

Mortality among Human Immunodeficiency Virus Type 2-Positive Villagers in Rural Guinea-Bissau Is Correlated with Viral Genotype

NICHOLAS C. GRASSLY,^{1*} ZHENG XIANG,² KOYA ARIYOSHI,³ PETER AABY,⁴
HENRIK JENSEN,⁴ MAARTEN SCHIM VAN DER LOEFF,⁴ FRANCISCO DIAS,⁵
HILTON WHITTLE,³ AND JUDITH BREUER²

Wellcome Trust Centre for the Epidemiology of Infectious Diseases, Department of Zoology, University of Oxford, Oxford,¹ and Department of Medical Microbiology, St. Bartholomew's and The Royal London School of Medicine and Dentistry, London,² United Kingdom; MRC Laboratories, Medical Research Council, Banjul, The Gambia, West Africa³; Department of Epidemiology, Staten Serum Institute, Copenhagen, Denmark⁴; and Laboratório Nacional de Saúde Pública, Bissau, Guinea Bissau⁵

Received 30 January 1998/Accepted 17 June 1998

We present the results of a 6-year study of 131 human immunodeficiency virus (HIV) type 2 (HIV-2)-infected individuals from a rural population in Guinea-Bissau. Proviral DNA sequences 1.3 kb in length were obtained from each individual and, together with clinical data, including proviral load and CD4 and CD8 levels, were used to assess whether viral genotype influences clinical outcome. With a phylogenetic model, a correlation was found between viral genotype and mortality; this correlation was not due to confounding factors, such as age-specific viral strains or cohabitation of patients. The data provide strong evidence for the involvement of viral genetic factors in determining HIV disease progression in vivo. The pattern of association found suggests that virulence factors are multiple and scattered throughout the HIV-2 genome and can be rapidly gained or lost by the virus through a combination of mutation and recombination. These findings may lead to the identification of viral determinants of HIV disease progression.

The majority of individuals infected with human immunodeficiency virus (HIV) type 1 (HIV-1) develop AIDS within 10 years. However, approximately 10% of cases progress rapidly (2 to 3 years) to AIDS, while 5 to 10% of cases are slow or nonprogressive, with patients remaining clinically asymptomatic after 10 years (20, 34). HIV type 2 (HIV-2) is claimed to have a longer incubation period between infection and overt AIDS than HIV-1 (29) but has recently been clearly characterized as having a similar diversity of clinical outcome (36). Among adults under the age of 45 years, the ratio of mortality in HIV-2-infected to uninfected individuals has been shown to be 5:1. However, unlike mortality with HIV-1, which rises with age (6), the mortality ratio for older people with HIV-2 infection is close to 1:1 (36), suggesting a dichotomy between rapid progression to AIDS and slow progression or even nonprogression.

Susceptibility to HIV-1 infection and the rate of progression of disease have been strongly associated with polymorphisms in host genes coding for the CCR5 (11, 32) and CCR2 (39) chemokine receptors, which also function as HIV coreceptors. In addition, there is some effect from host immunoregulatory factors, including the major histocompatibility complex (30). In contrast, although HIV-1 and HIV-2 have been shown to differ in biological properties in vitro (8, 38), little evidence has accumulated concerning the importance of viral genetic factors

in determining disease progression in vivo (33, 40, 41). A reason for the current lack of certainty about the existence and role of viral virulence factors is the difficulty in classifying clinical outcome for HIV-1 infection, due to the continuum in the lengths of incubation period between infection and full-blown AIDS (7, 34). In contrast, the clear dichotomy in clinical outcome for HIV-2 infection (36) provides an opportunity to carry out systematic investigations into the relative roles of host versus viral genetic factors in determining the rate of progression to AIDS. Such investigations not only are important for the understanding and control of HIV-2 but also may reveal important features concerning the mechanisms leading to AIDS in chronic retroviral infections. In addition, there is some evidence that HIV-2 infection may protect against HIV-1 infection (2, 19, 23, 42). If viral genetic determinants of nonprogressive clinical outcome could be found for HIV-2, then this may have implications for the development of vaccines against both HIV-1 and HIV-2.

This paper presents results from a prospective study of HIV-2-infected residents of a village in rural Guinea-Bissau who, together with uninfected controls, have been monitored with clinical, virological, and immunological investigations for over 6 years (37, 43). These HIV-2-infected residents have suffered approximately three times the mortality rate of uninfected controls (37) and show a dichotomy in the rate of disease progression (1), in accordance with the findings for a similar cohort of HIV-2-infected individuals monitored in Bissau, the capital of Guinea-Bissau (36). Using novel phylogenetic methods, we investigated whether the mortality rate of HIV-2-infected individuals within this cohort depends upon viral genotype.

* Corresponding author. Mailing address: Wellcome Trust Centre for the Epidemiology of Infectious Diseases, Department of Zoology, University of Oxford, South Parks Rd., Oxford OX1 3PS, United Kingdom. Phone: 44 1865 271263. Fax: 44 1865 310447. E-mail: nicholas.grassly@zoo.ox.ac.uk.

MATERIALS AND METHODS

Subjects. A total of 134 HIV-2-positive subjects were originally entered into a cohort study between March and June 1991 and followed up with annual censuses until 1997 (37). All subjects were of the same ethnic origin, and none had received blood transfusions or blood products. Subjects were examined clinically, and blood samples were taken in both 1991 and 1997 and cryopreserved. The CD4 and CD8 levels in the cryopreserved blood samples taken in 1991 were measured previously (37, 43). In addition, a quantitative PCR was used to assess proviral load (4). Postmortem interviews were conducted with the relatives of all those who died. Lymphocyte DNA from 131 of the 134 HIV-2-positive subjects from whom samples were obtained in 1991 was available for analysis.

PCR amplification and sequencing. Proviral DNA was extracted from one 200- μ l aliquot of frozen lymphocytes obtained in 1991, and the *gag*, *pol*, *env*, and long terminal repeat (LTR) regions were amplified by the PCR. Primer sequences (with International Union of Pure and Applied Chemistry mixed base codes) and positions (in nucleotides) (numbered according to HIV-2_{ROD}) were as follows: *Gag* 1, AGGTTACGGCCCGCGGAAAGAAAA (603 to 627); *Gag* 2, CCTACTCCTGACAGCCGTCAGCATTCTTC (1612 to 1581); *Gag* 3, GGATTAGCAGAGAGCCTGTTGGA (677 to 700); *Gag* 4 (44), CCTTAAGCTTTGTAGAATCTATCTACATA (1466 to 1437); *Pol* 1, AATATTATAGT AGAYTCACARTATGT (603 to 627); *Pol* 2, CCWAYCYCTGRCADGYXG TYARCATYTCYTC (1612 to 1581); *Pol* 3, TATGTTGCATGGGTCCAG CCC (3397 to 3992); *Pol* 4, TTCATTGCTTCCACTACTCC (1279 to 1258); *Env* 1, TRCTATDARRTTTAGRTAYTG (6807 to 6851); *Env* 2, TCYCIDGGAG GCAAATAHAC (7340 to 7316); *Env* 3, TAGRTAYTGTGCCACRCCRRGG (6911 to 6935); *Env* 4, TSCCTACTYTRTGCCAVGTRT (9301 to 9277); *LTR* 1, TAACCAAGGGAGGACATGGG (9411 to 9431); *LTR* 2, TGGTGAGAGT CTAGCAGGG (9592 to 9574); *LTR* 3, AGGAGCTGGTGGGGAACGCCCT (9432 to 9453); and *LTR* 4 (5), AACACCAGGCTCTACCTGCT (9573 to 9553).

A reaction mixture (50 μ l) containing 200 to 400 ng of extracted proviral DNA, 1.5 mM MgCl₂, 5 μ l of 10 \times reaction buffer (Perkin Elmer), 200 mM each deoxynucleoside triphosphate, 50 pmol of each primer, and 1 U of *Taq* polymerase (Perkin Elmer) was subjected to hot-start denaturation at 95°C for 3 min followed by 35 cycles of 94°C for 1 min, 40°C for 2 min, and 72°C for 1.5 min and a final cycle of 94°C for 1 min, 40°C for 2 min, and 72°C for 2 min. Nested PCRs were carried out under the same conditions, except that 0.5 μ l of the first-round product was added to the reaction mixture and the annealing temperature was raised to 55°C. Nested PCR products were electrophoresed on a 1.5% low-melting-temperature agarose gel, and a product of the correct size was extracted with the Wizard DNA purification system (Promega). The mixture was ethanol precipitated and directly cycle sequenced in both directions on an ABI 377 sequencer with fluorescent dye terminators. The *env* and *LTR* products were sequenced with the PCR primers. Additional primers used to sequence *gag* and *pol* were as follows: *Gag* 5, CACGACAGAGAAAGTGAAA (810 to 830); *Gag* 6, TCTACTGTGCTGTGTTCCCTG (1279 to 1258); *Pol* 5, TATGTTGC ATGGGTCCAGCCC (3971 to 3992); and *Pol* 6, TTCATTGCTTCCACTACTCC (4521 to 4502). In the majority of cases, the consensus sequence was verified by sequencing of products from two separate PCRs.

Phylogenetic trees. The consensus sequences were aligned with ClustalV (21). Maximum-likelihood (ML) phylogenetic trees were constructed with a test version of PAUP* (d54) made available by David Swofford. Trees were constructed for all sequences independently and for a joint alignment of *gag*, *LTR*, and *pol* sequences. These trees were used when assessing the correlation between viral phenotype (see below) and various clinical features. In the case of any missing individual clinical data, trees were constructed with the relevant number of taxa. The *env* sequences were not included in a joint alignment, as in only 79 subjects were sequences available for all four regions. The substitution model used to construct the ML phylogenetic trees was F84 (for a description, see reference 16). ML estimates of the transition/transversion ratio were obtained for each tree. In the case of the joint *gag*, *LTR*, and *pol* alignment, different relative substitution rates were allowed for each gene.

Phylogenetic models. To assess whether there are viral genetic determinants of clinical aspects of HIV-2 infection (which we term the viral phenotype), models of how viral phenotype changes with underlying viral genotype are required. The phenotypic characters modelled in this paper are the status of the infected individual (dead or alive), CD4 and CD8 levels (and their ratio), and viral load. The genotype of the virus is the consensus sequence obtained by direct sequencing from the infected individual and is represented by its position in the phylogenetic tree.

In a very simple model, an HIV-2 isolate can be said to be either benign (found in nonprogressors) or fatal. If the virus changes its state at rate α , then the probability (P) of change between benign and fatal states over any time t is given by

$$P_{ij}(t) = \begin{cases} f_i + f_j e^{-\alpha t}, & i = j \\ f_i - f_j e^{-\alpha t}, & i \neq j \end{cases}$$

where f is the equilibrium frequency of benign or fatal isolates in the viral population. Given that a virus is in state i at time 0 and state j at time t , the likelihood (L) of this situation for a given rate α is calculated simply as $L(\alpha, t|ij) = P_{ij}(t)$.

A measure of the time of separation of two viral isolates is simply the number of substitutions per site that have occurred in their nucleotide sequences, estimated with an appropriate substitution model. Thus, given two viral nucleotide

sequences (e.g., the *gag* gene) and their viral phenotypes, the likelihood for this simple model of phenotypic change for a given rate α can be calculated. If substitutional information about the time of separation of the two viral isolates is discarded and the rate of phenotypic change and time are treated as a compound parameter αt , the ML estimate of this parameter for two viral isolates given their viral phenotypes can be found. For more than two isolates, substitutional information can be used to construct a phylogenetic tree which best represents their relatedness (as is done with PAUP*). The ML for a distribution of benign and fatal isolates on this phylogenetic tree can be found with a "pruning" algorithm (13–15), which relies only upon the accuracy of the reconstructed phylogenetic topology (not branch lengths).

With a similar procedure, the ML for a distribution of continuous phenotypic characters (e.g., CD4 and CD8 levels and viral load) on a given phylogeny can be calculated. The model implemented in this paper is that of Brownian motion (see, e.g., reference 14). In other words, the mean phenotype undergoes a random diffusion on an infinite linear scale. Thus, the probability (P) that a character with value x_1 has value x_2 after time t follows the normal density function

$$P_{x_1x_2}(t) = \frac{1}{\sqrt{2\pi\mu t}} \cdot \exp \left[-\frac{(x_2 - x_1)^2}{2\mu t} \right]$$

where μ is the rate parameter. As for the discrete model, this probability is equivalent to the likelihood $L(\mu, t|x_1, x_2) = P_{x_1x_2}(t)$. Again, the compound parameter μt is considered and substitutional information is used only to construct a phylogenetic tree relating the sequences. The ML for a distribution of continuous characters on a phylogeny can be calculated with a pruning algorithm similar to that used for discrete characters but in which the length of the internal branch is altered following pruning (14).

The significance of each ML value calculated for the phylogenetic model relating viral genotype to clinical features was assessed by first calculating the MLs for 100 randomizations of the clinical data. The observed ML was then ranked among these likelihoods to obtain a P value. If the observed ML fell within or was better than the best five randomized likelihoods, then it was concluded that the data fit the model significantly better than would be expected by chance at the 5% level.

ML values have limited statistical power in differentiating hypotheses, since even randomized data may have an ML value that is quite high (12). Tests for significance based on ML values therefore tend to be conservative. The likelihood ratio is a far more powerful test statistic but relies on the existence of two competing hypotheses for which ML values can be calculated. An alternative hypothesis against which to test the models of genotype-phenotype correlation described here is difficult to develop. It is possible to use the normal distribution as an alternative hypothesis for the observed continuous characters and the binomial distribution for the discrete characters. However, care would need to be taken in interpreting nonsignificant likelihood ratios (i.e., ratios for which the hypothesis of genotype-phenotype correlation cannot be rejected), since the validity of such results would depend upon the validity of the alternative hypothesis. Thus, at present, only the ML statistic is used.

Investigation of other correlates of mortality. Possible correlates of mortality, other than viral genotype, were investigated by survival analysis because of their suggested importance in a previous study (36). Factors investigated included subject age, gender, area of residence, spouse HIV-2 status, and activity as a prostitute. A Cox proportional-hazards model was used (9), with time since blood sampling as the underlying time scale. For analyses other than that of age of subject, age was controlled for as a background variable.

Identification of motifs and base changes associated with pathogenic viruses. For each site in the *gag*, *pol*, *LTR*, and *V3* alignments, the minimum number of changes in the respective ML phylogenetic trees necessary to explain the distribution of sequence diversity was calculated with MacClade (28). These changes were then subdivided into those occurring in viral lineages whose descendants were exclusively associated with mortality and those occurring in all other lineages. By use of a program written in C it was possible to obtain plots of the frequencies of nucleotide changes along the aligned sequences similarly subdivided. These plots displayed the site-specific changes associated with mortality against the background of changes occurring in all other lineages and hence allowed identification of any changes and/or motifs associated with mortality. A nonphylogenetic method for detecting "signature patterns" of query sequences was also used to assess whether there is a correlation of certain nucleotides with mortality (27). A signature pattern is comprised of nucleotides for which the ratio of their frequencies in the query sequence alignment and the background alignment is above a certain threshold. A computer program, VESPA (27), implementing this method was used to determine the frequencies of the nucleotides comprising the putative signature of sequences associated with mortality and their frequencies in the alignment of sequences isolated from individuals alive at the end of the follow-up period.

Nucleotide sequence accession numbers. The following sequences were deposited in the EMBL database under accession no. AJ008441 and AJ011191, 539 and 222 for *gag*, respectively; AJ008540, 667 for *pol*; AJ008283 and AJ011223, 316 and 272 for *env*, respectively; and AJ008317, 440 for the *LTR*.

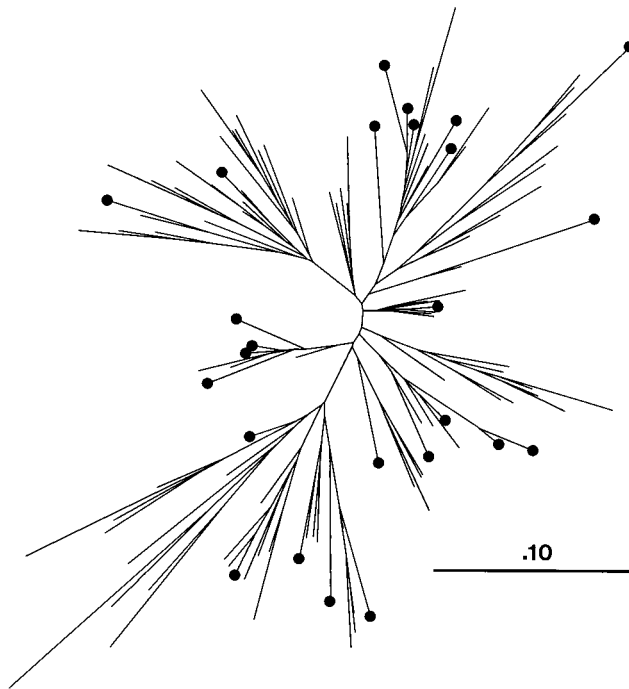


FIG. 1. ML phylogenetic tree of the HIV-2 isolates for the joint alignment of *gag*, *pol*, and LTR sequences, showing sequences isolated from people who subsequently died (marked by black circles). The estimated relative rates of substitution for the three regions of the HIV-2 genome were 1.19, 0.77, and 1.02, respectively. The estimated transition/transversion ratio was 1.65. The scale bar indicates the length of a branch with one substitution per 10 nucleotide sites.

RESULTS

Subjects and nucleotide sequences. During the study, 26 people died. Proviral *gag* (350 bp), *pol* (415 bp), *env* (V3; 357 bp), and LTR (164 bp) sequences were amplified from 123, 127, 84, and 124 patients, respectively. In 117 of the 131 subjects analyzed, including 25 of the 26 patients who died, at least three regions (*gag*, *pol*, and LTR) were amplified. The other patient who died and from whom sequences of the *gag* and V3 regions only were available was a 32-year-old woman who had a CD4 count of 32% and who died 4 days postpartum. Post-mortem interviews conducted with the relatives of the other 25 subjects who died could not exclude HIV as a cause of death in all cases, with the possible exception of a 44-year-old man.

However, all 25 patients who died were included in the analyses.

All sequences were found to be HIV-2 subtype A, in agreement with a previous analysis of a subset of the *gag* gene data set (44). The sequences have been deposited in the EMBL sequence database.

Phylogenetic trees. Figure 1 shows an ML phylogenetic tree for the 117 sequences of the joint alignment of *gag*, *pol*, and LTR sequences; sequences isolated from people who subsequently died are marked.

Phylogenetic models. A correlation was found between the death of HIV-2-infected individuals and viral genetic identity (Table 1). This correlation was highly significant for the joint alignment of LTR, *gag*, and *pol* sequences (P , <0.01) but became nonsignificant when the *gag*, *pol*, and LTR genes were examined individually (P , 0.45, 0.07, and 0.07, respectively). This result may reflect the increased phylogenetic resolution provided by use of a joint alignment with appropriate models for each gene (22) (although this effect may be counteracted by recombination occurring between the genes). A previous cohort study of HIV-2-infected individuals showed that deaths early in the follow-up period are more likely to involve rapid progressors (36). This trend was recently confirmed for the cohort studied here, with no excess mortality occurring in the HIV-2-positive cohort after 1995 (mortality rate ratio, 0.94, versus 3.38 before 1995 [1]). Analysis of only those deaths occurring prior to 1995 revealed a significant correlation with viral genotype for all but the *gag* sequences (Table 1).

No correlation was found between mortality and the V3 region of the envelope gene (P was 0.5 for both the analysis of all deaths and the analysis of deaths prior to 1995). This result may reflect the smaller number of V3 sequences than of *gag*, *pol*, and LTR sequences available for analysis. It may also reflect a lack of statistical power of the ML statistic.

It was possible to use these sequences to predict the in vitro replication phenotype of the HIV-2 isolates, since this phenotype was recently shown to correlate with certain amino acid changes in the V3 loop (3) as has been previously demonstrated for HIV-1 (18). The syncytium-inducing phenotype was, however, suggested by amino acid changes in only two subjects and was not found to correlate with mortality (χ^2 test of independence: χ^2 , 0.683; P , 0.41). These results suggest the importance of some other viral virulence factor(s) and confirm the observation that the syncytium-inducing phenotype tends to be more prevalent only close to the onset of AIDS (26, 38).

No correlation was found between viral genotype and high proviral load or low CD4 count (Table 1), despite previous findings of a relationship between the latter two and progres-

TABLE 1. Log likelihoods for a model of clinical outcome dependent upon viral genotype represented by different genes for a range of clinical features, together with P values denoting the probability of obtaining a likelihood this good by chance^a

Gene or region	Log likelihood (P) for:						
	All deaths ($n = 25$)	Pre-1995 deaths ($n = 15$)	Deaths of those less than 65 yr old ($n = 16$)	% CD4 level	% CD8 level	Ratio of % CD4 level to % CD8 level	Viral load
<i>gag</i>	-35.95 (0.45)	-26.07 (0.06)	-27.53 (0.07)	-141.92 (0.31)	-220.23 (0.41)	-244.59 (0.91)	-457.23 (0.55)
<i>pol</i>	-32.36 (0.07)	-21.69 (<0.01) ^b	-27.17 (0.10)	-175.62 (0.96)	-250.15 (0.91)	-142.85 (0.08)	-621.47 (0.10)
LTR	-30.36 (0.07)	-25.60 (0.05) ^b	-28.94 (0.35)	-144.78 (0.24)	-120.01 (0.17)	-256.39 (0.32)	-621.53 (0.36)
Joint alignment	-23.00 (<0.01) ^b	-17.26 (<0.01) ^b	-22.33 (<0.01) ^b	-138.61 (0.36)	-177.68 (0.26)	-143.07 (0.33)	-584.38 (0.89)

^a P values were obtained by ranking the actual likelihood among the likelihoods of 100 randomizations of the clinical data for each analysis (see Materials and Methods for details). A less negative log likelihood indicates a better model fit. However, the likelihood values are not comparable across the table owing to differences in phylogenetic topology and slight variations in the number of people included in the analysis due to missing data (see the text). n , number of deaths included in the analysis of mortality.

^b The P value was considered significant at the 5% level.

TABLE 2. Mortality of HIV-2-positive individuals subdivided according to age group relative to the oldest age group

Age (yr)	Mortality rate ratio	95% Confidence interval	P
15–34	0.27	0.08–0.93	0.039
35–54	0.15	0.04–0.52	0.003
55–64	0.17	0.05–0.58	0.005
≥65	1		

sion to AIDS in HIV-2-infected people (4). This result may be explained by the fact that of the 25 deaths included in our analysis, only 15 were associated with a high viral load (>100 copies/10⁵ CD4 cells) and low CD4 levels (<29%) at the time of blood sampling (in accordance with previous definitions for these categories [37]). Furthermore, the levels of HIV-2 provirus varied considerably, and the mean level was high even in subjects with high CD4 levels. Thus, it appears that these markers for disease progression are not consistent if the onset of AIDS is not imminent. A more suitable marker allowing early prognosis of the rate of progression to death in HIV-2-positive individuals may therefore be plasma viral load, as is the case for HIV-1 (31).

Investigation of other correlates of mortality. Using the Cox proportional-hazards model (9), we found no correlation between mortality and subject gender, area of residence, spouse HIV-2 status, or prostitution (*P*, 0.385, 0.984, 0.056, and 0.433, respectively). A significant correlation was found, however, between mortality and the age of HIV-2-positive individuals when the latter was treated as a continuous variable (mortality rate ratio, 1.06 per year [95% confidence interval, 1.03 to 1.09]; *P*, <0.001). Further analysis revealed that, although mortality was higher in those over 65 years old, there was no increase in mortality with age in those under 65 years old (test for trend: *P*, 0.593) (Table 2). This finding agrees with the dichotomy in clinical progression for HIV-2 infection (36) and previous observations for the same community (37).

Identification of motifs and base changes associated with pathogenic viruses. Consistent patterns of base changes in lineages leading to the pathogenic viruses were not found by use of the phylogenetic approach described. Similarly, no signature patterns were found to define the sequences associated with mortality. Table 3 shows the putative signature nucleotides of pathogenic sequences and their frequencies in both these pathogenic sequences and sequences isolated from patients alive at the end of the follow-up period. These frequencies are not very different, indicating no clear correlation between a particular viral motif or single nucleotide change and mortality in the regions sequenced. This finding probably reflects the fact that a relatively small proportion (1,300 bases) of the entire HIV-2 genome (~10,000 bases) was sequenced.

DISCUSSION

The data reported here reveal a correlation between viral genotype and mortality in HIV-2 infection in vivo. Furthermore, the fact that no distinct viral phylogenetic group (clade) is associated with mortality (Fig. 1) suggests that several viral virulence factors found within or linked to the *gag*, *pol*, and LTR regions sequenced can be gained or lost by the virus. Because such motifs were not detected within the sequences analyzed with the methods described here, linkage seems the most likely explanation.

It is possible that the correlation was confounded by other factors, which may covary in some way with both viral genotype

and mortality. Gender, area of residence, spouse HIV-2 status, and prostitution, previously suggested as risk factors (36), were all found not to correlate with mortality. However, age was found to correlate with the mortality of HIV-2-positive individuals when treated as a continuous variable (*P*, <0.001). In addition, age was found to correlate with viral genotype for the *gag* and LTR sequences when assessed with the likelihood methods described previously (*P*, 0.02 and 0.03, respectively). It is therefore possible that certain viral strains were circulating at certain times, infecting specific age groups when they were sexually active, and that the death of older subjects caused a correlation between mortality and viral genotype. However, although mortality was higher in those over 65 years old, there was no correlation between mortality and age in those under 65 years old (Table 2). Since consideration of the deaths of only people under 65 years old still resulted in a significant correlation with viral genotype (Table 1), age could not be confounding the in vivo correlation between viral genotype and mortality.

The possibility that discrete viral genetic elements are associated with mortality in HIV-2 infections is supported by experimental evidence derived from strains of simian immunodeficiency virus. Acquisition of virulence by some strains of simian immunodeficiency virus has been associated with changes occurring in the *env* (17, 24, 35), *gag* (25, 35), *tat* (17), and *nef* (35) genes and the LTR (24, 35). Furthermore, deletions in the HIV-1 *nef* gene have been described for a cohort of unrelated long-term nonprogressors who received transfusions from a single donor, supporting the notion that viral determinants may influence disease progression (10).

In conclusion, this paper provides evidence for heritable viral virulence factors important in determining the rate of disease progression in vivo. These results are compatible with the role of other determinants of virulence, including both host factors and viral genetic changes that may occur within individual hosts (intra-host evolution). Given the rapid mutation rate and turnover of HIV-2 within humans, the latter may be of

TABLE 3. Putative signature patterns for the V3, *gag*, LTR, and *pol* sequences associated with mortality, as detected with VESPA (27)

Signature	Position	Proportion of individuals	
		Dead	Alive
V3			
T	81	0.52	0.48
A	208	0.52	0.30
G	216	0.33	0.14
G	330	0.57	0.48
A	336	0.57	0.43
A	355	0.62	0.43
<i>gag</i>			
G	114	0.56	0.45
A	117	0.52	0.35
A	143	0.48	0.44
T	190	0.64	0.49
A	444	0.44	0.41
LTR			
T	34	0.56	0.44
T	71	0.52	0.24
<i>pol</i>			
G	140	0.52	0.39
A	192	0.56	0.33

importance, in particular with regard to escape from immune recognition or the use of a broader range of receptors. However, the recognition of heritable virulence factors now prompts the need for their precise identification and further research toward an understanding of the mechanisms by which they are acquired and interact with host determinants of disease progression.

ACKNOWLEDGMENTS

We are indebted to the people who took part in the community-based study and to Andrew Wilkins for his contribution to the design of the study. We also thank Pedro Biri Gomes for additional clinical data, Thiru Surentheran for technical help, and Eddie Holmes, Paul Harvey, and Robin Weiss for constructive criticism regarding the manuscript.

This work was supported by the MRC (J.B.) and BBSRC (grants to N.C.G.).

REFERENCES

- Aaby, P., H. Jensen, and M. Schim van der Loeff. Unpublished data.
- Aaby, P., A. G. Poulsen, O. Larsen, C. B. Christiansen, H. Jensen, M. Melbye, A. Naucner, and F. Dias. 1997. Does HIV-2 protect against HIV-1 infection? *AIDS* 11:939-940.
- Albert, J., P. Stalhandske, S. Marquina, J. Karis, R. A. M. Fouchier, E. Norrby, and F. Chiodi. 1996. Biological phenotype of HIV type-2 isolates correlates with V3 genotype. *AIDS Res. Hum. Retroviruses* 12:821-828.
- Ariyoshi, K., N. Berry, A. Wilkins, D. Ricard, P. Aaby, A. Naucner, P. T. N'Gum, O. Jobe, S. Jaffar, F. Dias, R. S. Tedder, and H. Whittle. 1996. A community-based study of human immunodeficiency virus type 2 provirus load in a rural village in West Africa. *J. Infect. Dis.* 173:245-248.
- Berry, N., K. Ariyoshi, O. Jobe, P. T. N'Gum, T. Corrah, A. Wilkins, H. Whittle, and R. Tedder. 1994. HIV type-2 proviral load measured by quantitative polymerase chain reaction correlates with CD4+ lymphopenia in HIV type 2 infected individuals. *AIDS Res. Hum. Retroviruses* 10:1031-1037.
- Biggar, R. J. 1990. AIDS incubation in 1891 HIV seroconverters from different exposure groups. *AIDS* 4:1059-1066.
- Cao, Y. Z., L. M. Qin, L. Q. Zhang, J. Safrin, and D. D. Ho. 1995. Virological and immunological characterization of long-term survivors of human immunodeficiency virus type-1 infection. *N. Engl. J. Med.* 332:201-208.
- Cheng-Mayer, C., D. Seto, M. Taten, and J. A. Levy. 1988. Biological features of HIV-1 that correlate with virulence in the host. *Science* 240:80-82.
- Cox, D. R., and D. Oakes. 1984. Analysis of survival data. Chapman & Hall, Ltd., London, England.
- Deacon, N. J., A. Tsykin, A. Solomon, K. Smith, M. Ludfordmenting, D. J. Hooker, D. A. McPhee, A. L. Greenway, A. Ellett, C. Chatfield, V. A. Lawson, S. Crowe, A. Maerz, S. Sonza, J. Learmont, J. S. Sullivan, A. Cunningham, D. Dwyer, D. Dowton, and J. Mills. 1995. Genomic structure of an attenuated quasi-species of HIV-1 from a blood transfusion donor and recipients. *Science* 270:988-991.
- Dean, M., M. Carrington, C. Winkler, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert, S. P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels, and S. J. O'Brien. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* 273:1856-1862.
- Edwards, A. W. F. 1992. Likelihood (extended ed.). The John Hopkins University Press, London, England.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368-376.
- Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35:1229-1242.
- Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25:471-492.
- Felsenstein, J., and G. A. Churchill. 1996. A hidden Markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* 13:93-104.
- Flaherty, M. T., D. A. Hauer, J. L. Mankowski, M. C. Zink, and J. E. Clements. 1997. Molecular and biological characterization of a neurovirulent molecular clone of simian immunodeficiency virus. *J. Virol.* 71:5790-5798.
- Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker. 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J. Virol.* 66:3183-3187.
- Greenberg, A. E., S. Z. Wiktor, K. M. Decock, P. Smith, H. W. Jaffe, and T. J. Dondero. 1996. HIV-2 and natural protection against HIV-1 infection. *Science* 272:1959.
- Haynes, B. F., G. Pantaleo, and A. S. Fauci. 1996. Towards an understanding of the correlates of protective immunity to HIV infection. *Science* 271:324-328.
- Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1992. ClustalV: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* 8:189-191.
- Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* 11:152-158.
- Kanki, P. J., G. Eisen, K. U. Travers, R. G. Marlink, M. E. Essex, C. C. Hsieh, and S. Mboup. 1996. HIV-2 and natural protection against HIV-1 infection—response. *Science* 272:1959-1960.
- Karlsson, G. B., M. Halloran, J. Li, I. W. Park, R. Gomila, K. A. Reimann, M. K. Axthelm, S. A. Iliff, N. L. Letvin, and J. Sodroski. 1997. Characterization of molecularly cloned simian human immunodeficiency viruses causing rapid CD4+ lymphocyte depletion in rhesus monkeys. *J. Virol.* 71:4218-4225.
- Kimata, J. T., and J. Overbaugh. 1997. The cytopathicity of a simian immunodeficiency virus Mne variant is determined by mutations in *gag* and *env*. *J. Virol.* 71:7629-7639.
- Koot, M., I. P. M. Keet, A. H. V. Vos, R. E. Y. DeGoede, M. T. L. Roos, R. A. Coutinho, F. Miedema, P. T. A. Schellekens, and M. Tersmette. 1993. Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann. Intern. Med.* 118:681-688.
- Korber, B., and G. Myers. 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res. Hum. Retroviruses* 8:1549-1560.
- Maddison, W. P., and D. R. Maddison. 1995. MacClade, 3.05 ed. Sinauer Associates, Inc., Sunderland, Mass.
- Marlink, R., P. Kanki, I. Thior, K. Travers, G. Eisen, T. Siby, I. Traore, C. C. Hsieh, M. C. Dia, E. Gueye, J. Hellinger, A. Gueyendiaie, J. L. Sankale, I. Ndoe, S. Mboup, and M. Essex. 1994. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* 265:1587-1590.
- McNeil, A. J., P. L. Yap, S. M. Gore, R. P. Brettell, M. McCall, R. Wyld, S. Davidson, R. Weightman, A. M. Richardson, and J. R. Robertson. 1996. Association of HLA types A1-B8-Dr3 and B27 with rapid and slow progression of HIV disease. *QJM Monthly J. Assoc. Physicians* 89:177-185.
- Mellors, J. W., C. R. Rinaldo, P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 272:1167-1170.
- Michael, N. L., G. Chang, L. G. Louie, J. R. Mascola, D. Dondero, D. L. Birx, and H. W. Sheppard. 1997. The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. *Nat. Med.* 3:338-340.
- Mocroft, A., C. A. Sabin, and A. N. Phillips. 1996. Lower prevalence and incidence of HIV-1 syncytium-inducing phenotype among injecting drug users relative to homosexual men—a response. *AIDS* 10:344-345.
- Munoz, A., A. J. Kirby, Y. D. He, J. B. Margolick, B. R. Visscher, C. R. Rinaldo, R. A. Kaslow, and J. P. Phair. 1995. Long-term survivors with HIV-1 infection: incubation period and longitudinal patterns of CD4+ lymphocytes. *J. AIDS Hum. Retrovirol.* 8:496-505.
- Novembre, F. J., P. R. Johnson, M. G. Lewis, D. C. Anderson, S. Klumpp, H. M. McClure, and V. M. Hirsch. 1993. Multiple viral determinants contribute to pathogenicity of the acutely lethal simian immunodeficiency virus *SIV_{smmPBj}* variant. *J. Virol.* 67:2466-2474.
- Poulsen, A.-G., P. Aaby, O. Larsen, H. Jensen, A. Naucner, I. M. Lisse, C. B. Christiansen, F. Dias, and M. Melbye. 1997. 9-year HIV-2-associated mortality in an urban community in Bissau, West Africa. *Lancet* 349:911-914.
- Ricard, D., A. Wilkins, P. T. N'Gum, R. Hayes, G. Morgan, A. P. Da Silva, and H. Whittle. 1994. The effects of HIV-2 infection in a rural area of Guinea-Bissau. *AIDS* 8:977-982.
- Schulz, T. F., D. Whitby, J. G. Hoad, T. Corrah, H. Whittle, and R. A. Weiss. 1990. Biological and molecular variability of human immunodeficiency virus type 2 isolates from The Gambia. *J. Virol.* 64:5177-5182.
- Smith, M. W., M. Dean, M. Carrington, C. Winkler, G. A. Huttley, D. A. Lomb, J. J. Goedert, T. R. O'Brien, L. P. Jacobson, R. Kaslow, S. Buchbinder, E. Vittinghoff, D. Vlahov, K. Hoots, M. W. Hilgartner, and S. J. O'Brien. 1997. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. *Science* 277:959-965.
- Soto-Ramirez, L. E., B. Renjifo, M. F. McLane, R. Marlink, C. O'Hara, R. Sutthent, C. Wasi, P. Vithayasai, V. Vithayasai, C. Apichartpiyakul, P. Auewarakul, V. P. Cruz, D. S. Chui, R. Osathanondh, K. Mayer, T. H. Lee, and M. Essex. 1996. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* 271:1291-1293.
- Spijkerman, I. J. B., M. Koot, M. Prins, I. P. M. Keet, A. Vandenhoeck, F. Miedema, and R. A. Coutinho. 1995. Lower prevalence and incidence of HIV-1 syncytium-inducing phenotype among injecting drug users compared with homosexual men. *AIDS* 9:1085-1092.
- Travers, K., S. Mboup, R. Marlink, A. Gueyendiaie, T. Siby, I. Thior, I. Traore, A. Diengarr, J. L. Sankale, C. Mullins, I. Ndoe, C. C. Hsieh, M. Essex, and P. Kanki. 1995. Natural protection against HIV-1 infection provided by HIV-2. *Science* 268:1612-1615.
- Wilkins, A., D. Ricard, J. Todd, H. Whittle, F. Dias, and A. P. Da Silva. 1993. The epidemiology of HIV infection in a rural area of Guinea-Bissau. *AIDS* 7:1119-1122.
- Xiang, Z., K. Ariyoshi, A. Wilkins, F. Dias, H. Whittle, and J. Breuer. 1997. HIV type 2 pathogenicity is not related to subtype in rural Guinea Bissau. *AIDS Res. Hum. Retroviruses* 13:501-505.