Deep Haplotype Divergence and Long-Range Linkage Disequilibrium at Xp21.1 Provide Evidence That Humans Descend From a Structured Ancestral Population

Daniel Garrigan,*,† Zahra Mobasher,* Sarah B. Kingan,*
Jason A. Wilder*,† and Michael F. Hammer*,†,1

*Genomic Analysis and Technology Core, University of Arizona, Tucson, Arizona 85721 and †Department of
Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721*

ABSTRACT

Fossil evidence links human ancestry with populations that evolved from modern gracile morphology in Africa 130,000–160,000 years ago. Yet fossils alone do not provide clear answers to the question of whether the ancestors of all modern *Homo sapiens* comprised a single African population or an amalgamation of distinct archaic populations. DNA sequence data have consistently supported a single-origin model in which anatomically modern Africans expanded and completely replaced all other archaic hominin populations. Aided by a novel experimental design, we present the first genetic evidence that statistically rejects the null hypothesis that our species descends from a single, historically panmictic population. In a global sample of 42 X chromosomes, two African individuals carry a lineage of noncoding 17.5-kb sequence that has survived for >1 million years without any clear traces of ongoing recombination with other lineages at this locus. These patterns of deep haplotype divergence and long-range linkage disequilibrium are best explained by a prolonged period of ancestral population subdivision followed by relatively recent interbreeding. This inference supports human evolution models that incorporate admixture between divergent African branches of the genus Homo.

WHETHER the transition from the archaic to the anatomically modern human (AMH) form occurred in a single, isolated population or in a subdivided archaic population remains a contentious question in the study of human evolution. The answer to this question has important implications for understanding the extent of reproductive isolation among transitional forms of early *Homo sapiens* and the manner in which adaptations associated with anatomically modern traits were assembled in the human genome. For example, if *H. sapiens* evolved from a single isolated population, then our ancestors may have adapted to a narrower set of conditions than if our ancestors evolved in different ecological settings over a wider geographical range. There is now good fossil evidence that transitional forms lived over an extensive geographic area in widely separated parts of Africa at least 130 thousand years ago (KYA), from Morocco to South Africa (WOLPOFF 1999; MCBREARTY and BROOKS 2000). The earliest fossils with modern features were recently found in Ethiopia, dating from 154–160 KYA (WHITE *et al.* 2003), whereas slightly later specimens (*i.e.*, 90–130 KYA) are found throughout North, East, and South

Africa (STRINGER and ANDREWS 1988; BRÄUER 1992; KLEIN 1999; MCBREARTY and BROOKS 2000). The appearance of many early modern fossils across much of eastern and southern Africa, as well as the Levant, raises the possibility that the transition to the AMH phenotype may have occurred over a broad geographical range.

An alternative to interpreting the fossil record is to examine patterns of genetic variation in the extant human population and use evolutionary models to estimate the probability of historical events. The most abundant data on extant human genetic variation come from mitochondrial DNA (mtDNA). Variation at this maternally inherited haploid locus suggests that AMH descend from a single African population that rapidly expanded from a small effective population size (N_e) of ~10,000 individuals (CANN *et al.* 1987; HARPENDING *et al.* 1998). The time to a most recent common ancestor (TMRCA) for the mtDNA locus has been estimated at slightly over 170 KYA (INGMAN *et al.* 2000), which is coincidental with the fossil evidence for the emergence of the AMH phenotype. Similarly, variation in the paternally inherited haploid Y chromosome appears to trace back to an African common ancestor that lived only ~100 KYA (WILDER *et al.* 2004). These two sex-specific haploid loci have played a pivotal role in supporting the "recent African replacement" (RAR) model of human origins. The RAR model posits that AMH

evolved from a single African population and that this population subsequently expanded and replaced all contemporaneous forms of Homo, both inside and outside Africa, with negligible levels of interbreeding (CANN *et al.* 1987; HAMMER *et al.* 1998; EXCOFFIER 2002).

Genetic variation in the nonhaploid regions of the nuclear genome presents a more ambiguous picture of human history and, in some cases, suggests that human population structure may predate the AMH phenotype. While deep lineages of mtDNA are found in some African hunter-gatherer populations, these lineages trace back only to the time of the global most recent mtDNA ancestor (*i.e.*, <170 KYA). In contrast, nucleotide variation at the X-linked *PDHA1* locus is partitioned into two lineages that are estimated to have diverged in Africa ~1.9 million years ago (MYA) (HARRIS and HEY 1999). This TMRCA is older than expected for an X-linked locus based upon the value of N_e estimated from mtDNA and most nuclear variation (EXCOFFIER 2002; TAKAHATA *et al.* 2001). One viable explanation for this unexpected antiquity is that the AMH population is descended from multiple archaic subpopulations. One expected consequence of ancient population subdivision is to increase the historical N_e and, thus, push back the TMRCA. This increased N_e may not be visible at all loci, depending upon the proportion of the genome descending from multiple subpopulations. It should also be noted that because the *PDHA1* locus encodes an enzyme essential for glucose oxidation, the possibility remains that natural selection has maintained *PDHA1* lineages for prolonged periods of time (HARDING 1999; HARRIS and HEY 1999). Other loci with unusually deep TMRCAs have also been discovered (HARDING *et al.* 1997; ZHAO *et al.* 2000; ZIETKIEWICZ *et al.* 2003; STEFANSSON *et al.* 2005). One interesting example comes from an X-linked pseudogene, *RRM2P4*, with a 2-million-year-old lineage that is found almost exclusively in East Asia (GARRIGAN *et al.* 2005).

In this report, we present novel sequence data from a 17.5-kb X-linked locus, Xp21.1, that exhibits ancient divergence among lineages. We analyze levels of haplotype divergence and linkage disequilibrium (LD) in the framework of models predicting patterns of nucleotide variation expected as a consequence of admixture between historically isolated subpopulations (WALL 2000; NORDBORG 2001). No previous human polymorphism study has been specifically designed to utilize these measures to assess the probability of a single population origin for AMH. The Xp21.1 locus was chosen for sequencing because it is located in a noncoding region of the X chromosome with moderately high levels of recombination and very low gene density. These criteria serve to minimize the potential impact of natural selection acting on linked functional sites. The unusually high levels of haplotype divergence and LD observed at the Xp21.1 locus fit the expectations under a model of isolation and admixture. Additionally, Monte Carlo
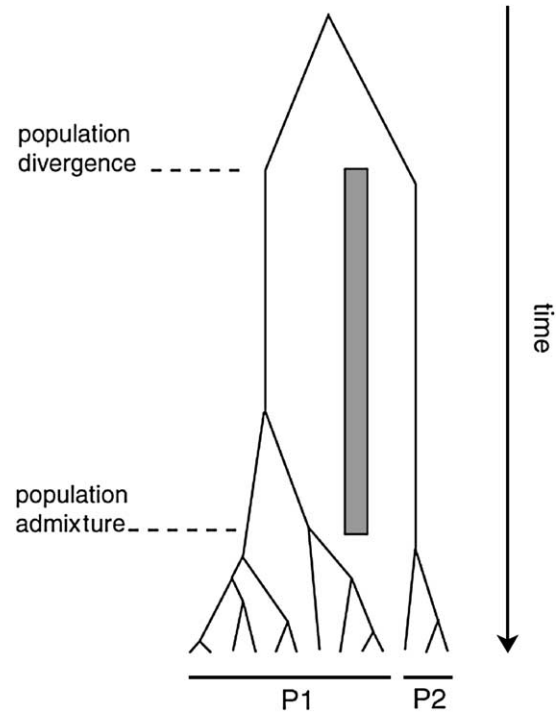

FIGURE 1.—Tracing the hypothetical ancestry of a sample of 13 chromosomes under the isolation-and-admixture model. Three of the 13 chromosomes, labeled P2, descend from a historically isolated subpopulation. The longer subpopulations spend in isolation (delimited by the shaded vertical bar), the longer is the proportion of genealogical time occupied by only two lineages. The resulting basal branches are expected to be longer under this model of subdivision compared with those of a single randomly mating population. Longer branches are expected to accumulate more mutations that will partition the data set into two distinct haplotype lineages. Upon population admixture, long-distance linkage disequilibrium is expected, which decays as time goes on.

computer simulations show that it is highly improbable that this pattern of nucleotide variation could arise in a single, panmictic population, as predicted by the RAR model. We consider several plausible alternative hypotheses and conclude that ancient population structure in the evolutionary lineage leading to AMH is the most likely explanation for the Xp21.1 data.

SUBJECTS AND METHODS

Experimental design: Our study design is based on the model first proposed by NORDBORG (2001) and extended by WALL (2000). This model of population subdivision seeks to predict patterns of nucleotide variation resulting from recent admixture between two divergent subpopulations. While this is not the only plausible model of ancient human subdivision, it does make explicit, testable predictions that distinguish it from the null hypothesis of panmixia. Figure 1 illustrates the expected changes in the genealogical history of a hypothetical genetic sample induced by the combination

a

```
                                         nucleotide position
                    1 1 1                8 8 9 9 9 9 9      1 1 1 1 1 1 1 1 1 1
            1 2 4 7 0 2 3                4 5 2 3 4 4 6      5 5 5 5 6 6 6 6 6 7 7
            5 1 3 3 6 9 1                9 0 7 1 0 9 3      5 5 5 6 6 0 4 6 6 3 3
 Haplotype  5 7 0 8 9 0 6                1 2 4 6 4 5 8      4 4 6 8 0 6 2 3 4 4 6
                                                           2 3 8 0 0 6 5 1 8 5          frequency
                                                                                       Af  nA  total
 Chimp      A A C G C A G   A A C A G A A   T G G A C T T C A T                          
 A          . . G . T C .   . G . . . . G   . . . . T . . . T C                          2   0    2
 B          . . . T . . A   C . A . . . .   . A A . . . . A . .                          4   9   13
 C          . C . T . . .   . . A . . . .   C A A . . . . A . .                          1   0    1
 D          . . . T . . .   . . A . . . .   . A A . . . . A . .                          1   0    1
 E          . . . T . . .   C . A . . . .   . A A . . . . A T C                          1   0    1
 F          . . . T . . .   . . A . . . .   . A A . . . . A T C                          1   0    1
 G          C . . T . . .   . . A . . . G   . A A . . . G A T C                          0   1    1
 H          C . . T . . .   . . A . T G .   . A A . . . G A T C                          0   1    1
 I          C . . T . . .   . . A . . G .   . A A . . . G A T C                          0  11   11
 J          C . . T . . .   . . A . . . .   . A A . . . G A T C                          0   1    1
 K          C . . T . . .   . . A . . G .   . A A . . . . A T C                          0   1    1
 L          C . . T . . .   . . A . . G .   . A A G . . G A T C                          0   1    1
 M          . . . T . . A   C . A . . . .   . A A . . C . A . .                          0   1    1
 N          . . . T . . .   C . A G . . .   . A A . . . . A . .                          0   3    3
 O          . . . T . . .   . . A G . . .   . A A . . . . A . .                          0   1    1
 P          . . . T . . .   C . A . . . .   . A A . . . . A . C                          0   1    1
 Q          C . . T . . .   . . A G . G .   . A A . . . G A T C                          0   1    1
            + + + +         + +             + +     +     +
```
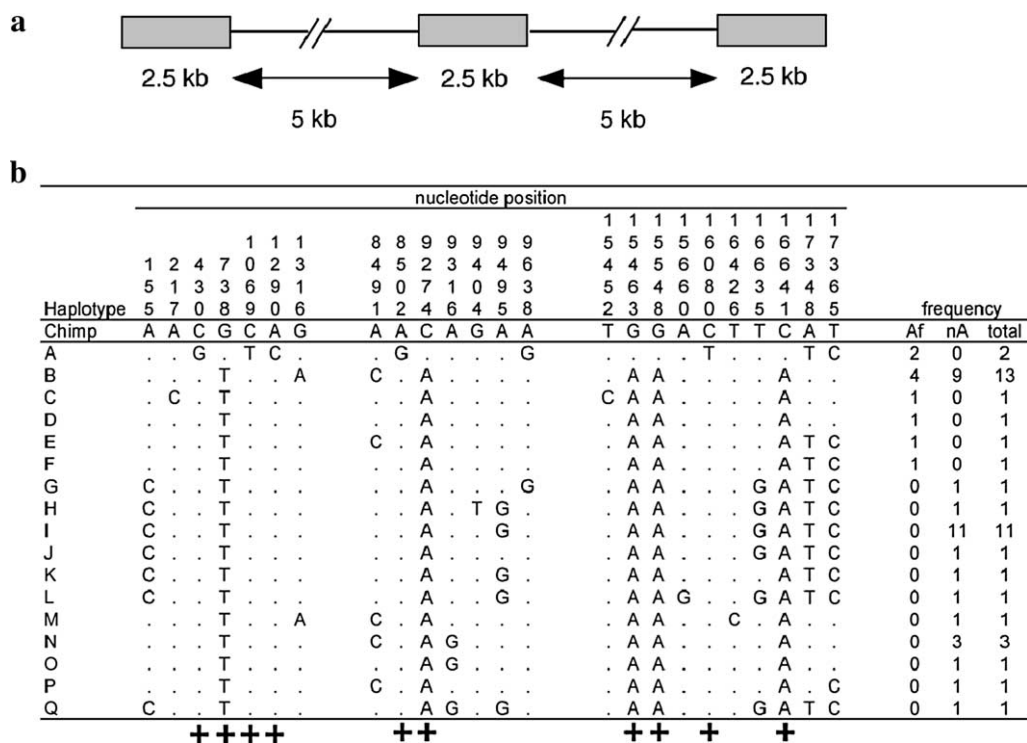
FIGURE 2.—The Xp21.1 polymorphism data set. (a) Schema of the locus trio design to assay nucleotide divergence among haplotypes and linkage disequilibrium. (b) Nucleotide variation found at Xp21.1. A dot indicates identity with the chimpanzee and gorilla outgroup sequences, which are assumed to be the ancestral human state. Congruent sites are labeled with "+." The last three columns list the frequency of each haplotype in African, non-African, and the total sample. The mutation occurring at site 9638 occurs in a guanine/cytosine dinucleotide repeat, making it unclear whether this is a recombination event or a parallel mutation.

of ancient population subdivision followed by recent admixture. Three parameters are of particular interest in this model: the time of onset for population divergence, the time of admixture, and the admixture proportion. We hereafter refer to this model specifically as the "isolation-and-admixture" (IAA) model. In the IAA model, the two basal branches are expected to be longer than those of a panmictic population, depending upon the length of time subpopulations spend in isolation. Neutral mutations arising in each isolated subpopulation would be unable to recombine with one another until after the admixture event, resulting in long-range LD among nucleotide sites in the extant sample. Within the context of this model, WALL (2000) has shown that measures of sequence LD are powerful indicators of past population subdivision. One such measure of LD is the number of mutations occurring on the two basal branches of the genealogy (*i.e.*, polymorphisms that, on a pairwise basis, result in only two haplotypes). Sites that partition the data in this way are referred to here as "congruent" sites after WALL (1999). The observed number of congruent sites is quantified by the variable l_b. The maximum physical distance between congruent sites is also a powerful measure to distinguish between panmixia and the IAA model (WALL 2000). This physical distance among sites is quantified by the variable g_d.

The power to reject single-population ancestry is improved when LD is examined over a minimum of ~20 kb of sequence, depending upon local recombination and mutation rates (J. D. WALL, personal communication). This conclusion suggests that other loci showing deep haplotype divergence, such as *PDHA1* (4 kb) (HARRIS and HEY 1999) and *RRM2P4* (2.4 kb) (GARRIGAN *et al.* 2005), are too short to have sufficient power to reject the null hypothesis of panmixia if it was, in fact, false. The Xp21.1 locus consists of three 2.5-kb nucleotide sequence fragments, each separated by an intervening 5 kb of unsequenced DNA, thus spanning a total of 17.5 kb (Figure 2a). This economical "locus trio" design is being implemented by our laboratory in an ongoing study of noncoding polymorphism in regions of the genome with higher than average levels of recombination and low gene density. The Xp21.1 locus is the first complete data set to emerge from this long-term study.

Population sampling and DNA sequencing: Human genomic DNAs were isolated from lymphoblastoid cell lines established by the Y Chromosome Consortium (Y CHROMOSOME CONSORTIUM 2002) at the New York Blood Center from blood donated by volunteers who gave informed consent. The New York Blood Center and the University of Arizona Human Subjects Committees approved all sampling protocols. A total of 42 hemizygous males were sampled, including 10 Africans (2 Tsumkwe San from Nambia, 1 West Bantu Herero, 1 East Bantu Pedi, 1 East Bantu Sotho, 2 Biaka Pygmies from Central African Republic, and 3 Mbuti Pygmies from Zaire), 12 Asians (3 Han Chinese, 2 Siberian Yakuts, 1 Cambodian, 3 Japanese, 1 Kashmir, and 2 Nasioi from Melanesia), 10 Europeans/Near Easterners (2 Ashkenazi Jews, 1 British, 1 Adygean from Krasnodar, 3 Germans, 2 Western Russians, and 1 Turk), and 10 Americans (1 Navajo, 1 Tohono O' Odham, 1 Poarch Creek, 2 Karitianas and 2 Surui from Brazil, 1 Mayan, and 2 Amerindians of unknown tribal affiliation). A

single male common chimpanzee (*Pan troglodytes*) was sequenced from DNA provided by O. Ryder to provide information on the ancestral state of human polymorphisms. As an additional outgroup, gorilla (*Gorilla gorilla*) DNA was also sequenced, which was provided by the Lincoln Park Zoo in Chicago.

Genomic DNA was PCR amplified in a 25-μl volume reaction with 40 cycles. Variable conditions for each amplification and primer sequences are available from the authors upon request. Internal primers (also available upon request) were used to generate overlapping sequence runs on an ABI 3730XL automated sequencer. Contiguous haplotype sequences were assembled for each individual and aligned using the computer application Sequencher (GeneCodes, Ann Arbor, MI).

Data analysis and coalescent simulations: The population mutation parameter θ (= $3N_e\mu$, where N_e is the diploid effective population size and μ is the rate of neutral mutation per site per generation) and the population recombination parameter ρ (= $2N_e r$, where r is the rate of recombination per site per generation) were simultaneously estimated by the full-likelihood, importance-sampling method (FEARNHEAD and DONNELLY 2001) that approximates an optimal proposal density of parameters for the ancestral recombination graph. We use the effective sample size (ESS) as a diagnostic of the importance sampling efficiency and accept parameter estimates when the ESS $\geq$100. It should be noted that the parameters θ and ρ are estimates of mutation and recombination rates in an ideal population that conforms to the assumptions of the Wright-Fisher model. Polymorphism statistics and the minimum number of recombination events were calculated with the computer application DnaSP (ROZAS *et al.* 2003). Lineage divergence times (t_1) were estimated by taking the ratio of the net proportion of nucleotide substitutions (NEI 1987) between lineages (d_1) to that of humans and the chimpanzee and gorilla outgroups (d_o). We assume that the divergence time (t_o) for the human-chimpanzee comparison is 6 MYA and that it is 8 MYA for the human-gorilla comparison. We can then estimate lineage divergence time as $t_1 = (d_1 t_o)/d_o$. Similarly, the neutral mutation rate per year (μ) was estimated as $\mu = d_o/2t_o$.

Coalescent simulations of panmixia were performed with a modified version of the computer program "ms" (HUDSON 2002). Simulated 17.5-kb data sets were generated with the maximum-likelihood estimates of ρ and θ. An additional computer program was written to calculate l_b and g_d from the output of ms (program available from the authors upon request). Because the simulated data sets were not constrained to produce the observed numbers of segregating sites, we standardized the number of congruent sites by calculating the proportion of total segregating sites that can be classified as congruent. Additional simulations that mimic the locus trio sequencing design yielded results qualitatively equivalent to the contiguous 17.5-kb simulations (Figure S1 at http://www.genetics.org/supplemental/). Be-
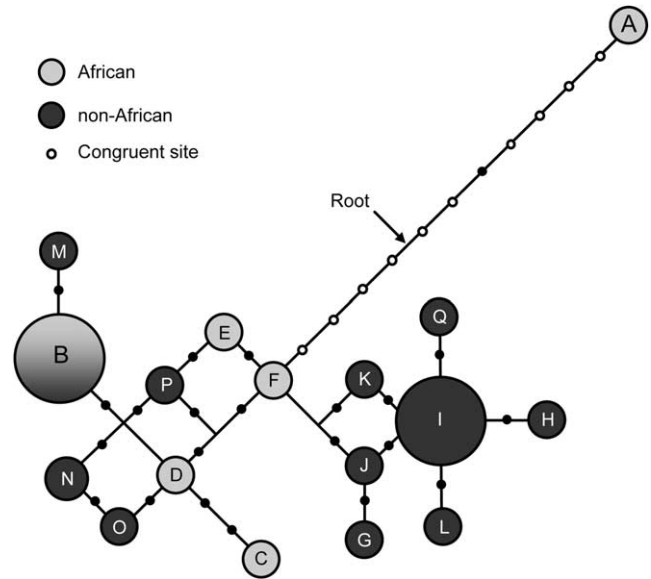

FIGURE 3.—Minimum-spanning haplotype network for the Xp21.1 locus trio. Haplotypes found exclusively in African populations are represented by shaded circles, while haplotypes found only in non-African populations are shown as solid circles. Solid dots along branches denote mutation events, while open dots denote mutation events that are classified as congruent sites in the text. The chimpanzee outgroup sequence indicates the root is located in the position pointed to by the arrow. Reticulating branches are likely the result of historical recombination events.

cause both l_b and g_d are sensitive to the assumed value of ρ, to be conservative, we also simulated panmictic genealogies with values of ρ that are lower than the maximum-likelihood estimate (Figure S2 at http://www.genetics.org/supplemental/).

RESULTS

Patterns of polymorphism and linkage disequilibrium: A total of 24 polymorphic nucleotide sites were found at the Xp21.1 locus trio in a global sample of 42 male individuals. Ten of these polymorphic sites result in only two haplotypes on a pairwise basis and can therefore be defined as congruent ($l_b = 10$; Figure 2b). These 10 congruent sites occur on the internal branches of the haplotype network shown in Figure 3, separating the A haplotype from haplotypes B–Q. Haplotype A is present only in two (of the three sampled) Mbuti Pygmy individuals. At the 5′ end of the Xp21.1 locus trio, the first congruent site occurs at position 430 and, at the 3′ end, the last congruent site occurs at position 16641 (Figure 2b). There is a single observable recombination event between the A haplotype and haplotypes B–Q at the 3′ end of the sequence, between sites 16641 and 17348. There is no such signature of a recombination event at the 5′ end of the sequence. Therefore, $g_d \geq 16.2$ kb. Among the less diverged haplotypes B–Q, there is evidence for a minimum of three recombination events. Previous analysis of LD among human nucleotide poly-
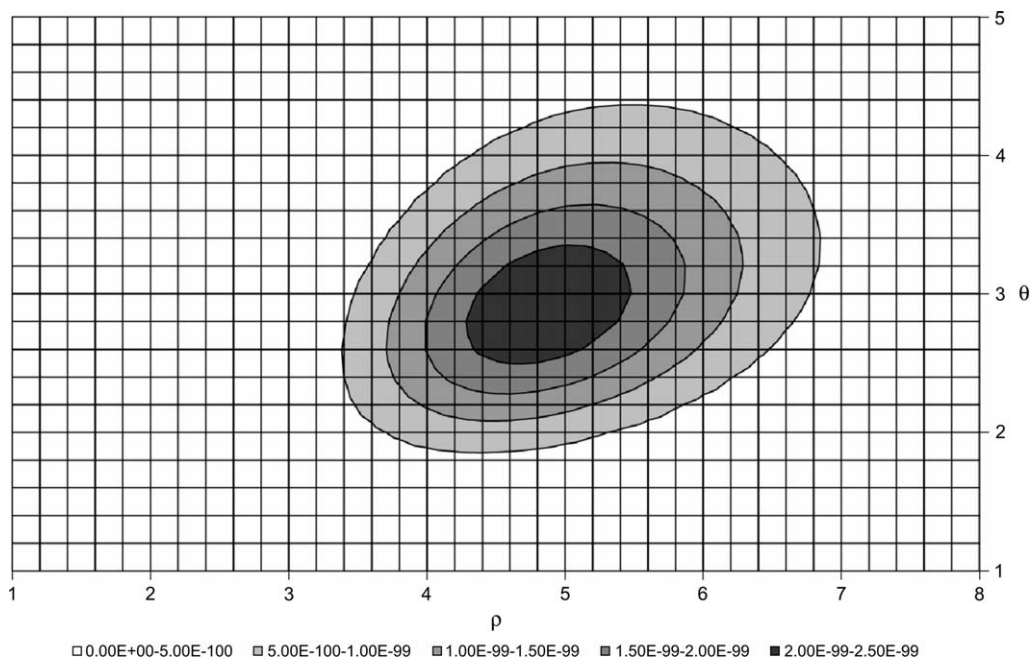
FIGURE 4.—Likelihood surface for the population mutation (θ) and recombination (ρ) parameters estimated from the Xp21.1 data set. The maximum-likelihood point estimate is $\theta = 2.8$ and $\rho = 4.8$. The value of θ is for the 7.5 kb of actual sequence, while the value of ρ is for the entire 17.5-kb region.

□ 0.00E+00-5.00E-100 ▨ 5.00E-100-1.00E-99 ▥ 1.00E-99-1.50E-99 ▦ 1.50E-99-2.00E-99 ■ 2.00E-99-2.50E-99

morphisms demonstrated that the half-length of complete LD typically extends <5 kb in African populations (REICH *et al.* 2001), suggesting that our observation of complete LD between congruent sites separated by at least 16.2 kb in Africa is unusual.

With homologous sequence data from a common chimpanzee outgroup, we estimate that the net proportion of nucleotide substitutions between humans and chimpanzees is $d_o = 0.597\%$ per nucleotide. The net proportion of substitutions between the A lineage and lineages B–Q is $d_1 = 0.189\%$. If we assume that all substitutions and polymorphisms are neutral with respect to natural selection and the mutation rate has remained constant, the ratio of net substitutions, $0.189/0.597 = 0.317$, suggests that the A lineage originated slightly less than one-third of the time since the human-chimpanzee divergence or $\sim$1.899 MYA. This corresponds to a neutral mutation rate of 4.844×10^{-10}, which is much lower than the average value estimated from 15 regions on the X chromosome (1.661×10^{-9} per nucleotide site per year; HAMMER *et al.* 2004). For the human-gorilla comparison, $d_o = 1.161\%$, and the time for the origin of the A lineage becomes 1.302 MYA. This corresponds to a neutral mutation rate of 1.451×10^{-9} per nucleotide per year, which is very close to the mean X chromosome value. We also examined a larger 500-kb window of a human-chimpanzee sequence alignment centered around the Xp21 locus trio and found very similar, low levels of divergence (0.581%; data not shown), consistent with previous findings on the X chromosome (EBERSBERGER *et al.* 2002).

Population recombination and mutation rates were simultaneously estimated with an importance-sampling full maximum-likelihood method. The population recombination rate per nucleotide site ($\rho = 2.743 \times 10^{-4}$) and the population mutation rate per site ($\theta = 3.733 \times$

10^{-4}) were estimated under the assumption of a constant-sized panmictic population (Figure 4). If we divide our estimated value of θ by three times the neutral mutation rate, calculated with the chimpanzee outgroup, we obtain $N_e \approx 28,000$ for the Xp21.1 locus trio. Last, the frequency spectrum of mutations occurring at the Xp21.1 locus trio does not significantly deviate from that expected under mutation-drift equilibrium, as summarized by Tajima's *D*-statistic ($D = -0.378$; $P = 0.431$).

Probability of panmixia: What is the probability that the A haplotype has persisted for >1 million years in a single randomly mating population and still retained complete LD over at least 16.2 kb of sequence? This probability was estimated using Monte Carlo simulation of the coalescent process, producing samples of 42 chromosomes that evolve in a constant-sized panmictic population with mutation and recombination parameters estimated from the data. With this Monte Carlo procedure it is straightforward to obtain the expected values for l_b and g_d. From 10^5 replicate coalescent simulations of panmixia, $E(l_b) = 4.49$ and $E(g_d) = 8.13$ kb. To be conservative, the individual univariate probabilities are $P(l_b \geq 10|\theta, \rho, n) = 0.0248$ and $P(g_d \geq 16.2|\theta, \rho, n) = 0.0159$, where $n = 42$ is the sample size. The probability of jointly observing the two summary statistics is $P(l_b \geq 10, g_d \geq 16.2|\theta, \rho, n) = 0.0011$ (Figure 5). The significance of these results is robust to uncertainty in the Xp21.1 recombination rate. Figure S2 shows that the probability of the Xp21.1 data evolving in a single panmictic population is <0.05 for all nonzero values of ρ.

DISCUSSION

Recent approaches to inferring ancestral demography from human DNA sequence polymorphism have been indirect, based primarily on estimates of the long-
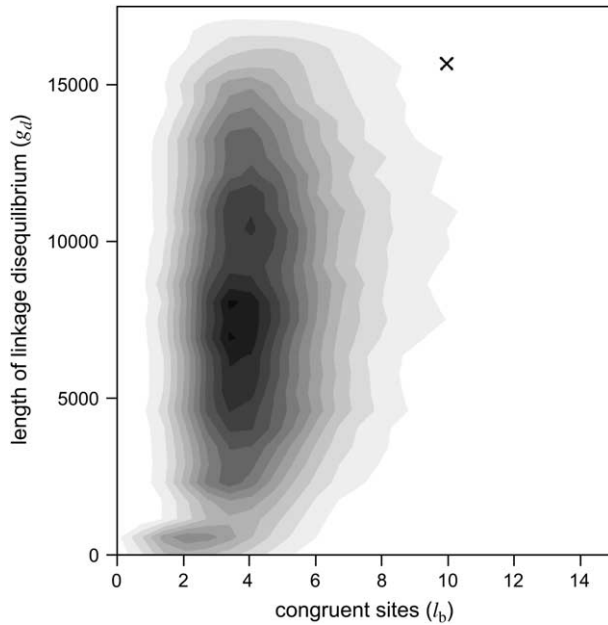
FIGURE 5.—The probability density for the number of congruent sites (l_b) and the maximum physical distance for complete linkage disequilibrium among congruent sites (g_d) expected in a single panmictic population. The shaded area encompasses the estimated 95% confidence interval and the observed Xp21.1 data point is marked with an X.

term N_e and from the shapes of gene trees (CANN *et al.* 1987; HARPENDING *et al.* 1998; HARRIS and HEY 1999; TAKAHATA *et al.* 2001; EXCOFFIER 2002). The design of our study is novel in that it combines genealogical patterns in sequence data with measures of LD to test specific hypotheses about ancient human population structure. Our computer simulations demonstrate that the probability of observing levels of haplotype divergence and LD over the 17.5 kb of sequence from the Xp21.1 locus trio is unacceptably low in a single, randomly mating population. Beyond ancient subdivision, we consider three plausible alternative explanations for the Xp21.1 data: single population growth, a single population bottleneck, and locus-specific natural selection.

The human population has clearly undergone a recent large expansion, yet our coalescent simulations assumed a constant-sized population. With respect to this alternative, our approach can be considered conservative because allowing the population to expand will only diminish the expected proportion of total genealogical time during which only two basal lineages exist (MARJORAM and DONNELLY 1994). Such a reduction in basal branch lengths is expected to result in fewer congruent sites. We performed additional simulations of a single population with 24 polymorphic sites that remains at constant N_e until ~100 KYA, at which time it expands exponentially 100-fold. An additional 10^5 replicates of this computer simulation yield $E(l_b) = 2.47$, much lower than the 4.49 expected under constant size.

While a single population expansion is unlikely to

produce the pattern observed at Xp21.1, a recent reduction from a large ancestral N_e may result in the persistence of divergent lineages. Additional coalescent simulations of a 100-fold population reduction occurring ~100 KYA yield $E(l_b) = 9.70$. However, this alternative model is unlikely to be viable for the human population since the effects of a reduced N_e (in the absence of population structure) are expected to be visible throughout the entire genome, and no consistent evidence for an ancient African bottleneck has been detected in surveys of nucleotide variation at other loci (HARDING *et al.* 1997; HARRIS and HEY 1999; INGMAN *et al.* 2000; WALL and PRZEWORSKI 2000; ZHAO *et al.* 2000; ALONSO and ARMOUR 2001; FRISSE *et al.* 2001; HAMMER *et al.* 2004; HARDING and McVEAN 2004).

The effects of natural selection on patterns of nucleotide variation tend to be localized to small chromosomal regions (*i.e.*, sites linked to those subject to selection) due to the dissociating effect of recombination. Given that Xp21.1 maps to a region of moderately high recombination (~1.9 cM/Mb; http://genome.ucsc.edu) and low gene density (the closest known gene is >500 kb away), there is no *a priori* reason to suspect that this noncoding locus trio is influenced by balancing selection, which is the most likely selective force known to elongate genealogical history (GARRIGAN and HEDRICK 2003). Under long-term balancing selection, a skew toward an excess of intermediate-frequency polymorphisms is expected, which is not observed at the Xp21.1 locus trio (*e.g.*, Tajima's $D = -0.378$). Additionally, TAKAHATA (1990) has shown that when balancing selection is symmetrical, there is no expected change in the proportion of genealogical time occupied by only two lineages; rather, the entire neutral genealogy is elongated. Most importantly, however, neither a population bottleneck nor balancing selection will prevent lineages from recombining with one another, which is a key feature of our inference. For example, under balancing selection, recombination will reduce LD and polymorphism around a selected site (HUDSON and KAPLAN 1988). Therefore, given the inconsistencies arising under alternative hypotheses, we are led to conclude that a model of ancient population subdivision is the most likely explanation for the pattern of Xp21.1 polymorphism.

Previous analyses of human genomic nucleotide diversity have demonstrated that the observed variances in both TMRCA and the mutation frequency spectra are too large to be compatible with simple demographic models, such as single-population growth or bottlenecks (WALL and PRZEWORSKI 2000; PLUZHNIKOV *et al.* 2002; HAMMER *et al.* 2004). This large observed variance in TMRCA throughout the human genome has recently prompted some investigators to propose a metapopulation model for historical human demography (HARDING and McVEAN 2004; WAKELEY 2004). Under such models of historical population structure, WAKELEY (1999) found that the likelihood of the pattern of 44 un-

linked autosomal polymorphisms is maximized when the ancestral human population was more strongly structured than it has been in more recent history. Although we find that levels of divergence and LD at the Xp21.1 locus are consistent with the IAA model of subdivision, the above metapopulation family of models may also be compatible with the data. While both are models of historical subdivision, they differ with respect to the timescale over which gene flow between subpopulations is assumed to occur, a difference that may have important implications for understanding the origin of the AMH phenotype. Forthcoming evidence from additional X-linked and autosomal loci will undoubtedly aid in distinguishing between these two models and enable estimation of parameters, such as the admixture proportion or the rate of gene flow.

An interesting feature of our inference is that the putative isolation and admixture event likely occurred between ancient African subpopulations. The question of hominin admixture has typically focused on events between either AMH and Neanderthals in Europe or AMH and *H. erectus* in Asia. Given recent fossil evidence, Africa may have provided the greatest opportunity for admixture between archaic subpopulations of Homo, simply because Africa harbored the highest levels of hominin taxonomic diversity (Wood 2002; Tattersall 2003). If the IAA model is correct, it implies that subpopulations of archaic Homo existed in allopatry within Africa for much of the Pleistocene. Regardless, the Xp21.1 data reiterate the key role of African hominin diversity in the evolution of our species (Jolly 2001).

If the AMH genome contains *any* degree of dual ancestry (*i.e.*, archaic and modern), the recent African replacement model in its strictest definition (*i.e.*, that of complete replacement) must be rejected. While the majority of the AMH genome may descend from a single African population, if further studies corroborate the inferences made from the Xp21.1 data, it would imply that the evolutionary lineage leading to AMH did not evolve reproductive isolation from other archaic hominin subpopulations and, thus, cannot be considered a distinct biological species. Several models of human origins propose periods of admixture between the lineage leading to AMH and endemic premodern populations (Bräuer 1992; Eswaran 2002). Although, it should be noted that the Xp21.1 data yield little information on whether the admixture may have occurred before or after the emergence of the AMH phenotype. It may be that the AMH phenotype arose first and admixture occurred as a consequence of the rapid expansion of that population. Alternatively, the AMH phenotype may, itself, be the by-product of such admixture events.

In conclusion, it is unlikely that the deep divergence between Xp21.1 haplotypes and the complete LD between polymorphisms defining these haplotypes arose in a single population. This inference is robust to uncertainty in the estimated recombination rate for the Xp21.1

locus trio. If one accepts the argument that balancing selection is unlikely to have generated such patterns, some form of prolonged population subdivision must be invoked to explain the data. Our experimental design and hypothesis-testing framework are based upon the isolation-and-admixture model of Nordborg (2001) and Wall (2000), and we find that the Xp21.1 polymorphism data do indeed conform to the predictions of this model. However, additional models of sustained population subdivision (such as the metapopulation models) also remain feasible; further theoretical predictions based upon these models must be obtained before the specific nature of the putative historical subdivision can be fully addressed. While a single locus can cast doubt on the recent African replacement model, full resolution of the question of interbreeding awaits future data collected from additional loci. One speculative question that would be particularly relevant to human biology is whether the assimilated archaic genetic material is limited to noncoding regions, such as Xp21.1, that introgressed into the AMH population by random genetic drift or whether functional regions may have also introgressed due to natural selection.

LITERATURE CITED

Alonso, S., and J. A. Armour, 2001 A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. Proc. Natl. Acad. Sci. USA **98:** 864–869.

Bräuer, G., 1992 Africa's place in the evolution of *Homo sapiens*, pp. 83–98 in *Continuity or Replacement? Controversies in Homo sapiens Evolution*, edited by G. Bräuer and F. H. Smith. Balkema, Rotterdam, The Netherlands.

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

Ebersberger, I., D. Metzler, C. Schwarz and S. Pääbo, 2002 Genomewide comparisons of DNA sequences between humans and chimpanzees. Am. J. Hum. Genet. **70:** 1490–1497.

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

Excoffier, L., 2002 Human demographic history: refining the recent African origin model. Curr. Opin. Genet. Dev. **12:** 675–682.

Fearnhead, P., and P. Donnelly, 2001 Estimating recombination rates from population genetic data. Genetics **159:** 1299–1318.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack *et al.*, 2001 Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. **69:** 693–709.

Garrigan, D., and P. W. Hedrick, 2003 Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution **57:** 1707–1722.

Garrigan, D., Z. Mobasher, T. Severson, J. A. Wilder and M. F. Hammer, 2005 Evidence for archaic Asian ancestry on the human X chromosome. Mol. Biol. Evol. **22:** 189–192.

Hammer, M. F., T. Karafet, A. Rasanayagam, E. T. Wood, T. K. Altheide *et al.*, 1998 Out of Africa and back again: nested

cladistic analysis of human Y chromosome variation. Mol. Biol. Evol. **15:** 427–441.

HAMMER, M. F., D. GARRIGAN, E. WOOD, J. A. WILDER, Z. MOBASHER *et al.*, 2004 Heterogeneous patterns of variation among multiple human X-linked loci: the possible role of diversity-reducing selection in non-Africans. Genetics **167:** 1841–1853.

HARDING, R. M., 1999 More on the X files. Proc. Natl. Acad. Sci. USA **96:** 2582–2584.

HARDING, R. M., and G. McVEAN, 2004 A structured ancestral population for the evolution of modern humans. Curr. Opin. Genet. Dev. **14:** 667–674.

HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX *et al.*, 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60:** 772–789.

HARPENDING, H. C., M. A. BATZER, M. GURVEN, L. B. JORDE, A. R. ROGERS *et al.*, 1998 Genetic traces of ancient demography. Proc. Natl. Acad. Sci. USA **95:** 1961–1967.

HARRIS, E. E., and J. HEY, 1999 X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **96:** 3320–3324.

HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18:** 337–338.

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

INGMAN, M., H. KAESSMANN, S. PÄÄBO and U. GYLLENSTEN, 2000 Mitochondrial genome variation and the origin of modern humans. Nature **408:** 708–713.

JOLLY, C. J., 2001 A proper study for mankind: analogies from the Papionin monkeys and their implications for human evolution. Yearb. Phys. Anthropol. **44:** 177–204.

KLEIN, R. G., 1999 *The Human Career: Human Biological and Cultural Origins.* University of Chicago Press, Chicago.

MARJORAM, P., and P. DONNELLY, 1994 Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics **136:** 673–683.

McBREARTY, S., and A. S. BROOKS, 2000 The revolution that wasn't: a new interpretation of the origin of modern human behavior. J. Hum. Evol. **39:** 453–463.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NORDBORG, M., 2001 On detecting ancient admixture, pp. 123–136 in *Genes, Fossils and Behavior: An Integrated Approach to Human Evolution.* IOS Press, Amsterdam.

PLUZHNIKOV, A., A. DI RIENZO and R. R. HUDSON, 2002 Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics **161:** 1209–1218.

REICH, D. E., E. M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

STEFANSSON, H., A. HELGASON, G. THORLIEIFSSON, V. STEINTHORSDOTTIR, G. MASSON *et al.*, 2005 A common inversion under selection in Europeans. Nat. Genet. **37:** 129–137.

STRINGER, C. B., and P. ANDREWS, 1988 Genetic and fossil evidence for the origin of modern humans. Science **239:** 1263–1268.

TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87:** 2419–2423.

TAKAHATA, N., S. H. LEE and Y. SATTA, 2001 Testing multiregionality of modern human origins. Mol. Biol. Evol. **18:** 172–183.

TATTERSALL, I., 2003 Once we were not alone. Sci. Am. (Summer): 20–27.

WAKELEY, J., 1999 Nonequilibrium migration in human history. Genetics **153:** 1863–1871.

WAKELEY, J., 2004 Metapopulation models for historical inference. Mol. Ecol. **13:** 865–875.

WALL, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

WALL, J. D., 2000 Detecting ancient admixture in humans using sequence polymorphism data. Genetics **154:** 1271–1279.

WALL, J. D., and M. PRZEWORSKI, 2000 When did the human population size start increasing? Genetics **155:** 1865–1874.

WHITE, T. D., B. ASFAW, D. DEGUSTA, H. GILBERT, G. D. RICHARDS *et al.*, 2003 Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature **423:** 742–747.

WILDER, J. A., Z. MOBASHER and M. F. HAMMER, 2004 Genetic evidence for unequal effective population sizes of human females and males. Mol. Biol. Evol. **21:** 2047–2057.

WOLPOFF, M. H., 1999 *Paleoanthropology.* McGraw-Hill, Boston.

WOOD, B., 2002 Hominid revelations from Chad. Nature **418:** 133–135.

Y CHROMOSOME CONSORTIUM, 2002 A nomenclature system for the tree of Y chromosomal binary haplogroups. Genome Res. **12:** 339–348.

ZHAO, Z., L. JIN, Y. X. FU, M. RAMSAY, T. JENKINS *et al.*, 2000 Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. USA **97:** 11354–11358.

ZIETKIEWICZ, E., V. YOTOVA, D. GEHL, T. WAMBACH, I. ARRIETA *et al.*, 2003 Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am. J. Hum. Genet. **73:** 994–1015.