A new ε globin HincII variant fragment length in a South African Negroid family

MICHELE RAMSAY, JENNIFER A THOMSON*†, AND TREFOR JENKINS
From the MRC Human Ecogenetics Research Unit, Department of Human Genetics, School of Pathology, South African Institute for Medical Research and University of the Witwatersrand; and *Department of Genetics, University of the Witwatersrand, Johannesburg, South Africa.

SUMMARY A new HincII ε globin variant is reported in a South African Negroid family. The usual HincII ε globin fragment lengths are 8·0 and 3·7 kb and the variant described here is 14·0 kb in length. The 14·0 kb fragment was generated by a site alteration removing the 3’ HincII site on a chromosome that already lacked the 5’ HincII site. It appears that there are differences among races with regard to the frequencies of the 8·0 and 3·7 kb alleles. The 3·7 kb allele is the more common form in Caucasoids whereas the 8·0 kb allele occurs in the majority of Negroid and Khoisan subjects.

A vast amount of ‘normal’ variation exists between people. Initially this variation was studied in proteins but now it is possible to detect silent or neutral variation in the DNA of coding sequences, intervening sequences, and regions flanking genes. Normal variation has been extensively studied in the β globin gene cluster embracing a region of approximately 35 kilobases on chromosome 11.1 2 It has been demonstrated that a small number of common sequence variations exist within the β globin cluster and these are identified by the presence or absence of certain restriction endonuclease sites leading to differences in DNA fragment lengths. Where the rarer fragment length occurs at a frequency greater than 1%, a restriction fragment length polymorphism is present at that locus. At present, 12 loci in the β globin gene cluster conform to this definition and three of these, the TaqI, HinfI, and HpaI RFLPs, are private and only occur at polymorphic frequencies in Negroids.

RFLPs are useful for tracing the origin and spread of mutations, for demonstrating the multiple origins of identical mutations (for example, β globin gene cluster3–5), for identifying chromosomes containing mutant genes, and for elucidating the molecular basis of inherited disorders.

Southern African populations are at present being screened for globin associated RFLPs in order to establish whether the same allelic morphs occur in these populations and whether the frequencies for these morphs are similar to those found in other populations. A random sample of Bantu speaking Negroids from the Venda chieftainship, a group of Bantu speaking Negroids of mixed origin (resident in Johannesburg, RSA), and a group of San (‘Bushman’) were screened for the HincII ε globin RFLP.

The HincII ε globin RFLP was first described in 1982 by Antonarakis et al6 in Greeks, Italians, Indians, Turks, and American Blacks. The 3·7 kb fragment occurred at relatively high frequencies in the Mediterraneans and Indians, whereas the 8·0 kb fragment showed a higher frequency in American Blacks. In the three population groups presented here, the 8·0 kb fragment also occurred at a higher frequency than the 3·7 kb fragment. These data indicate that there is a marked shift in allele frequency among Caucasoids when compared with Negroids.

Materials and methods

POPULATIONS STUDIED
Venous blood samples were collected in ACD (acid citrate dextrose) vacutainer tubes after informed consent had been obtained from the subjects. The 20 randomly chosen Venda subjects were school children from the Matangari School in the Sibasa district of Venda. The 18 Bantu speaking Negroids of mixed origin (Zulu, South Sotho, Xhosa, and Swazi) were employees of the South African Insti-
The 17 randomly chosen San ('Bushman') were participants in a general haematological survey conducted at Tsumkwe in Bushmanland, Namibia. Haematological indices were obtained on all the subjects and were within the normal range.

**Isolation of DNA**
The buffy coats were removed from the blood samples after centrifugation and were frozen until required for DNA extraction. After thawing, the samples were digested overnight at 37°C with proteinase K (100 μg/ml) and SDS (0.5% v/v) and the protein was removed by repeated phenol:chloroform extractions. The samples were then treated with RNase (100 μg/ml) and again digested with proteinase K (100 μg/ml). Phenol:chloroform extractions were again performed and the DNA was resuspended in 10 mmol/l Tris, pH 8, 1 mmol/l EDTA.

**Southern blot hybridisation**
Approximately 10 μg of DNA from each subject was digested with 25 units of HincII (Amersham) and electrophoresed in 0.6% agarose. The DNA was transferred to Schleicher and Schuell nitrocellulose by the Southern gel transfer method and hybridised to an ε globin DNA probe. The probe is designated pe 1-3 and a 1.3 kb EcoRI/BamHI insert of genomic ε globin DNA was used for labelling with 32P dCTP by nick translation. After extensive washing, the filters were autoradiographed using Trimax XD x-ray plates and exposed for one to four days.

**Results**
During the screening of the Venda population, a previously unreported HincII ε globin fragment length of 14.0 kb was observed (fig 1b). The 14.0 kb fragment was shown to be inherited in a Mendelian fashion in the family shown in fig 1a. In order to determine whether the 14.0 kb HincII ε globin fragment was due to a HincII site alteration or a DNA rearrangement, single and double digests were performed using HincII, HindIII, and BamHI. The latter two were chosen because each has one

---

**FIG 1**  
(a) Pedigree of the Venda family with the 14-0 kb HincII ε globin fragment. (b) Autoradiograph of DNA from the Venda family (lanes 8 to 11: 1.1, 1.2, 2.1, and 2.2 respectively) and from other randomly selected Venda subjects (lanes 1 to 7) digested with HincII and hybridised to the 1-3 kb EcoRI/BamHI fragment of the ε globin probe pe 1-3.
A new ε globin HincIII variant fragment length in a South African Negroid family

![Diagram](https://example.com/diagram.png)

**FIG 2** (a) Restriction map of the ε globin gene region on chromosome II (Hd, HindIII; Hc, HincII; B, BamHI; *polymorphic HincII sites). (b) Illustration of HincII sites giving rise to chromosomes with a 14-kb, 8-kb, and 3.7-kb HincII fragment respectively. The haplotypes denote the presence or absence of the two HincII sites.

site close to the ε globin gene and another distant from it: the distant HindIII site is on the 5' side while the distant BamHI site is on the 3' side (fig 2a). The Venda DNA sample was shown to be heterozygous for the 8-kb and 14-kb HincII fragments. As a control, DNA from a Caucasoid, heterozygous for the 8-kb and 3.7-kb HincII fragments, was used. The HindIII fragment was shown to be approximately 8-kb and the BamHI fragment about 3.7-kb in both subjects (fig 3, table 1), in agreement with the findings of Trent et al.² but not in agreement with the observation of 0.6 and 25-kb BamHI fragments by Fritsch et al.³ In addition, a double digest of BamHI and HindIII generated a fragment of 1.96-kb in both samples indicating that the 14-kb HincII ε globin fragment was not due to a DNA rearrangement. Double digests of HincII/HindIII and HincII/BamHI generated fragments of different sizes in the Caucasoid control and the Venda heterozygote, confirming that the 14-kb HincII fragment is, in fact, due to a HincII site alteration. The control had two HincII/HindIII fragments but only one HincII/BamHI fragment, indicating that the 3' HincII sites were identical on both chromosomes but that the 5' HincII sites were different. This confirms the result of Antonarakis et al.⁶ that the mutation giving rise to the 8-kb HincII fragment occurred in the 5' HincII site. Conversely the Venda heterozygote had two HincII/BamHI fragments and one HincII/HindIII fragment indicating that in the case of the 14-kb fragment the 3' HincII site had been lost, resulting in a (--) haplotype (fig 2b). The 2.8-kb HincII/BamHI fragment corresponds, therefore, to the (++) haplotype and the 8.5-kb HincII/BamHI fragment to the chromosome giving rise to the (--) haplotype. Because the HincII/BamHI fragment is only 8.5-kb in length and not 13.1-kb, it appears that the 3' HincII site mutation occurred on a chromosome that also contained the 5' HincII site mutation. This
latter observation seems likely because the 8-0 kb \textit{HincII} fragment is the more common allele in the Venda population.

In order to establish the frequencies of the 14-0, 8-0, and 3-7 kb \textit{HincII} \(\epsilon\) globin fragments in the Venda, the South African Negroids of mixed origin, and the San populations, and to determine whether the 14-0 kb fragment was present in the latter two populations, DNA from 20 randomly chosen Venda, 18 South African Negroids of mixed origin, and 17 randomly chosen San were screened for the \textit{HincII} variant. The 14-0 kb \textit{HincII} fragment was found in only one of the Venda subjects giving it a frequency of 0.025 in this population (table 2). The 14-0 kb \textit{HincII} fragment was not found in the South African

\textbf{TABLE 1} \(\epsilon\) globin specific restriction fragment lengths in the Caucasoid control and the Venda heterozygote.

<table>
<thead>
<tr>
<th>Restriction endonuclease</th>
<th>Fragments in kb in DNA from Caucasoid control</th>
<th>Fragments in kb in DNA from Venda heterozygote</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{HincII}</td>
<td>8-0, 3-7</td>
<td>14-0, 8-0</td>
</tr>
<tr>
<td>\textit{BamHI}</td>
<td>13-7</td>
<td>13-7</td>
</tr>
<tr>
<td>\textit{HindIII}</td>
<td>8-1</td>
<td>8-1</td>
</tr>
<tr>
<td>\textit{BamHI/HindIII}</td>
<td>1-9</td>
<td>1-9</td>
</tr>
<tr>
<td>\textit{HincII/BamHI}</td>
<td>2-8</td>
<td>2-8</td>
</tr>
<tr>
<td>\textit{HincII/HindIII}</td>
<td>7-4, 2-9&quot;</td>
<td>7-4, 2-9&quot;</td>
</tr>
</tbody>
</table>

DNA in lanes 1 and 2 was digested with \textit{HincII/BamHI}, lanes 3 and 4 with \textit{HincII/HindIII}, lanes 5 and 6 with \textit{BamHI}, lanes 7 and 8 with \textit{HindIII}, and lanes 9 and 10 with \textit{HindIII/BamHI}. Lane 11 contains \(\lambda\) phage DNA digested with \textit{HindIII} as a size marker.
Negroids, and San populations compared with those in other populations tested; the 14-0 kb HincII fragment length was not detected in any of these.

Discussion

It has previously been estimated that approximately one in every 100 nucleotides in the β globin gene cluster differs between persons. The abundance of RFLPs in this cluster is, therefore, not surprising, nor is the fact that other RFLPs, like the PvuII site 3' to the γ globin gene, are still being found.

An interesting fact which appears to be emerging is that for certain of these RFLPs there are differences in the frequencies of the allelic morphs in different populations. There are thus a number of common sequence variants, the study of which may be useful in tracing the origin and spread of mutations. To date the new HincII ε globin variant has been found in only one population, the Venda, at a frequency of 0.025. It is therefore unlikely that this allele will be important for linkage studies as other polymorphic variants in the β globin gene cluster occur more commonly. The high frequency of both the 3-7 and 8-0 kb alleles in Negroids does, however, make the HincII RFLP useful as a linked marker in prenatal diagnosis. Although this new HincII site variant was absent in the 18 Bantu speaking Negroids of mixed origin and the 17 San subjects studied, this does not exclude the possibility of its presence at a low frequency in these populations. Due to the small sample size it cannot yet be concluded that the 14-0 kb fragment is 'polymorphic'; it may be a rare variant in the Venda population. From the data available one could hypothesise that the mutation creating this variant fragment length occurred in the Bantu speaking Negroes after their divergence from the San and possibly the West African Negroes, as this mutation has not been reported in American Blacks. It is interesting to note that the 8-0 kb fragment is a more common allele in the Venda, the mixed Bantu speaking group, and American Blacks and it has almost reached fixation in the San where the 3-7 kb fragment frequency is only 0.03. In Mongoloids and Caucasoids, with the exception of the Greek population, the 3-7 kb fragment is the more common allele. It appears, therefore, that there is a significant population difference in the frequencies of the alleles at this locus. The altered HincII sites are in the DNA flanking the ε globin gene, approximately 1 kb each from the 5' and 3' extremities of this gene, indicating that they may possibly have some functional significance in the regulation of gene expression. It is possible that selection may, therefore,

![Diagram illustrating the polymorphic HincII sites giving rise to the (−−), (−+), (++), and (+−) haplotypes of the ψβ1 gene region. (Adapted from Antonarakis et al).](http://jmg.bmj.com/)

**TABLE 2 Frequencies of the HincII ε globin gene fragment lengths.**

<table>
<thead>
<tr>
<th>Population</th>
<th>Total no of chromosomes studied</th>
<th>Frequencies</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>8-0 kb</td>
<td>3-7 kb</td>
</tr>
<tr>
<td>Venda</td>
<td>40</td>
<td>0.675</td>
<td>0.300</td>
</tr>
<tr>
<td>Bantu speaking Negroids</td>
<td></td>
<td>0.706</td>
<td>0.294</td>
</tr>
<tr>
<td>of mixed origin</td>
<td>36</td>
<td>0.900</td>
<td>0.100</td>
</tr>
<tr>
<td>American Blacks</td>
<td>42</td>
<td>0.970</td>
<td>0.030</td>
</tr>
<tr>
<td>Greek</td>
<td>30</td>
<td>0.470</td>
<td>0.530</td>
</tr>
<tr>
<td>Italian</td>
<td>41</td>
<td>0.270</td>
<td>0.730</td>
</tr>
<tr>
<td>Indian</td>
<td>12</td>
<td>0.170</td>
<td>0.830</td>
</tr>
<tr>
<td>Turks</td>
<td>10</td>
<td>0.140</td>
<td>0.660</td>
</tr>
<tr>
<td>Asian Indian</td>
<td></td>
<td>0.220</td>
<td>0.780</td>
</tr>
<tr>
<td>South-east Asian</td>
<td></td>
<td>0.280</td>
<td>0.720</td>
</tr>
</tbody>
</table>

*Total number of chromosomes studied not indicated.
account for the racial differences, though the possible selective agent remains unknown. Despite the fact that the 14-0 kb HincII ε globin fragment has only been found in one family, it is interesting that there are two HincII site mutations, one on either side of the ε globin gene (fig 2), and this occurrence may indicate that HincII is possibly in the class ofMspI and TaqI which have been claimed to generate many more RFLPs than other enzymes studied to date.13 BothMspI and TaqI recognise the CpG dinucleotide which is often methylated in the human genome at the cytosine residue. Deamination of 5-methyl cytosine can lead to this residue being replaced by thymidine,14 thus resulting in the removal of the restriction site in which the CpG dinucleotide occurred. The recognition site for HincII is GTPyPuAC, and should the two central bases be CpG then the above argument would apply to the HincII sites as well. Further evidence for the possible increased base pair substitution rate at HincII sites is provided by the presence of two HincII polymorphic sites within and flanking the ψβ globin geneb (fig 4). In this area all four possible haplotypes generated by the two polymorphic HincII sites have been observed, whereas only three of the possible haplotypes around the ε globin gene have been observed. The (+−) haplotype that would generate an ε globin HincII fragment of 9-7 kb has not been found, but the number of subjects tested is too small to conclude that it does not occur.

We thank Dr T Maniatis for ρε 1-3 and Dr C Mathew for critical reading of the manuscript.

Addendum

In subsequent studies the 14 kb HincII ε globin fragment was found in the heterozygous state in a Xhosa subject and also in one American Black family. The latter was observed by C Boehm at the Johns Hopkins Hospital.

References


Correspondence and requests for reprints to Dr M Ramsay, Department of Human Genetics, South African Institute for Medical Research, Hospital Street, PO Box 1038, Johannesburg 2000, South Africa.