Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1805–1809, March 1994
Evolution

Molecular and population genetic analysis of allelic sequence diversity at the human β-globin locus

(Melanesia/allele-specific amplification/recombination/gene conversion/evolution)

S. M. FULLERTON*†, R. M. HARDING*, A. J. BOYCE‡, AND J. B. CLEGG*

*Medical Research Council Molecular Haematology Unit, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, United Kingdom; and ‡Institute of Biological Anthropology, University of Oxford, 58 Banbury Road, Oxford, OX2 6QS, United Kingdom

Communicated by David Weatherall, October 25, 1993 (received for review September 24, 1993)

ABSTRACT **Allelic sequence polymorphism at the β-globin locus was investigated in a group of 36 Melanesians. A 3-kilobase fragment containing the gene and its flanking regions was sequenced in 60 normal (β^A) and 12 thalassemic (intron 1, position 5, G → C) chromosomes. Haplotype relationships between linked polymorphisms were derived by allele-specific PCR amplification and sequencing. Seventeen nucleotide polymorphisms and 2 length variants were identified, and these sites segregated as 17 sequence haplotypes in the normal chromosomes. This haplotype diversity is higher than that expected on the basis of the nucleotide polymorphism observed and is probably due to recombination and gene conversion. Nucleotide diversity at synonymous sites in the sample is 0.14%, suggesting an average age of sequence divergence of ≈450,000 years, consistent with that expected for a neutrally evolving human nuclear locus.**

Molecular population genetic analysis of nuclear loci was pioneered by Kreitman in a study of the alcohol dehydrogenase (*Adh*) locus in *Drosophila melanogaster* (1). That survey of 11 alleles revealed extensive nucleotide polymorphism, which has since been shown to reflect the effects of both recombination (2) and balancing selection (3) on the *Adh* gene. Although allelic sequencing is the ideal approach to investigating both molecular mechanisms and population genetic questions, the stratagem of using isochromosomal species lines to simplify analysis of diploid loci is not applicable to humans. Studies of human DNA polymorphism have therefore been mostly limited to sequence analysis of haploid mitochondrial DNA (mtDNA) and to scoring of restriction fragment length polymorphisms (RFLPs) at nuclear loci (4). Allelic sequence analysis of nuclear genes has remained, largely for technical reasons, an elusive aim.

The polymerase chain reaction (PCR) (5) and accompanying innovations have made rapid sequencing of genomic DNA commonplace, but because PCR simultaneously amplifies both alleles of diploid loci, linkage relationships in compound heterozygotes are obscured. In most instances, amplified DNA must be subcloned, a time-consuming procedure with the potential for introducing sequence errors that are unacceptable in population-genetic analyses. We have used allele-specific PCR amplification (6) to circumvent these difficulties and extend the analysis of human nuclear genes by determining allelic sequence variation in a 3-kilobase (kb) region surrounding the human β-globin locus.

The β-globin locus has been one of the most intensively studied of all human loci, not least because of its association with severe inherited hemoglobinopathies such as sickle cell anemia and β-thalassemia, some of the commonest genetic diseases in the world. The earliest examples of prenatal diagnosis of these disorders using DNA analysis relied on linkage disequilibrium of β-globin gene mutations with adjacent RFLPs, over 20 of which have now been identified (7). A comprehensive set of β-globin gene cluster RFLP haplotypes has been amassed for many populations (8), and it was the availability of these data that first prompted analysis of population relationships using DNA markers (9). This well-characterized locus is thus an ideal subject for detailed analysis of allelic sequence variation in human populations.

Here, we report the results of our initial survey of nucleotide polymorphism at the β-globin locus§ in a group of Melanesians from northern Vanuatu, an island archipelago in the southwest Pacific Ocean.

MATERIALS AND METHODS

Population Sampled. Genomic DNAs, extracted from the umbilical cord bloods (10) of 36 unrelated Melanesians (24 normal individuals born on the island of Espiritu Santo in northern Vanuatu and 12 β-thalassemia heterozygotes from a neighboring island, Maewo), were used.

Amplification and Sequencing of Diploid Template. A 3.067-kb fragment containing the human β-globin gene and 1.4 kb of flanking sequence was PCR-amplified and directly sequenced (11). Genomic template (400 ng) was subjected to 35 cycles of amplification with primers corresponding to positions 10547–10572 and 13592–13613 of the human β-globin gene reference sequence (12). Ten picomoles of each primer (3′ primer biotinylated), 50 mM KCl, 10 mM Tris chloride (pH 8.4), 1.5 mM $MgCl_2$, 0.001% gelatin, 200 mM dNTPs, 1.85 μg of gene *32* protein (Pharmacia) (13), and 2 units of *Taq* polymerase were included in each 100-μl reaction. Cycle conditions were 1 min at 94°C, 2 min at 60°C, and 5 min at 72°C.

Four reactions per sample were combined, and single-stranded template was prepared by using streptavidin-coated magnetic beads (Dynal, Great Neck, NY) for capture and purification of the biotinylated strand. This template was sequenced with Sequenase T7 DNA polymerase (United States Biochemical) with eight internal sequencing primers (18–24 bases long) spaced at ≈350-base-pair (bp) intervals. Identical sequences from different individuals were run out in parallel, with corresponding dideoxy-NTP (ddNTP) termination reactions loaded in adjacent wells (Fig. 1).

Allele-Specific Amplification for Linkage Determination. To determine linkage relationships in compound heterozygotes, the amplification refractory mutation system (ARMS) (6) was used. An ARMS primer, incorporating an additional internal (position −3) mismatch (14), was used to permit *Taq* polymerase extension from only one of the two bases present at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RFLP, restriction fragment length polymorphism; ARMS, amplification refractory mutation system.

†To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L26462–L26478).

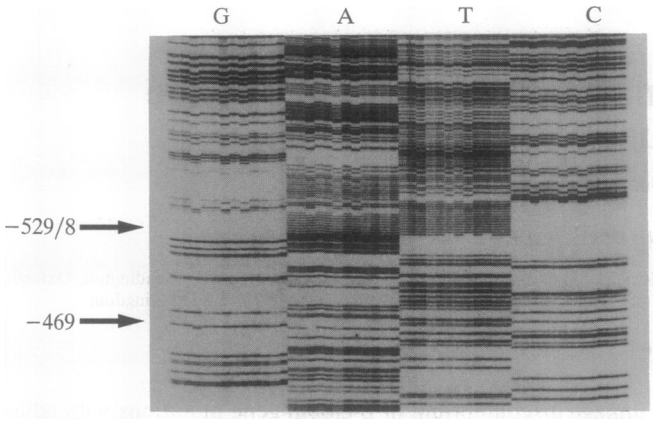


FIG. 1. DNA sequence of 12 allele-specific templates amplified by the ARMS (6) method. Termination reactions are loaded in parallel and from left to right are ddGTP (G) 1–12, ddATP (A) 1–12, ddTTP (T) 1–12, and ddCTP (C) 1–12. Nucleotides −636 to −428 of the β-globin gene cap site, 3′ → 5′ strand, are shown. The AT repeat region polymorphism (−529 and −528) is indicated, as is a downstream A → G difference in sample 3 at position −469. In heterozygous samples, the region upstream of the purine–pyrimidine repeat is uninterpretable because of length differences associated with the insertion polymorphism.

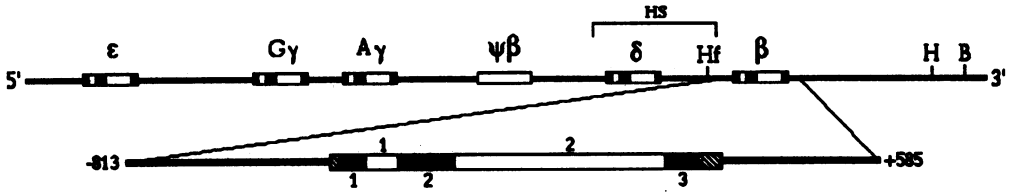
a heterozygous site. Only the sequence that lies between the preferred base and the reverse primer is amplified, resulting in an allele-specific PCR product. Sites of high heterozygosity, with proximity to either the 5′ or 3′ end of the 3.0-kb fragment (i.e., positions −703, +183, +339, +492 in Fig. 2*B*), were used in the ARMS amplifications. Allele-specific products ranging in size from 2.6 to 2.9 kb were routinely amplified to high yield and specificity, giving unambiguous DNA sequence (Fig. 1).

Genomic DNAs were amplified as described above by using the appropriate ARMS primer and complementary biotinylated 5′ or 3′ primer (annealing temperature was adjusted as required for specificity). Template preparation and sequencing was as described for diploid amplifications. With the sequence of one allele directly determined, most of the sequence of the corresponding allele was inferred by reference to the original diploid sequence. The second allele was specifically amplified and sequenced only to resolve the length polymorphisms at positions −521 and −529. In each case there were no discrepancies between the haploid and diploid sequences obtained from a given individual.

RFLP Analysis of Adjacent Sites. Amplified genomic DNA was digested with restriction enzymes to score RFLP sites outside the sequenced region. PCR primers for the *Hin*fI site 5′ to the β-globin gene were as described (17). Primers used to type the 3′ *Hin*dIII and *Bam*HI sites will be described elsewhere (S.M.F. and J.B.C., unpublished data).

Test for Population Subdivision. To evaluate subdivision, we used the single subpopulation test of Strobeck (18). This test compares two estimates of diversity, $\theta = 4N\mu$, where N is the population size and μ is the mutation rate. When assuming the infinite "alleles" model, θ is estimated from the

A



B

Ref	Haplotype Frequency Santo	Maewo	-909 Hinf I	-703 T	-551 T Rsa I	-541 T	-529 -	-528 T	-521 C	-469 T	-340 T	-316 T	Exon1 6 C	Intron 2 16 C Ava II	74 G	81 C	666 T	101 G	181 G	183 C	339 A Hinf I	492 G	~7000 Hind III	~8000 Bam HI
A1	10	2	+/-	C	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+/-	+
A2	5	3	+/-	C	C	•	AT	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	-,+	+/-,+
A3	1		-	C	C	•	AT	A	•	•	•	C	•	•	•	•	•	•	•	•	•	•	-	+
A4	2	2	+	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+/-	+
B1	9	2	+/-	C	C	•	AT	A	•	•	•	•	•	•	T	•	•	•	•	•	T	T	+	-
B2	2		+/-	•	C	•	AT	A	•	•	•	•	•	•	T	•	•	•	•	•	T	T	+	-
B3	1	1	+	C	C	•	AT	•	•	•	•	•	•	•	T	•	•	•	•	•	T	T	+	-
B4	1		+	C	C	•	•	•	•	•	•	•	•	•	T	•	•	•	•	•	T	T	+	-
B5	2		+	•	•	•	•	A	•	C	•	•	•	•	T	•	•	•	•	•	T	T	+	+/-
B6	1		+	•	•	•	AT	A	-(2)	•	C	•	•	•	T	•	•	•	•	•	T	T	+	-
C1	6		+/-	•	•	•	•	A	•	C	C	•	T	G	T	•	C	•	•	A	•	T	-	+
C2	1		+	•	•	C	•	A	•	C	C	•	T	G	T	•	C	•	•	A	•	T	-	+
C3	1		+	•	•	•	•	A	-(1)	•	C	•	T	G	T	•	C	•	A	A	•	T	-	+
C4	2	1	+/-	C	C	•	AT	A	•	•	C	•	T	•	T	•	C	•	•	A	•	T	-	+
D1	2	1	+/-	•	•	•	•	•	•	•	C	•	T	G	T	T	C	C	•	A	•	T	-	+
D2	1		+	•	•	•	•	•	•	•	C	•	•	G	T	T	C	C	•	A	•	T	-	+
D3	1		-	C	•	•	•	•	•	•	C	•	T	G	T	T	C	C	•	A	•	T	-	+

FIG. 2. (*A*) The 3-kb sequenced region and its relationship to the human β-globin gene cluster. The top line shows the β-globin gene cluster. RFLP sites are Hf, *Hin*fI; H, *Hin*dIII, and B, *Bam*HI. HS indicates the approximate boundaries of a previously identified 9.1-kb hotspot for recombination (32). Below the cluster, in expanded form, is the 3-kb genomic region that was sequenced. □, Introns (numbered above the boxes); ■, exons (numbered below the boxes); and ▨, untranslated regions of the β-globin gene. (*B*) Sequence variation observed in the two island samples of normal β^A chromosomes from northern Vanuatu. Variant positions are numbered at the 5′ end in relation to the β-globin gene cap site and at the 3′ end in relation to its polyadenylylation site, while polymorphisms within the gene are labeled according to functional region (i.e., intron 2: 81 is the 81st nucleotide of the second intron of the gene). Sequence haplotypes are arranged in the order of sequence similarity and labeled A, B, C, or D (see text). Stars indicate identity with the β-globin reference sequence (12) as defined on the first line, and dashes indicate site absence. At −521, (1) and (2) refer to the number of nucleotides deleted in the sequence. The RFLP sites flanking the sequenced region are shown, as are recognized RFLPs associated with observed nucleotide differences.

observed number of alleles or sequence haplotypes. When assuming the infinite "sites" model, θ is also estimated from the number of polymorphic sites (s_n) observed in a sample of n DNA sequences:

$$\theta = s_n \Big/ \sum_{i}^{n-1} (1/i).$$

This estimate of θ is used to calculate the expected number of sequence haplotypes, k:

$$k = 1 + \theta/(\theta + 1) + \ldots \theta/[\theta + (n - 1)].$$

Whereas the observed number of haplotypes should not differ significantly from the expected number when there is random mating and no intragenic recombination, population subdivision leads to fewer than the expected number of haplotypes being observed.

RESULTS

Nucleotide Polymorphism in Northern Vanuatu. We examined nucleotide polymorphism at the β-globin locus in 60 normal (β^A) chromosomes and 12 thalassemic (intron 1, position 5, G → C) chromosomes from 36 unrelated Melanesians. Seventeen dimorphic nucleotide positions were identified within the 3 kb sequenced (Fig. 2): 7 in the 5′ flanking region (814 bp total length), 1 (synonymous substitution) in the exons of the gene (444 bp), 4 in its intron sequences (980 bp), and 5 in the 3′ flanking DNA (582 bp). No polymorphisms were found in either the 5′ or 3′ untranslated regions of the gene. Ten of the 17 base changes observed were transitions, only 1 of which results from mutation at a CpG dinucleotide site (position +181). Two length polymorphisms were also observed in the sequenced region: a 2-bp AT insertion at −529 and either a 1- or 2-bp deletion at position −521. The AT dinucleotide insert accompanies either the presence or absence of a T → A replacement at −528, hence its previous report as an +ATA, −T event (19).

The dimorphisms observed at positions −541, −469, −316, +101, +181, +183, and +492 have not been reported previously to our knowledge, although the +101, +183, and +492 variants are present in Caucasians (S.M.F., unpublished data). The remaining 12 polymorphisms, including all of those within the gene, as well as the −521 and −529 length variants, have been observed in Caucasian, Asian, and African populations (16, 19–24).

There is an average of one nucleotide polymorphism every 175 bp and 1 in 700 bp vary between any two randomly chosen sequences in the Santo sample (n = 48)—i.e., nucleotide diversity at the locus, estimated from the number of observed polymorphic sites (25), is 0.14 ± 0.05%, a value not dissimilar to that estimated from RFLPs in the β-globin gene cluster (26). This level of β-globin nucleotide polymorphism is consistent with estimates of diversity at other loci. A recent survey using published cDNA and genomic sequences for 49 human nuclear genes found an average nucleotide diversity at silent sites of 0.076% (27), which is not substantially different from our own estimate of 0.14%. Although nucleotide polymorphisms are apparently nonuniformly distributed across the sequenced region, there is no significant difference in nucleotide diversity of "effectively silent" (synonymous or noncoding) sites (1) between coding, noncoding, and flanking regions sequenced (Fig. 3). However, nucleotide diversity in the introns is lower than in any other portion of the sequenced region.

Sequence Haplotype Polymorphism in Northern Vanuatu. The linkage relationships between the nucleotide polymorphisms observed in the two island populations are given in Fig. 2*B*. In the Espiritu Santo sample, the 19 variable positions were found to cosegregate as 17 sequence haplotypes ranging in frequency from 2% to 21%. Seven of these haplotypes accounted for the 12 normal β^A chromosomes from Maewo. The distribution of normal haplotypes on both islands is similar, despite the different sample sizes. The β-thalassemia mutation, which is found in Vanuatu almost exclusively on Maewo (29), was present on the same sequence haplotype background (*A4*) in all 12 cases.

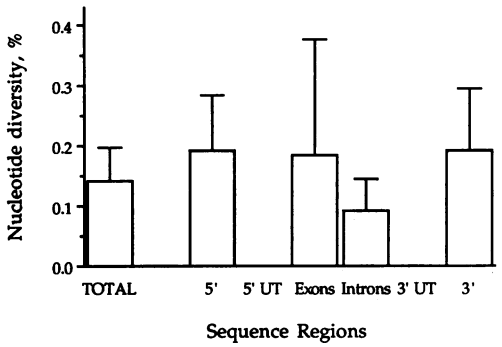


FIG. 3. Bar graph showing nucleotide diversity (25) relative to the number of effectively silent (1) sites in the 3-kb region. Nucleotide diversity within the 5′ and 3′ flanking regions and in combined coding (exons) and noncoding (introns) regions of the gene is shown to the right of the overall (total) estimate for the sequence. UT, untranslated region. Brackets represent the standard error of each estimate.

Haplotype polymorphism at the β-globin locus has previously been described in terms of gene "frameworks" or RFLP linkage groups. The 17 Melanesian sequence haplotypes fall into four groups, designated A (42% of the β^A chromosomes), B (32%), C (18%), and D (8%), whose within-gene polymorphisms correspond respectively to the four previously described β-globin gene frameworks 1, 2, 3-Asian, and 3 (20). The same haplotype groups are identified independently by unweighted pair group method with arithmetic mean (UPGMA) analysis (data not shown). Each group is distinguished by a characteristic pattern of linked polymorphic sites in the gene and its 3′ flanking region.

Our survey is unusual in observing four haplotype groups in one population. Other studies have only ever identified three gene frameworks, hence groups, in a particular ethnic sample: A, B, and D in Mediterraneans and A, B, and C in Asian Indians, Southeast Asians, Chinese, and American Blacks (7). Group D (framework 3) gene sequences, which have previously only been observed in Caucasian populations, make up the smallest proportion of the Melanesian sequence haplotypes. However, earlier analyses, which largely relied on RFLP analysis for identification of gene frameworks, could not have distinguished C and D type sequences, so the presence of group D β-globin gene sequence haplotypes in Asian populations may have been inadvertently overlooked.

Fig. 2 also shows the relationship of the 17 sequence haplotypes to surrounding restriction enzyme sites that make up the 3′ β-globin gene cluster RFLP haplotype. The *Hin*dIII and *Bam*HI sites that lie, respectively, 7 and 8 kb 3′ to the gene, form four haplotypes: ++, −−, −+, and +−. The C and D sequence groups are together associated with the −+ RFLP haplotype. Another pattern, +−, exclusively defines B-type sequences, although *B5* is also associated with the ++ haplotype. However, A-type sequence haplotypes, *A2* in particular, are found with three RFLP haplotypes, −+, ++, and −−. The *Hin*fI polymorphism, although lying only 175 bp 5′ of the start of the 3-kb region, is randomly associated with sequences from each haplotype group, as noted previously (30). Clearly, while RFLP analysis tends to confirm major distinctions observed at the nucleotide level, sequence anal-

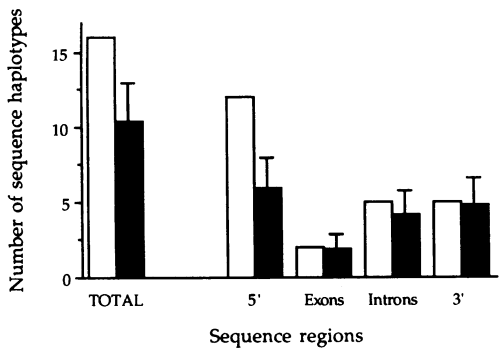


FIG. 4. Bar graph showing observed (□) versus expected (■) haplotype diversity in the Espiritu Santo sample as calculated by the Strobeck test (18). Results for the total sequence are given in the left-most comparison. Similar comparisons, calculated separately for the 5′ and 3′ flanking DNA and for combined coding (exons) and noncoding (introns) regions of the gene, are also shown. Brackets represent the standard error of each estimate. Length polymorphism was not taken into account in these comparisons.

ysis reveals chromosomal relationships that would have been overlooked or misinterpreted if only RFLP analysis had been used.

5′ RFLP haplotypes found in this study are similar to those identified in a previous sample from Vanuatu (29) and appear to be in linkage equilibrium with the 3′ haplotypes (data not shown).

Comparison of Nucleotide and Sequence Haplotype Polymorphism. We looked for evidence of population subdivision by comparing the observed number of sequence haplotypes with the number expected given the nucleotide diversity (18) (Fig. 4). However, rather than observing a reduction in haplotype diversity consistent with population subdivision, we found significantly more sequence haplotypes in Espiritu Santo than expected. The smaller Maewo sample was not analyzed. When similar comparisons are made for particular regions of the sequence (Fig. 4), haplotype diversity is only excessive in the 5′ flanking DNA. This nonuniform relationship between the nucleotide and haplotype polymorphism observed is suggestive of a molecular rather than demographic process; indeed, excess haplotype diversity is expected at loci undergoing recombination (31).

DISCUSSION

We used DNA sequence analysis to determine nucleotide polymorphism at the human β-globin locus in a Melanesian population. Although nucleotide diversity in Vanuatu is consistent with global estimates derived from cDNA and RFLP studies, the remarkable feature of the data is the considerable haplotype variation revealed by sequencing. Not surprisingly, previous approaches gave no intimation of this complexity, primarily because excess haplotype diversity, rather than resulting from the accumulation of new allelic variants via point mutation, is probably due to the shuffling of existing nucleotide polymorphisms by recombination, a scenario first envisaged by Watt (28) and later developed theoretically by Strobeck and Morgan (31).

The conclusion that recombination in the 5′ flanking region has generated much of the observed haplotype diversity is supported by the distribution of nucleotide sites among the sequence haplotypes. Haplotype groups, which most likely represent distinct sequence lineages, are each polymorphic for the same sites in the region 5′ to the β-globin gene. The likelihood of so many nucleotide dimorphisms recurring in different lineages as a consequence of random mutation is very small; they are better explained as resulting from interallelic recombination and gene conversion. Similar "recurrences" affecting sites that lie within the β-globin gene are present in sequences *C4* and *D2*. However, other differences within groups, such as those defining *A3* (−316), *C2* (−541), and *C3* (+181), are better ascribed to new substitutions arising on existing sequence haplotype backgrounds.

The excess diversity attributable to recombination is found predominantly in the 5′ flanking DNA of our 3-kb region. This is almost certainly not the full extent of recombinatorial activity, however, as the 5′ *Hin*fI site is also randomly associated with the haplotype groups defined by sequencing. Conceivably, then, the region immediately 5′ to the β-globin gene may be strongly influenced by, if not actually an extension of, an upstream recombination "hotspot" previously defined by linkage analysis (32) (see Fig. 2*A*). Recombination rates for this hotspot have been estimated to be 3–30 times higher than those of surrounding regions (33). Although probably initiating recombination outside the 3-kb sequenced region, the hotspot could be responsible for the recombination-driven haplotype diversity we have observed. Certainly, there are examples in both cultured mammalian cells and yeast of homologous recombination and gene conversion occurring at considerable distances from a recombinogenic origin (refs. 34 and 35 for example).

Although most of the interallelic exchanges occur in DNA 5′ of the β-globin gene, there are exceptions, such as sequence haplotypes *C4* and *D2*, which may have resulted from short-tract gene conversion around single polymorphic sites within the gene itself. Similar examples of possible conversion-generated diversity may be found among common hemoglobinopathies such as HbS, HbE, and some β-thalassemias. Many of these mutations are found on diverse 3′ RFLP haplotype backgrounds despite evidence that they have only recently reached high frequencies by malarial selection (8). While this has been ascribed to multiple independent mutations, interallelic gene conversion of single original mutations seems a more plausible explanation, especially in view of the fact that most of these putative conversions are found in the 5′-most (recombinogenic) part of the gene (36).

At the 3′ end of the sequenced region, where recombination and gene conversion have not been acting to increase sequence haplotype diversity, there is no evidence from the Strobeck test for increased genetic drift induced by population substructure or expansion from a recent population bottleneck. All of the previously identified β-globin gene frameworks are present in our northern Vanuatu sample, and the majority of the polymorphisms observed are known to be globally distributed. Such variation belies the archipelago's remote Pacific location and its relatively recent [within the past 3500 years (37)] settlement history. In contrast, limited evidence for lower average heterozygosity of both mtDNA loci (38) and RFLP sites (15), suggestive of increased genetic drift, has been obtained in studies that have compared Melanesians (from Papua New Guinea and the Solomon Islands) with other ethnic groups.

The number of polymorphic sites in our sample from northern Vanuatu provides a recombination-independent estimate of the age of the observed variation. As 1.6% difference (39) has accumulated at the β-globin locus during the 5 million years (40) separating chimpanzees from humans, 0.14% nucleotide diversity corresponds to an average age of sequence divergence of ≈450,000 years. Similarly, the pairwise difference between *A3* and *D1*, the two most divergent haplotypes in the sample, suggests that the 3′ nonrecombining portions of these sequences may have persisted for more than twice that time, ≈1.1 million years. These ages, based on data from a single population, are consistent with the coalescence time expected for a neutrally evolving human nuclear locus, given a long-term effective population size of 10,000 (41).

Therefore, our data cannot support the suggestion that the human species underwent a population bottleneck 200,000 years ago, as is controversially suggested by mtDNA analysis (42). The discrepancy between these estimated ages is likely to reflect the 4-fold greater impact of genetic drift, which proportionately reduces coalescence time for mtDNA compared with neutral nuclear DNA (41). To explore the genetic contribution to the modern gene pool of regional populations prior to 200,000 years ago, then, requires investigation of human nuclear loci.

Although only one locale has been sampled, two important conclusions have been drawn from our data. First, a substantial part of the haplotype diversity observed has been generated by recombination and gene conversion. Second, observed nucleotide diversity suggests an average age of sequence divergence of at least 450,000 years, supporting coalescence times predicted from neutral theory. This study of allelic variation at the human β-globin locus thus demonstrates the value of sequence analysis for the investigation of molecular and population processes. A fuller appreciation of the nature and extent of β-globin nucleotide polymorphism awaits similar analyses of other human populations. Global investigation of β-globin diversity alone will not solve the puzzle of human origins, however, and many other loci will need to be studied before consensus regarding the evolutionary history of *Homo sapiens* is realized.

We thank Dr. Mary Ganczakowski for providing the samples used in our analysis and Drs. Rhona Borts, Peter Donnelly, Jonathan Flint, Adrian Hill, Ed Louis, and an anonymous reviewer for helpful comments on the manuscript. S.M.F. was supported by grants from the Rhodes and Wellcome Trusts.

1. Kreitman, M. (1983) *Nature (London)* **304,** 412–417.
2. Stephens, J. C. & Nei, M. (1985) *J. Mol. Evol.* **22,** 289–300.
3. Hudson, R. R., Kreitman, M. & Aguade, M. (1987) *Genetics* **116,** 153–159.
4. Devor, E. J., ed. (1992) *Molecular Applications in Biological Anthropology* (Cambridge Univ. Press, Cambridge, U.K.).
5. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
6. Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C. & Markham, A. F. (1989) *Nucleic Acids Res.* **17,** 2503–2516.
7. Antonarakis, S. E., Kazazian, H. H., Jr., & Orkin, S. H. (1985) *Hum. Genet.* **69,** 1–14.
8. Flint, J., Harding, R. M., Clegg, J. B. & Boyce, A. J. (1993) *Hum. Genet.* **91,** 91–117.
9. Wainscoat, J. S., Hill, A. V. S., Boyce, A. J., Flint, J., Hernandez, M., Thein, S. L., Old, J. M., Lynch, J. R., Falusi, A. G., Weatherall, D. J. & Clegg, J. B. (1986) *Nature (London)* **319,** 491–493.
10. Old, J. M. & Higgs, D. R. (1983) in *The Thalassemias: Methods in Hematology*, ed. Weatherall, D. J. (Churchill Livingstone, London), pp. 74–102.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Poncz, M., Schwartz, E., Ballantine, M. & Surrey, S. (1983) *J. Biol. Chem.* **258,** 11599–11609.
13. Schwarz, K., Hansen-Hagge, T. & Bartram, C. (1990) *Nucleic Acids Res.* **18,** 1079.
14. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. & Sninsky, J. J. (1990) *Nucleic Acids Res.* **18,** 999–1005.
15. Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. & Cavalli-Sforza, L. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 839–843.
16. Moschonas, N., de Boer, E. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10,** 2109–2120.
17. Sutton, M., Bouhassira, E. E. & Nagel, R. L. (1989) *Am. J. Hematol.* **32,** 66–69.
18. Strobeck, C. (1987) *Genetics* **117,** 149–155.
19. Semenza, G. L., Malladi, P., Surrey, S., Delgrosso, K., Poncz, M. & Schwartz, E. (1984) *J. Biol. Chem.* **259,** 6045–6048.
20. Orkin, S. H., Kazazian, H. H., Jr., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G. & Giardina, P. J. V. (1982) *Nature (London)* **296,** 627–631.
21. Semenza, G. L., Delgrosso, K., Poncz, M., Malladi, P., Schwartz, E. & Surrey, S. (1984) *Cell* **39,** 123–128.
22. Cheboune, Y., Pagnier, J., Trabuchet, G., Faure, C., Verdier, G., Labie, D. & Nigon, V. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4431–4435.
23. Trabuchet, G., Elion, J., Baudot, G., Pagnier, J., Bouhass, R., Nigon, V. M., Labie, D. & Krishnamoorthy, R. (1991) *Hum. Biol.* **63,** 241–252.
24. Lapouméroulie, C., Dunda, O., Ducrocq, R., Trabuchet, G., Mony-Lobé, M., Bodo, J. M., Carnevale, P., Labie, D., Elion, J. & Krishnamoorthy, R. (1992) *Hum. Genet.* **89,** 333–337.
25. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
26. Kazazian, H. H., Jr., Chakravarti, A., Orkin, S. H. & Antonarakis, S. E. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 137–146.
27. Li, W.-H. & Sadler, L. A. (1991) *Genetics* **129,** 513–523.
28. Watt, W. B. (1972) *Am. Nat.* **106,** 737–753.
29. Hill, A. V. S., Bowden, D. K., O'Shaughnessy, D. F., Weatherall, D. J. & Clegg, J. B. (1988) *Blood* **72,** 9–14.
30. Kohen, G., Philippe, N. & Godet, J. (1982) *Hum. Genet.* **62,** 121–123.
31. Strobeck, C. & Morgan, K. (1978) *Genetics* **88,** 829–844.
32. Antonarakis, S. E., Boehm, C. D., Giardina, P. J. V. & Kazazian, H. H., Jr. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 137–141.
33. Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Waber, P. G., Boehm, C. D. & Kazazian, H. H., Jr. (1984) *Am. J. Hum. Genet.* **36,** 1239–1258.
34. Wahls, W. P., Wallace, L. J. & Moore, P. D. (1990) *Mol. Cell. Biol.* **10,** 785–793.
35. Symington, L. S., Brown, A., Oliver, S. G., Greenwell, P. & Petes, T. D. (1991) *Genetics* **128,** 717–727.
36. Matsuno, Y., Yamashiro, Y., Yamamoto, K., Hattori, Y., Yamamoto, K., Ohba, Y. & Miyaji, T. (1992) *Hum. Genet.* **88,** 357–358.
37. Bellwood, P. (1989) in *The Colonization of the Pacific: A Genetic Trail*, eds. Hill, A. V. S. & Serjeantson, S. W. (Oxford Univ. Press, Oxford), pp. 1–59.
38. Stoneking, M. & Wilson, A. C. (1989) in *The Colonization of the Pacific: A Genetic Trail*, eds. Hill, A. V. S. & Serjeantson, S. W. (Oxford Univ. Press, Oxford), pp. 215–245.
39. Savatier, P., Trabuchet, G., Faure, C., Cheboune, Y., Gouy, M., Verdier, G. & Nigon, V. M. (1985) *J. Mol. Biol.* **182,** 21–29.
40. Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. & Takahata, N. (1992) *J. Mol. Evol.* **35,** 32–43.
41. Takahata, N. (1993) *Mol. Biol. Evol.* **10,** 2–22.
42. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325,** 31–36.