MOLECULAR BIOLOGY AND EVOLUTION

LETTER

Evidence for Archaic Asian Ancestry on the Human X Chromosome

Daniel Garrigan, Zahra Mobasher,* Tesa Severson,* Jason A. Wilder,* and Michael F. Hammer*†*

*Division of Biotechnology, and †Department of EEB, University of Arizona, Tucson

The human *RRM2P4* pseudogene has a pattern of nucleotide polymorphism that is unlike any locus published to date. A gene tree constructed from a 2.4-kb fragment of the *RRM2P4* locus sequenced in a sample of 41 worldwide humans clearly roots in East Asia and has a most-recent common ancestor approximately 2 Myr before present. The presence of this basal lineage exclusively in Asia results in higher nucleotide diversity among non-Africans than among Africans. A global survey of a single-nucleotide polymorphism that is diagnostic for the basal, Asian lineage in 570 individuals shows that it occurs at frequencies up to 53% in south China, whereas only one of 177 surveyed Africans carries this archaic lineage. We suggest that this ancient lineage is a remnant of introgressive hybridization between expanding anatomically modern humans emerging from Africa and archaic populations in Eurasia.

Introduction

Recently, Hammer et al. (2004) analyzed global human nucleotide variation at 15 X-linked loci, one of which stands out as unique in its pattern of polymorphism. In a global sample of 41 individuals, sequence variation at the ribonucleotide reductase M2 subunit pseudogene 4 (*RRM2P4*) is partitioned into two divergent, basal lineages. Both of these lineages are found in Asia, whereas only one is found in sub-Saharan Africa (fig. 1*B*). The two lineages differ by five fixed mutations, leading to an estimate for the time to a most-recent common ancestor (TMRCA) that is approximately 2 Myr before present (B.P.). Such long basal branches of the global *RRM2P4* genealogy are not expected under a neutral panmictic model (Wall 1999, 2000). Moreover, a gene tree that roots in Asia has not been previously observed in published human nucleotide polymorphism data sets (e.g., Takahata, Lee, and Satta 2001) and raises intriguing questions concerning the origin and expansion of anatomically modern humans. In this paper, we explore the possibility that polymorphism at *RRM2P4* recovers a history of admixture between *Homo sapiens* and an archaic Asian form of *Homo* before the latter went extinct in the late Middle Pleistocene (Swisher et al. 1996).

Methods

The methodology and samples used to assay nucleotide polymorphism at the *RRM2P4* locus has been previously reported by Hammer et al. (2004). Summary statistics were calculated with the computer application DnaSP version 3.99 (Rozas et al. 2003). The TMRCA was estimated by dividing the observed net pairwise nucleotide differences between lineages (D_A [Nei 1987]) by twice the mutation rate. We also used a maximum-likelihood approach to estimate the TMRCA with the computer program GENETREE version 9.0 (Griffiths 2002). Using the null demographic model of panmixia and constant population size, we estimated the parameter, θ_{ml} ($= 3N_e\mu$, where N_e is the effective population size and μ is the mutation rate), by generating a single-likelihood curve covering a wide range of possible values. Based on this value of θ_{ml}, we then estimated the N_e and TMRCA. The G/A single-nucleotide polymorphism at position 2020 was genotyped in a sample of 570 Africans and non-Africans with the TaqMan genotyping assay (Applied Biosystems, Foster City, Calif.), following the manufacturer's protocol. The sample included 22 Khoisan from Namibia, 39 Dogon from Mali, 47 Bantu from South Africa, 25 Bakola from Cameroon, 44 Dinka from Sudan, 46 Mongolians, 40 Sri Lankans, 11 Tibetans, 48 Baining from New Britain, 26 Japanese, 28 Han from northern China, 30 Yao from southern China, 28 Altaians, 24 Papua New Guineans, 45 Italians from central Italy, 39 Dutch, and 27 Iranians. All sampling protocols were carried out with the approval of the University of Arizona Human Subjects Committee.

Results and Discussion

Figure 1*A* depicts the 13 polymorphic sites we observe in a total of 2,385 bp of *RRM2P4* sequence, and table 1 lists summary statistics for the locus. Although levels of nucleotide diversity are only slightly higher than the genomic average (Yu et al. 2002), *RRM2P4* is one of the few surveyed loci that exhibits more variation in non-Africans than in Africans. It is worth noting that this RRM2 pseudogene is located in a region with a high rate of crossing-over (~3.6 cM/Mb) and a low gene density on Xq27.3 (Hammer et al. 2004). Despite its genomic context, our 2.4-kb fragment shows no direct evidence of historical recombination by the four-gamete test of Hudson and Kaplan (1985). This result is not unexpected, because complete linkage disequilibrium is known to extend beyond 2.4 kb in both African and non-African populations (Reich et al. 2001). However, this lack of recombination does permit reconstruction of a single, nonreticulating gene tree (fig. 1*B*). The genomic context of *RRM2P4* makes it less likely that the elongated basal branches are the result of linkage to a functional site(s) subject to some form of balancing selection. Additionally, there are more restricted conditions for a X-linked balanced polymorphism compared with the

Key words: hybridization, human origins, polymorphism, population subdivision, pseudogene.

E-mail: mfh@u.arizona.edu.

Mol. Biol. Evol. 22(2):189–192. 2005
doi:10.1093/molbev/msi013
Advance Access publication October 13, 2004

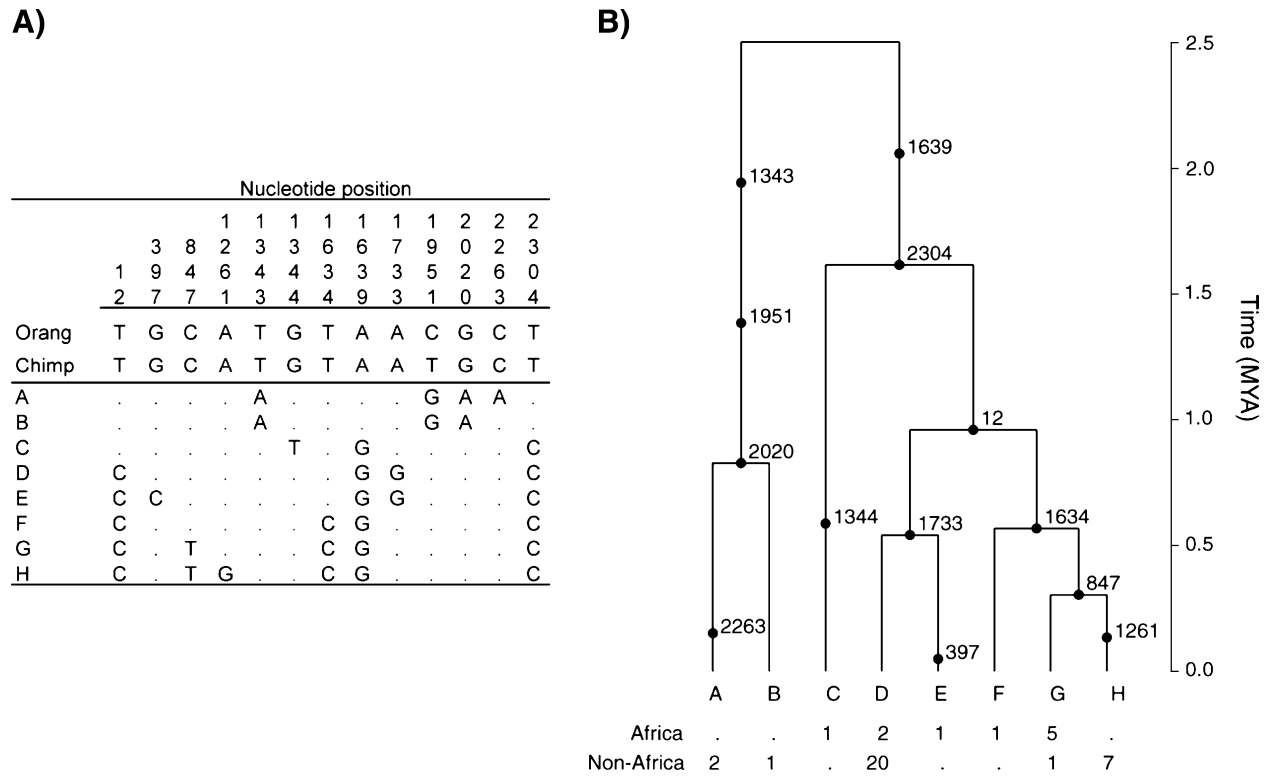

FIG. 1.—(*A*) Sequences and (*B*) gene tree for the eight *RRM2P4* haplotypes. Scale bar denotes time in units of millions of years before present. Mutations are labeled with a dot along the edges of the genealogy and are placed according to the maximum-likelihood estimate of their respective ages. Below the tips of the gene tree are the multiplicities of each haplotype in both the African and the non-African samples.

autosomes (Hedrick and Parker 1997). Indeed, the frequency spectrum of mutations at *RRM2P4* does not reject the null hypothesis of neutral mutation-drift equilibrium (Tajima's D statistic $= -0.206$, $P = 0.489$) (table 1).

Outgroup sequencing reveals that there are 18 nucleotide substitutions between human *RRM2P4* sequences and one common chimpanzee (*Pan troglodytes*) sequence and 62 substitutions between human sequences and one orangutan (*Pongo pygmaeus*) sequence. The human-chimpanzee comparison yields a neutral mutation rate of 7.4×10^{-10} substitutions per site per year (assuming a 6 Myr human-chimpanzee divergence time). This rate is only slightly slower than the mean 15-locus rate calculated with the data of Hammer et al. (2004) of 8.4×10^{-10} per site per year. Utilizing sequence from all three species, the relative rate test of Tajima (1993) does not reject a constant rate of nucleotide substitution at the *RRM2P4* locus ($\chi^2 = 0.39$, $P = 0.532$). If we combine the neutral mutation rate given above with an average of 8.25 nucleotide differences observed between the two human *RRM2P4* lineages, then we can estimate that the two lineages diverged approximately 1.96 MYA. This estimate is in reasonable agreement with a TMRCA of 2.44 Myr B.P. obtained from the coalescent-based maximum-likelihood method of Griffiths (2002), which assumes a panmictic population of constant size (fig.1*B*). This places the most-recent common ancestor of *RRM2P4* in the late Pliocene, a time when the genus *Homo* first appears in the fossil record (Wood and Collard 1999). The maximum-likelihood method also estimates that the effective population size of *RRM2P4* is approximately 25,000. Two other surveyed X-linked loci (*DMD44* and *PDHA1*) yield comparably large TMRCAs and effective population sizes (Hammer et al. 2004; Harris and Hey 1999) and may also provide evidence of ancient population structure (Harding 1999).

Table 1
Estimates of Population Genetic Parameters for the *RRM2P4* Locus

Population	n	S	θ_W(%)	π (%)	D
Global	41	13	0.127	0.119	−0.206
Africans	10	6	0.089	0.090	0.068
Non-Africans	31	11	0.115	0.124	0.224

NOTE.—n is the number of individuals sampled; S is the number of polymorphic sites; θ_W is Watterson's (1975) estimate of 3 Nm per nucleotide site (N is the effective population size and m is the neutral mutation rate); π is the average nucleotide heterozygosity; D–Tajima's (1989) neutrality statistic.

The Asian *RRM2P4* lineage is found exclusively in three of 31 sampled non-Africans (a Japanese, a Chinese, and a Melanesian individual), whereas the other, more diverse, lineage constitutes the entirety of African and the majority of non-African variation. To obtain a better estimate the frequency of the Asian lineage and to test for its presence in Africa, we genotyped the SNP occurring at position 2020 (fig. 1*A*; also see *Methods*) in a total sample of 570 African and non-African individuals representing 17 populations. Figure 2 shows the observed geographical distribution of the Asian lineage. There appears to be a decreasing frequency gradient centered on southern China, where the Asian lineage reaches a maximum of 53% in the

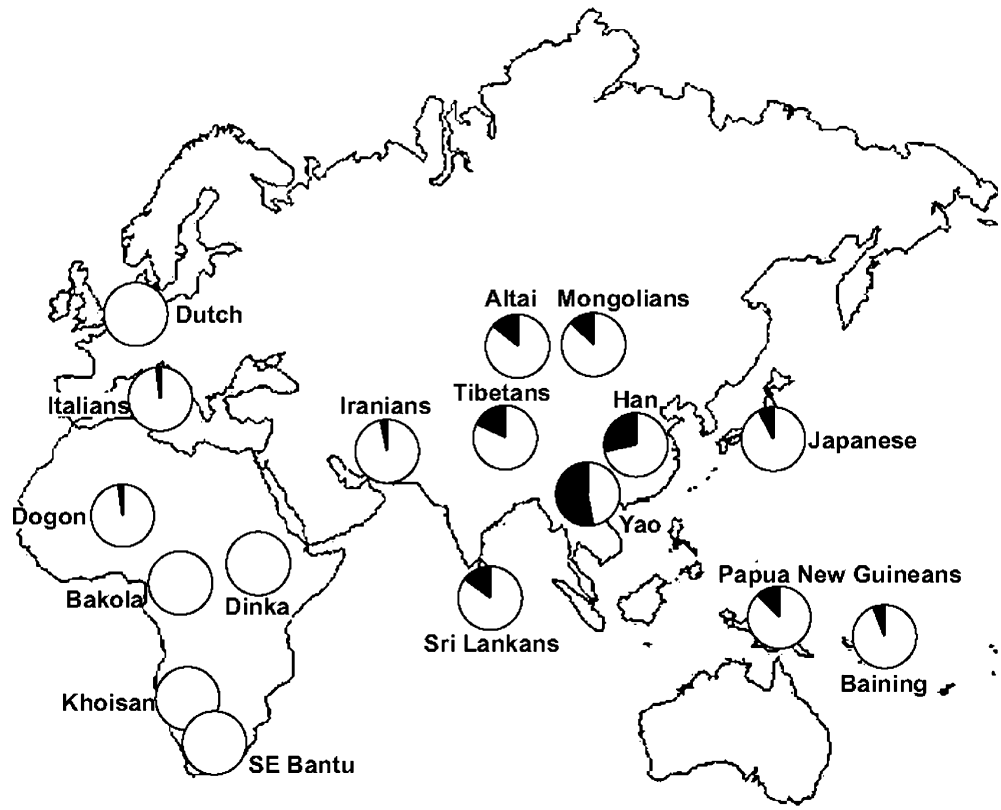

FIG. 2.—Geographical distribution of the archaic *RRM2P4* lineage in 17 population samples. Frequency data can be found in Supplementary Material online.

Yao of southern China and 29% in the Han Chinese. The Asian lineage is present in only one of 177 sampled Africans (0.6%). This geographical pattern is consistent with a Southeast Asian origin of this 2-Myr-old lineage.

If we accept that the mutations occurring at the *RRM2P4* locus are selectively neutral (and, therefore, influenced only by demographic history), then there are two alternative explanations for its unusual pattern of polymorphism. The first alternative is that the long basal branches are simply the chance result of genetic drift in a large, panmictic ancestral population. It is possible that the divergent, basal lineage was carried out of Africa and subsequently lost in Africa and/or increased in frequency in Asia by genetic drift. Second, it may be the result of recent admixture between two divergent populations; that is, the expanding anatomically modern human population and *Homo erectus*. Distinguishing between these two alternatives proves difficult, given the low power afforded by any single locus (Nordborg 2000; Wall 2000). To achieve greater single-locus power, it is preferable to examine patterns of polymorphism over physical distances approximately 20 to 30 kb in length (Wall 2000).

Conclusion

Polymorphism occurring at the *RRM2P4* locus is unique in that it clearly roots in East Asia, has an ancient TMRCA, and also yields higher non-African than African nucleotide diversity. Our SNP assay estimates that the Asian lineage is found at less than 1% in African populations. The distribution of the Asian lineage strongly suggests an Asian origin but should not be taken as definitive proof that it did not originate in Africa. The Asian lineage appears to have diverged from the globally distributed portion of the genealogy approximately 2MYA. It is interesting to note that this estimated divergence time is concordant with the age of the oldest *Homo erectus* fossils found outside of Africa (Gabunia et al. 2000). Following further lines of investigation, if panmixia at the *RRM2P4* locus can be rejected, it would have important implications for our view of *Homo sapiens* as a species. Any degree of dual ancestry in the modern human genome would either demonstrate that the transition to an anatomically modern form did not occur in an isolated, panmictic population (Cann, Stoneking, and Wilson 1987) or that replacement of preexisting hominid populations was incomplete (e.g., Brauer 1989; Smith, Falsetti, and Donnelly 1989).

Acknowledgments

We thank J. Wall for his fruitful collaboration and R. Sprissler for assistance with SNP genotyping. We also thank Tanya Karafet, Dashtseveg Tumen, Giovanni Destro-Bisol, Himla Soodyall, Andrea Novelleto, Peter de Knijff, Beverly Strassmann, William Wang, Jonathan Friedlaender, and Du Ruofu for providing DNA samples. Publication of this work was made possible by grant GM-53566 from the National Institute of General Medical

Sciences (to M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of the National Institutes of Health.

Literature Cited

Brauer, G. 1989. The evolution of modern humans: a comparison of the African and non-African evidence. Pp. 123–154 *in* P. Mellars, and C. Stringer, eds. The human revolution: behavioural and biological perspectives on the origins of modern humans. Edinburgh University Press, Edinburgh.

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

Gabunia, L., A. Vekua, D. Lordkipanidze et al. (14 co-authors). 2000. Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. Science **288**:1019–1025.

Griffiths, R. C. 2002. Ancestral inference from gene trees. Pp. 94–117 *in* M. Slatkin, ed. Modern developments in theoretical population genetics: the legacy of Gustave Malecot. Oxford University Press, New York.

Hammer, M. F., D. Garrigan, E. Wood, J. A. Wilder, Z. Mobasher, A. Bigham, J. G. Krenz, and M. W. Nachman. 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.

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

Hedrick, P. W., and J. D. Parker. 1997. Evolutionary genetics and genetic variation of haplodiploids and X-linked genes. Ann. Rev. Ecol. Syst. **28**:55–83.

Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–164.

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

Nordborg, M. 2000. On detecting ancient admixture. Genes, fossils, and behaviour: an integrated approach to human evolution. IOS Press, Amsterdam.

Reich, D. E., M. Cargill, S. Bolk et al. (11 co-authors). 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**:2496–2597.

Smith, F. H., A. B. Falsetti, and S. M. Donnelly. 1989. Modern human origins. Yrbk. Phys. Anthropol. **32**:35–68.

Swisher, C. C., W. J. Rink, S. C. Anton, H. P. Schwarcz, G. H. Curtis, A. Suprijo, and Widiasmoro. 1996. Latest Homo erectus of Java: potential contemporaneity with *Homo sapiens* in southeast Asia. Science **274**:1870–1874.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

———. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–607.

Takahata, N., S. H. Lee, and Y. Satta. 2001. Testing multiregionality of modern human origins. Mol. Biol. Evol. **18**: 172–183.

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

———. 2000. Detecting ancient admixture in humans using sequence polymorphism data. Genetics **154**:1271–1299.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. **7**: 256–276.

Wood, B., and M. Collard. 1999. The human genus. Science **284**:65–71.

Yu, N., F. C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S. K. Shyue, and W. H. Li. 2002. Larger genetic differences within Africans than between Africans and Eurasians. Genetics **161**:269–274.

Scott Edwards, Associate Editor

Accepted October 4, 2004