genotype $C = 8$ case.

$N_e = 6{,}978$ after the wave is reduced from its pre-wave value of 14,875, but not as much as in the replacement case. Nevertheless, many loci show non-significant, and some even positive, T_D-values. The importance of this pattern becomes evident in the next section, where we relate the signal of expansions to assimilation. The pattern observed in the empirical data is not much different from these last cases.

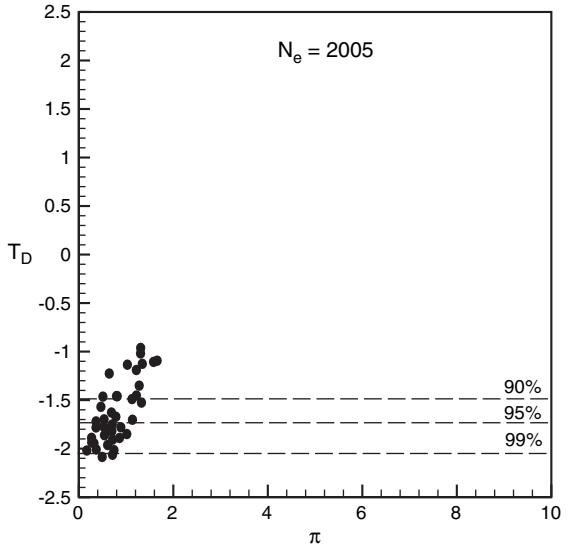

Fig. 8. Scatter plot of the Tajima D statistic (T_D) versus the mean pairwise difference (Π) at the nuclear loci after the wave for the replacement case. The pre-wave $N_e = 15{,}780$.

In a survey of 313 loci Stephens et al. (2001) found 90% had negative T_D-values (see Fig. 10), a pattern consistent with Fig. 9, suggesting that the shift to negative T_D-values in the simulations is remarkably similar to that in the empirical data, further supporting the proposition that the latter were shaped by a diffusion wave. In the figure, the horizontal axes show the nucleotide diversity, i.e., the mean pairwise difference (Π) normalized by the sequence length.

Expansions versus assimilation

For the same nuclear simulations as before, we now present scatter plots of T_D-values at each locus versus assimilation coefficient, A_s, at that locus. Recall that A_s is 0 when no assimilation occurs, and is 1 when there is "perfect" assimilation.

Figure 11 shows the results after the wave for the $C = 1$ case. Note that the A_s-values cluster around 1, and the T_D-values are biased slightly towards positive values because of population

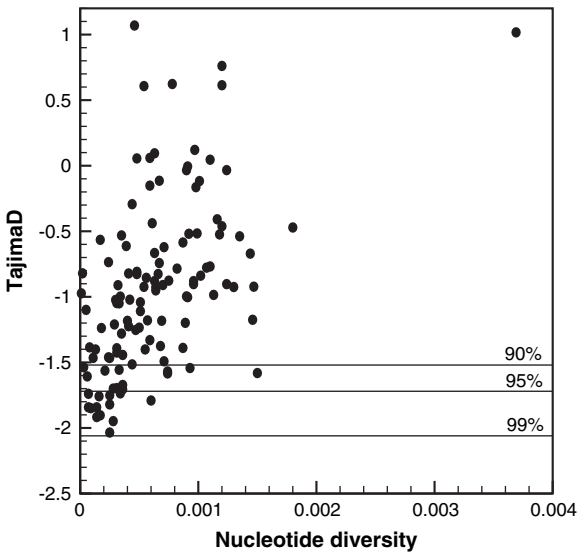

Fig. 10. Empirical data from Stephens et al. (2001) for 313 nuclear loci.

subdivision, as discussed above. The A_s-values confirm that low-complexity genotypes allow near-perfect assimilation, with the moderns of a region essentially inheriting the neutral genes of the archaics of the same region.

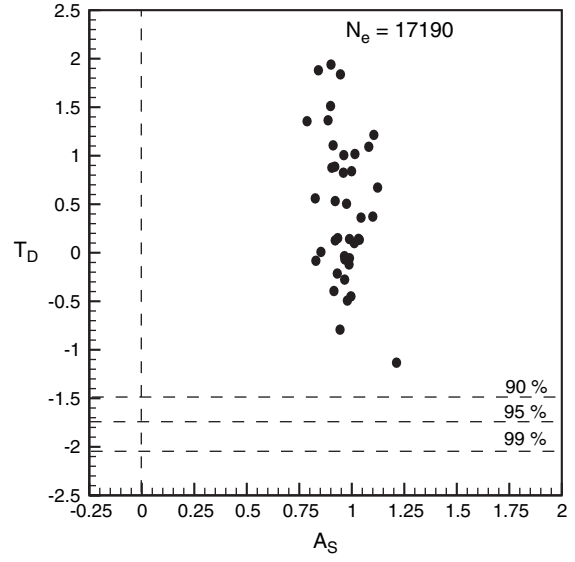

Fig. 11. A scatter plot of Tajima D (T_D) values versus the assimilation coefficient (A_s) for 40 simulated nuclear loci for the low complexity genotype $C =$, $P_f = 1$ case.

However, for the complex genotype $C = 8$ case shown in Fig. 12, there are many loci with A_s-values either 0 or close to it, and these are always associated with strongly negative T_D-values. Note that for cases of $A_s = 0$, the spread of T_D-values is similar to that of the "pure" replacement simulations. This is precisely as it should be; the neutral loci simulated are unlinked to each other and to the functional genes. Thus, a "replacement" locus will give Π- and T_D-values independent of whether or not the other loci are "replacement" or "assimilation," and determined only by other considerations such as the overall expansion factor.

The pattern is clear. The wavefront bottleneck causes a drastic reduction in worldwide diversity after a high-C wave. The bottleneck allows only a few alleles to survive at each locus. These alleles may, however, randomly be either African ones carried along by the wave, or non-African archaic variants assimilated along the wave path. Strongly negative T_D-values are likely to occur in loci where a few alleles—most likely African, but not necessarily so—fortuitously expanded into the entire global population, while loci in which different African and assimilated alleles are represented would show weaker signs of expansions, or none at all.

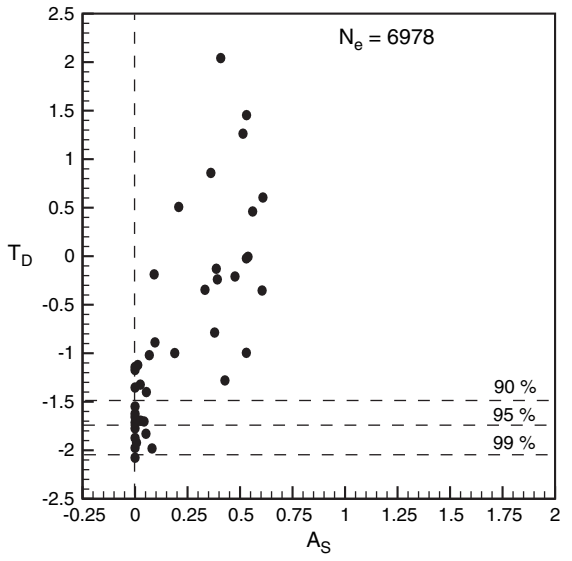

Fig. 12. Same as for Fig. 11, except the high-complexity genotype $C = 8$ case is plotted.

Thus, replacement at a locus is associated with strong signals of expansion, and assimilation with weaker, or no, signals of expansions. As the nuclear genome seems replete with loci offering contrasting signals of expansions, we must suspect that archaic alleles are endemic in the genome.

Discussion

How pervasive was assimilation?

The question immediately arises as to how widespread was archaic assimilation in the human genome. Even a rough estimate suggests that assimilation was surprisingly high—surprising because the debate until now has been whether there was any assimilation at all. The simulations conducted here have shown that significant T_D-values are strongly correlated with low assimilation, and conversely, that loci with high assimilation usually yield non-significant (but often negative) T_D-values.

The Tajima D statistic is ideally suited for detecting assimilation from archaic populations. Given the long history of humans living outside of Africa, there would have been significant geographical structure (which increases T_D) in the global human population at the time of the wave initiation, ca. 100,000 years ago. Thus, any assimilation from non-African archaic humans would inevitably increase T_D, reducing the possibility of a significant signal of expansion. Therefore, T_D is less likely to show expansions than other statistics if assimilation occurs. [See Yu et al. (2002) for a possible case in point.]

We can make a stronger assertion based on a comparison of nuclear loci with mtDNA. The deep differences between Neandertal and modern mtDNA, and the extremely low variability and geographical structure in modern mtDNA leaves little doubt that archaic mtDNA was largely, if not completely, replaced. The empirical T_D for mtDNA in non-African populations is -2.28 (Ingman et al., 2000). While such a low T_D-value is not obtained in the simulations of 10,000 individuals that we described above, it is well within the range of T_D-values obtained for simulated world populations of 30,000 or 50,000.

Recovery from the wavefront bottlenecks in these larger simulated populations leads to stronger signals of expansions, i.e., more negative values of T_D, even while the effective populations indicated by the mean pairwise differences remain below 10,000 at the simulated present day (as the pairwise differences have not fully recovered their equilibrium values after the wave). These latter simulations are, we believe, closer to the putative modern human diffusion wave, as they deliver highly negative T_D-values in the range empirically seen in mtDNA. However, the T_D-values for the replacement cases in these larger simulated populations nearly always fall within the 90% significance range ($T_D \leq -1.5$). This suggests that any loci with empirical T_D-values that are less than 90% significant are likely to have been affected by assimilation. For example, the "geography-based" sample of 437 loci presented by Ptak and Prezeworski (2002) shows fewer than 25% of the values are below -1.55, implying, by the above argument, that around 75% of the loci had significant assimilation. These are astoundingly high figures.

Other data, too, suggest pervasive assimilation. Scans of 624 STR loci by Storz et al. (2004) revealed that 13 of these had significantly reduced variability by their very strict criterion, which the authors attributed to selection, but which we think signals "replacement" loci. Even among the 13 loci, reduced variation occurred either in Europe or Asia, but rarely in both—which seems peculiar if selection were involved, but is entirely likely if random fixation of certain alleles carried by the wavefront independently occurred in the separate Asian and European waves. Only one of the 13 apparent sweeps was seen in Africa, where the diffusion wave model predicts some sub-Saharan populations would not have been subjected to the wave. Therefore, we believe the STR pattern better fits the diffusion wave hypothesis than the selectionist one. By looser criteria, approximately 25% of their loci showed this pattern of reduced variability outside of Africa—in either Europe or Asia, but not both. While Storz et al. (2004) proposed that the patterns are due to selection, it is important to note that they studied STRs with no known functional significance.

The immediate objection to the possibility of such pervasive assimilation in the nuclear genome may come from physical anthropologists. If such a degree of assimilation occurred during the modern human transition, they may ask, why did the modern human morphology remain essentially the same across the world, rather than showing physical signs of admixture with archaics from Asia and Europe? One possible answer to this question has already been given by Eswaran (2002), who argued that the modern morphology itself may have given the coadapted modern advantage—possibly due to reduced childbirth mortality—that propagated modernity. Thus, the modern morphology, and the alleles that "coded" for it, could be fixed in all modern populations even while the rest of the human genome carried a considerable number of assimilated archaic human alleles. Other aspects of modern human morphology could also have been selectively advantageous, but it is not immediately apparent what they could have been. Some correlated aspect of energy requirement and usage might be involved.

Assimilation at functional loci

With the exception of the *C* loci associated with modernity, the simulations presented here consider only neutral loci. Thus, the interpretation of the contrasting signals of expansions may be thought to hold only for such loci. To take an extreme view: is it possible that assimilation from archaics affected only neutral parts of the genome, while the functional parts were entirely derived from early modern Africans?

The evidence weighs heavily against this possibility. Many, even most, functional loci globally surveyed (e.g., β-globin, MC1R, PDHA1, Dys44, Y-chromosome, etc.) show deep structure both geographically and temporally, with coalescence times for non-African variation extending to much earlier time periods than the first emergence of modern humans in Africa. This strongly suggests that archaic assimilation affected these loci. The T_D values, too, are usually non-significant and even positive, not even remotely suggesting population expansions. Moreover, in functional loci, an expansion would cause strongly negative T_D-values in a random mating population, except in cases where deep phylogenetic structure was preserved by balancing selection before the expansion. Two of us (HCH, ARR) proposed balancing selection as an explanation for the lack of signals of expansions at functional loci (Harpending and Rogers, 2000), but such loci do not show some other signs of such selection (Wall and Przeworski, 2000). On the other hand, assimilation from archaic humans is a sufficient explanation for the missing signs of expansions, especially since loci with local selective advantage are more likely to be assimilated (Eswaran, 2002), and loci under balancing selection are also more likely to be assimilated.

Conclusions

The simulations presented here suggest resolutions for a number of crucial puzzles in the genetic data on modern human origins.

A diffusion wave of a complex genotype can explain why mismatch distributions of high-mutation-rate loci (such as mtDNA) show late Pleistocene expansions, while those of lower-mutation-rate SNPs show a contraction (Marth et al., 2003). It also explains why the more rapidly responsive site frequency spectra of SNPs show a bottleneck-and-expansion history (Marth et al., 2004). These explanations follow directly as consequences of a low assimilation-rate diffusion wave of moderns spreading out of Africa.

The same mechanism also explains why the expansions in Europe and Asia followed so late after the expansions in Africa (Harpending et al., 1993; Reich et al., 2001; Gabriel et al., 2002), while certain populations in sub-Saharan Africa show no signs of expansions (Excoffier and Schneider, 1999). The model suggests that these populations are directly descended from the first modern populations in the "core region," which would not be swept by the wave (Eswaran, 2002). But perhaps the most interesting explanations offered by the model concerns why—even among non-Africans—certain other loci do *not* show the characteristic bottleneck-and-expansion pattern and why, while most show mildly negative Tajima D values, there is so much variation in these values.

These empirical findings directly suggest that assimilation from archaic human populations accompanied the modern human transition across the world. The bottleneck at the wavefront, while greatly restricting the genetic diversity in the non-African (and north African) modern populations, randomly allowed—at least at neutral loci—either African polymorphisms to spread worldwide, or else allowed non-African polymorphisms to be assimilated and spread along with the wavefront. In the first case, we see strong signals of expansions, as in mtDNA; in the second case, we see less clear-cut signals of expansions, often accompanied by signs of deep population subdivision, significant numbers of unique non-African polymorphisms, and great time depths in non-African populations. The latter signs have been found in numerous nuclear loci studied in the last few years. We conjecture that as much as 80% of the nuclear genome is significantly affected by assimilation from archaic humans (i.e., 80% of loci may have some archaic admixture, *not* that the human genome is 80% archaic).

While each locus has its own history, the above reasoning suggests that African-dominated loci would all roughly tell the same story, while the others would each have its own—for assimilation would have varied in time and place in each case. Thus in the late 1990s, after a decade when most geneticists became convinced of the strict replacement recent African origin model, there was confusion when many nuclear loci—each in its own way—contradicted the patterns first seen in mtDNA. Yet, the following of that particular model remained strong, as there was no other theory that could explain the contrasting patterns. Now there is such a theory, and it tells us that while modern humans first emerged in Africa, living human populations carry within them a substantial genetic inheritance that had its origins in non-African archaics.

Acknowledgements

This paper was written while Vinayak Eswaran visited the University of Utah supported by a Fulbright fellowship.

References

Ambrose, S.H., 1998. Late Pleistocene human population bottlenecks, volcanic winter and differentiation of modern humans. J. Hum. Evol. 34, 623–651.

Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. Nature 325, 31–36.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Bertranpetit, J., 1997. Mitochondrial DNA variation and the origin of the Europeans. Hum. Genet. 99, 443–449.

Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M., Haines, G., Barch, D., 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148, 1269–1284.

Eswaran, V., 2002. A diffusion wave out of Africa: the mechanism of the modern human revolution? Curr. Anthropol. 43, 749–774.

Eswaran, V., 2003. On a diffusion wave as the genetic mechanism for the spread of modern humans: a reply to O.M. Pearson and A.C. Stone. Curr. Anthropol. 44, 559–561.

Excoffier, L., Schneider, S., 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc. Natl. Acad. Sci. 96, 10597–10602.

Felsenstein, J., 2003. Inferring Phylogenies. Sinauer Associates, Sunderland.

Fisher, R.A., 1937. The wave of advance of an advantageous gene. Ann. Hum. Genet. 7, 355–369.

Gabriel, S.B., Schaffner, H., Nguyen, J.M., More, E.A., Roy, J., 2002. The structure of the haplotype blocks in the human genome. Science 296, 2225–2229.

Haigh, J., Maynard Smith, J.M., 1972. Population size and protein variation in man. Genet. Res. 19, 73–89.

Harding, R.M., Fullerton, S.M., Griffiths, R.C., Bond, J., Cox, M.J., Schneider, J.A., Moulin, D.S., Clegg, J.B., 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. 60, 772–789.

Harding, R.M., Healy, E., Ray, A.J., Ellis, N., Flanagan, N.S., Todd, C., Dixon, C., Sajantila, A., Jackson, I., Birch-Machin, M.A., Rees, J.L., 2000. Evidence for variable selective pressure at MC1R. Am. J. Hum. Genet. 66, 1351–1361.

Harpending, H.C., Rogers, A.R., 2000. Genetic perspectives on human origins and differentiation. Annu. Rev. Genomics Hum. Genet. 1, 361–385.

Harpending, H.C., Sherry, S.T., Rogers, A.R., Stoneking, M., 1993. The genetic structure of ancient human populations. Curr. Anthropol. 34, 483–496.

Harris, E.E., Hey, J., 1999. X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. 96, 3320–3324.

Ingman, M., Kaessmann, H., Paabo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408, 708–713.

Jones, J.S., Rouhani, S., 1986. How small was the bottleneck? Nature 319, 449–450.

Kimmel, M., Chakraborty, R., King, J.P., Bamshad, M., Watkins, W.S., Jorde, L.B., 1998. Signatures of population expansion in microsatellite repeat data. Genetics 148, 1921–1930.

Manderscheid, E., Rogers, A.R., 1996. Genetic admixture in the late Pleistocene. Am. J. Phys. Anthropol. 100, 1–5.

Marth, G., Schuler, G., Yeh, R., Davenport, R., Agarwala, R., 2003. Sequence variations in the public human genome date reflect a bottlenecked population history. Proc. Natl. Acad. Sci. 100, 376–381.

Marth, G.T., Czabarka, E., Murvai, J., Sherry, S.T., 2004. Analysis of the allele frequency spectrum of genome-wide data reveal signals of differential demographic history in three large world populations. Genetics 166, 351–372.

Pesole, G., Sbisa, E., Preparate, G., Saccone, C., 1992. The evolution of the mitochondrial D-loop region and the origin of modern man. Mol. Biol. Evol. 9, 587–598.

Przeworski, M., Hudson, R.R., Rienzo, A.D., 2000. Adjusting the focus of human variation. Trends Genet. 16, 296–302.

Ptak, S.E., Prezeworski, M., 2002. Evidence for population growth in humans is confounded by fine scale population structure. Trends Genet. 18, 559–563.

Reich, D., Cargill, M., Bolk, S., Ireland, J., Sabeti, P., Richter, D.J., Lavery, T., Kuoyoumjian, R., Farhadian, S., Ward, R., Lander, E., 2001. Linkage disequilibrium in the human genome. Nature 411, 199–204.

Reich, D.E., Goldstein, D.B., 1998. Genetic evidence for a Paleolithic human population expansion in Africa. Proc. Natl. Acad. Sci. 95, 8119–8123.

Relethford, J., 2001. Genetics and the Search for Modern Human Origins. Wiley, New York.

Rogers, A.R., Harpending, H.C., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552–569.

Rogers, A.R., Jorde, L.B., 1995. Genetic evidence on modern human origins. Hum. Biol. 67, 1–36.

Smith, F.H., 1985. Continuity and change in the origin of modern humans. Z. Morphol. Anthropol. 75, 197–222.

Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharaya, T., 2001. Haplotype variation and linkage disequilibrium in 313 human genes. Science 293, 489–493.

Storz, J.F., Payseur, B.A., Nachman, M.W., 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol. Biol. Evol. 21, 1800–1811.

Stringer, C.B., 1992. Evolution of early modern humans. In: Jones, S., Martin, R., Pilbeam, D. (Eds.), The Cambridge Encyclopedia of Human Evolution. Cambridge University Press, Cambridge, pp. 241–251.

Tajima, F., 1989. Statistical method for testing the neutrality hypothesis by DNA polymorphism. Genetics 123, 585–595.

Templeton, A., 2002. Out of Africa again and again. Nature 416, 45.

Wall, J.D., Przeworski, M., 2000. When did the human population start increasing? Genetics 155, 1865–1874.

Wills, C., 1995. When did Eve live? An evolutionary detective story. Evolution 49, 593–607.

Wolpoff, M.H., Wu, X., Thorne, A.G., 1984. Modern *Homo sapiens* origins: a general theory of human evolution involving the fossil evidence of East Asia. In: Smith, F.H., Spencer, F. (Eds.), The Origins of Modern Humans: A World Survey of the Fossil Evidence. Alan Liss, New York, pp. 411–483.

Wright, S., 1932. The role of mutation, inbreeding, crossbreeding and selection in evolution. In: Jones D.F. (Ed.), Proceedings of the Sixth International Congress of Genetics, pp. 356–366.

Yu, N., Fu, Y.-X., Li, W.-H., 2002. DNA polymorphism in a world-wide sample of human X chromosomes. Mol. Biol. Evol. 19, 2131–2141.

Yu, N., Zhao, Z., Sambuughin, Y.-X.F., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L.B., Kuromori, T., Li, W.-H., 2001. Global patterns of human DNA sequence variation in 10-kb region on chromosome 1. Mol. Biol. Evol. 18, 214–222.

Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E., Pamilo, P., Trexler, M., Patthy, L., Jorde, L., Ramos-Onsins, S., Yu, N., Li, W.-H., 2000. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. 97, 11354–11358.

Zhivotovsky, L.A., Bennett, L., Bowcock, A.A., Feldman, M.W., 2000. Human population expansion and microsatellite variation. Mol. Biol. Evol. 17, 757–767.