Simple human DNA-repeats associated with genomic hypervariability, flanking the genomic retroposons and similar to retroviral sites

Evgeny I. Rogaev

Laboratory of Genetics, All Union Research Center for Mental Health, USSR Academy of Medical Sciences, Moscow 113152, Zagorodnoe sh. 2, USSR

Received August 8, 1989; Revised and Accepted November 3, 1989

ABSTRACT

Earlier we found a human hypervariable genomic region (GVR). The DNA hybridization probe isolated from this region detects multiple hypervariability of restriction DNA fragments from genomic loci. The sequencing data suggest that the genomic instability and variability are associated with tandem DNA repeats. The DNA hybridization probe contains two families of simple DNA repeats designated as 'apo' and 'tau'. The (TC)_n-rich family of DNA 'tau'-repeats bears some similarity to the simple transcribed repeats of *Drosophila virilis*, simple repetitive motifs of the human proenkephaline gene exon 1, and short sites of retroviral LTR ends.

Apo-repeats show an unusual similarity to Rauscher viral env gene site. Besides GVR, apo- and tau-like repeats are localized in other genomic loci and can form separate tandem clusters and terminal repeats flanking certain copies of retroposons (Alu-SINES).

INTRODUCTION

The locus-specific and dispersed variable repetitive regions were isolated from the human genome (1–13). These regions proved to be highly polymorphic effective genetic markers for linkage analysis in human pedigrees and for individual's identifications (9,13). The analysis of clusters of tandem variable repeats may represent fundamental importance to the study of fast genetic and genomic changes in the human microevolution. It is not known to what extent hypervariable DNA-elements from different genomic regions are related to each other in their structure, function, evolution (4) and origin.

The group of variable dispersed human tandem repeats ('minisatellites') were shown to share a common 'core' site similar to the recombination signal (Chi) of *Escherichia coli*. But, the unit lengths and nucleotide sequences of the 'minisatellite'-repeats from different genomic loci are very different (8). Earlier, we selected random human genomic fragments (10, 11) to search for polymorphic DNA-repeats and revealed genomic probe, detecting variability of human DNA-restriction fragments. The polymorphism was also revealed upon comparative analysis of DNA samples from chimpanzees. The structural and similar analysis of tandem repeats from this genomic probe are shown

here. We reported the families of short tandem repeats, associated with genomic hypervariability, that has similar length and sequence in different loci, and, firstly, revealed the homology of human genomic elements (apo-repeats) to region of more variable component of retroviral genome-env gene region.

MATERIALS AND METHODS

DNA-preparation and blot hybridization

Genomic DNA was isolated from human cells (14,15). PstI-digested DNA were separated in 0.7% agarose gel and transferred by blotting to nylon filters ('Hybond'). The resulting blots were hybridized with DNA GVR prepared by nick translation. Approximately 400 ng of pGVR DNA, PE2.2 or pEP500 was labeled with ³²P dCTP to a specific activity of 4 × 10⁸ cpm/μg of DNA. The hybridization procedure was carried out in the presence of 50% formamide, 5 × SSC, 1 mM NaH₂PO₄, 100 U/ml of heparin (16), 0.01% polyvinylpyrrolidone, 3% sarcosyl and 200 μg/ml of polyA-DNA (14 hours, 42°C). Blot washing was performed in 0.1 × SSC and 0.3% SDS (3 hours, 67°C).

Cloning and DNA-sequence analysis

The 2.85 kb GVR (PstI-TVRI-6 fragment (12)) was originally cloned as a random genomic fragment into pBR322 (11) and was subcloned into pUC19. Subcloning of EcoRI/PstI-EP500 fragment was also performed in pUC19. For sequencing the 12 clones with systematic deletions of GVR were obtained as described (17), except that SI-nuclease was used instead of DNaseI, which prevented the degradation of linearized molecules (18). Transformation of clones was performed in the rec A⁻ bacterial host HB101, BMH7118, JM109 strains of *E. coli*.

Sequencing of clones was carried out according to Maxam-Gilbert method from 3'-³²P-HindIII, -EcoRI, -XhoI, -BamHI sites of GVR (Fig. 4) or pUC19 polylinker.

To diminish compression of poly-G sites and to stimulate denaturation, 50–60% formamide was sometimes used instead of urea.

RESULTS

The aim of our previous studies was to search for unstable repeats of human DNA. To this end we screened a plasmid library having

random human DNA genomic PstI-fragments. An analysis of randomly selected evolutionary conservative clones demonstrated that one of them carrying a PstI-restriction genomic fragment (GVR 2.85 kb) showed an hypervariability (11) with regard to many polymorphic restriction fragments, although the number of polymorphic fragments was lower than that of Jeffreys 'minisatellites' (Fig. 1,2).

At the same time, no differences in DNA patterns were observed, when different tissues of the same individual were compared (Fig. 2).

Fragment GVR was separated into two subfragments, EP500 and PE2.2 (Fig. 3). Blot hybridization of the EP500 probe revealed only one or two PstI-bands that were similar in size to that of the cloned PstI-GVR fragment (Fig. 2c). This finding suggests that the EP500 fragment is a unique region of the genome.

Thus, the multiple hypervariability of the genetic loci appears

Figure 1. Interindividual variability of PstI-patterns detected in human genomic DNA with the GVR-region as a probe. Blot-hybridization with placental DNAs of random unrelated individuals (only more polymorphic region of 2.8–5.5 k.b. patterns is shown).

to be linked with the other part of the GVR-probe, i.e., PE2.2 (Fig. 3). Indeed, sequencing of the genomic variable region (GVR) (Fig. 3a, b) revealed that this part of the probe contains two non-homologous clusters of simple DNA tandem repeats termed by us as 'apo' and 'tau'.

Apo-repeats

The four copies of DNA apo-repeats (37–38 b.p.) are adjacent to the G/C enriched region. Another DNA apo-repeat is located at a distance of 67 b.p. from the apo-cluster. A comparison of our sequence with the nucleotide bank ('GenBank') sequences revealed six homologous tandem apo-repeats located in intron 3 of the apolipoprotein CII gene (19). When analysing the nucleotide sequence of the recently isolated R-ras gene (20), we happened to detect in the 5'-region of this gene ten very diverged tandem copies of apo-repeats. It is worthwhile mentioning a high degree of homology of apo-monomers in one locus, ApoCII, and a high degree of monomer divergence in the other, GVR and 5'-region of R-ras gene. Thus, all the five GVR apo-monomers appear to show a 50–78%, ten R-ras gene apo-like monomers a 50–60%, whereas the ApoCII apo repeats a 92–100% homology.

At the same time, the homology of 'consensus' sequences of these clusters is as high as 85–98%. Since the GVR apo-repeat depicted in Fig. 4 shows a far greater homology (96%) to the ApoCII apo-repeats than the units of its own GVR cluster, the probable explanation of this homology is that the copy of this repeat has transposed recently from the diverged apo-cluster of GVR to the ApoCII gene intron 3 and amplified there up to non-diverged subunits. Here it seems appropriate to mention the homology (80%) of the GVR apo-repeat with the env gene region (21) of the Rauscher virus (Fig. 4) and some, although not too great, similarity of the Rauscher virus region to the region adjacent to the apo-cluster (200 b.p. region having a 55% homology) (Fig. 3b). We also observed a 70% homology of DNA apo-repeats with long (45 b.p.) direct DNA repeats flanking the Alu-family member in the 5' region of the apolipoprotein E-gene

Figure 2. Interindividual variability and ontogenetic stability patterns detected in human genomic DNA with the human genomic GVR regions as probes. Blot-hybridization of the 32 P-labeled: (a) GVR-fragment containing apo- and tau-repeats with PstI-human DNA from muscle tissue (1) and brain (2) of one embryo and from white blood cells of unrelated random Caucasian (3); (b) PE2.2 fragment, containing apo- and tau-repeats, with placental tissues of two randomly selected individuals; (c) EP500 probe with PstI-digested DNA from organs of one individual: spleen (1), heart (2), kidney (3), various divisions of brain (4–6) and different organs of other individual: heart (7), lung (8), brain (9) and PstI-digested DNA pBR (PstI-GVR-fragment in pBR 322).

Figure 3. The structure of the genomic variable region (GVR). (a) Some structural peculiarities of GVR; Sp6-sites; G/C enriched regions; DNA apo-repeats; DNA tau-repeats; the region containing the EP500 unique genomic fragment. (b) The nucleotide sequence of the PE2.2 probe from GVR: 1–97 b.p.-region homologued to region (67%) of bovine β -crystalline subunit -B1 gene, 172–321 b.p. – apo-repeats, 387–423-apo-repeat, 367–565 b.p.-region homologued to Rauscher env gene region (55%), 1345–1372 b.p. – (TCTAGC)₄ microrepeats, 1422–1940-tau-repeats; 3–548, 939–1351 b.p. more long open reading frames of translation.

(ApoE) (22) located, together with ApoCII, on the 19th chromosome. The *D. virilis* DV 192 DNA-fragment two tandem repeats (36 b.p.) (23) bears a slight homology (60%) to apo-repeats (Table 1).

Tau-repeats

More than 20 tandem tau-repeated DNAs having a consensus structure, C₅ (TC)₄ TG₃ (TC)₂ TGT (Fig. 5) were localized at a distance of 1 kb from the DNA apo-cluster in GVR. The cluster composed of this tandem repeats bears a similarity to the

transcribed DNA repeats of DV 192 (23) and (TC)_n-rich repetitive motifs of the human proenkephaline gene exon 1 (area of exon 1060–1360 b.p.) (24). The regions of the complete exact homology in exon do not exceed 13–15 b.p. (TCTGTCTGTCTGT, CTCTGTCCCTCTCTC, TCTCT-CTGAGTCT).

When computer analysing the nucleotide sequences of the R-ras gene (20), besides apo-like repeats, we revealed tau-like repeats, located in intron 1 at a distance of 2 kb from apo-like repeats. The R-ras tau-consensus (C₄ (TC)₄ TG₃ (TC)₂ TGT),

Figure 4. Sequence homology between DNA apo-repeats in GVR and human apolipoprotein CII gene (ApoCII) intron 3, R-ras-gene intron 1, Rauscher mink cell focus forming virus env-gene and 45 b.p. direct repeats in the 5'-region of the human apolipoprotein E-gene (ApoE). The GVR apo-consensus was derived from five members of GVR apo repeats, the ApoCII apo-consensus — from six repetitive monomers of the ApoCII gene, R-ras-apo-consensus from ten monomers of the R-ras gene. The comparative analysis of the GVR apo-consensus and the Rauscher virus region also involved the A- and CCA-nucleotides, flanking monomer 5 in the GVR apo-repeat and duplication T present in this monomer. The GVR apo-repeat showing a far greater (96%) homology to the ApoCII apo-repeats than to the GVR cluster monomers (60–80%) is underlined.

that we estimated for eleven very diverged tau-like repeats (40–100% homology) is very similar to GVR-tau-consensus. In this case too, similar to the apo-like repeats in the ApoE-gene locus, the R-ras tau-repeats flanked the member of the Alu-family (Fig. 6). The results of a comparative analysis suggest that the tau-repeats at one flank of the Alu-repeating sequence in the R-ras gene intron bear a greater (81–100%) homology to the tau-repeats located in the corresponding position at the other flank than the tau-repeats both from left and right clusters (45–77% homology). This was paralleled with the formation of the identical terminal repeats (92% homology) having the length of 93 b.p. (Table 2, Fig. 6). The clusters of hypervariable tandem repeats can thus be regarded as 'hot' sites of retroposon insertions in the gene loci and/or may play the role of direct retroposon repeats.

Conventionally, a tau-consensus can be divided into two domains, one of which (vir) is identical to the *Drosophila* DV 192 simple repeats, whereas the other one (R) contains short sequences similar to the viral recombination ends (25–28) (Fig. 5).

DISCUSSION

High DNA-polymorphism associated with simple DNA repeats

The probe GVR or PE2.2, containing tau- and apo- repeats, allowed detection of polymorphic restriction fragments when

Table 1. Homology of GVR apo-consensus to apo-like elements from other genomic region ($P < 0.001$)

	ApoCII gene intron	R-ras gene intron	ApoE-gene 5'-region	DV 192 fragment <i>D. virilis</i>	Env gene region of R. virus
GVR	98%	86%	78%	60%	77%

The consensus of clusters are compared. Corresponding consensus are derived from 5 copies of 36–38 b.p. GVR repeats (unknown chromosomal localization), 6 copies of 37 b.p. Apo- CII-intron 1 repeats (chromosome 19 (19)), 10 copies of 5'-region R-ras gene repeats (chromosome 19 (20)), 2 copies of 45 b.p. 5'-ApoE-region repeats (chromosome 19 (22)), 2 copies of 36 b.p. *D. virilis* DV192 repeats.

hybridized to human DNA cleaved with each of restriction enzymes tested: BamHI, EcoRI, HindIII, HaeIII, SacI, MvaII, PstI, MspI. The most polymorphism is observed on the PstI-digested DNAs.

An examination of DNA samples from 21 individuals (Moscow Caucasians) revealed not less than 19 polymorphic PstI-fragments in the 0.6–12 kb size range. Pairwise comparisons of two unrelated individuals can permit the establishment of differences in 8–12 PstI-bands. Since GVR region does not have internal PstI-sites, the obtained data suggest that GVR probe detects several homologous loci, not less than 3 which are variable. The GVR-region itself has a small variability in the length of alleles

Figure 5. A GVR cluster of DNA tau-repeats and a tau-consensus. Only 11 monomers of more than 20 tau-repeats in the GVR region are shown. The Vir-domain is identical to the DNA tandem repeats of *D. virilis*; the R-domain contains the sites identical to retroviral recombination sites. Identical GGGTCT sites at the ends of 38 b.p. inverted repeats Tn3, 11 b.p. inverted repeats of the murine Moloney virus (Mo-Mu SV) and of other retroviruses and the sites located at the ends of HIV LTR R-repeats are shown.

(2.7–3.0 kb) determined probably by the difference in 1–10 copies of tau-repeats in different alleles.

The unremitting homology between regions of clusters of GVR apo-repeats and ApoCII-, Rras-apo-repeats reaches 20–60 b.p., and between GVR tau-repeats and Rras-tau-repeats — 25 b.p. This is enough to identify these loci by GVR-probe in hybridization experiments. The observed polymorphism can be caused either by these loci or by other areas of genome with unknown localization.

The reason for the hypervariability of certain loci is the difference in the VTR copy number. The situation in the GVR region seems to be very much alike (Fig. 1, 2). Moreover, tau-elements are characterized by dinucleotide deletions (CT or CC) which occur frequently. The independent nature of these deletions can be substantiated by the fact that most of other mutations in identical (with regard to deletions) monomers are different (Fig. 5).

The tandem periodicity of monomers and dinucleotide inside the monomers, the purine-pyrimidine asymmetry (5:1 for the tau-repeats) in the strands and presence of sites similar to the

retroviral sites seem to be responsible for local rearrangements of observed tandem repeats and for their distribution within genome.

Structural organization and evolution of simple tandem repeats

Tau- and apo-repeats are organized in clusters in different genomic loci, and are likely to concentrate on the 19th chromosome (ApoE-, ApoCII-, R-ras-genes). In some loci the repeated monomers are very diverged, in others — homogeneous. In some cases the clusters of tau- and apo- repeats flank Alu-family members. It is suggested that the clusters are 'hot spots' of inserts of Alu-repeats, the insertion of which duplicates a copy or a long block of several copies of simple tandem repeats. This proved, for example, by high homology of tau-monomers of the left flank of Alu-sequence to tau-monomers of right flank by general high divergency of Rras-tau-monomers.

We also observed a nonrandom association of non-homologous apo- and tau-like DNA tandem repeat families in GVR, R-ras gene, *D. virilis* genomic segment. The region in the ApoCII gene

Figure 6. The genomic structure of apo- (white arrows) and tau- (black arrows) like repeats. A tight linkage of apo- and tau-repeat clusters in the genomic loci and their ability to form midterminal repeats (MTR) at the ends of short retroposons are shown.

Table 2. Homology (%) of tau-like repeats (L) of left flank from Alu-sequence to tau-like repeats (R) of right flank from Alu-sequence in R-ras intron region.

	L1	L2	L3	L4	L5	L6
R1	50	78	99	59	50	50
R2	73	77	67	100	72	65
R3	50	65	50	68	100	42
R4	50	73	72	63	46	82

The more homologous tau-like monomers are framed. These monomers form middle-long terminal repeated sequences of Alu-retroposon copy.

also contains, apart from apo-repeats (36–38 b.p.), the repeats carrying a $(TC)_n$ motif, $(TG)_n (TC)_n$ (Fig. 6). An analysis of nucleotide sequences of two independently cloned alleles (19, 29) showed that in this case, too, the dinucleotide repeats, i.e., $(TC)_7(AC)_{22}$ in one and $(TC)_8(AC)_9$ in the other allele of the Apo CII gene, are variable.

An interesting finding is the similarity, despite certain genomic instability, of simple VNTR in phylogenetically unrelated organisms. This peculiarity can be revealed both by direct comparisons of primary structures (e.g., human tau- and apo-repeats *D. virilis* DV192 simple repeats) and by hybridization of VNTR (e.g., 'minisatellites' (30) and M13 (31, 32, 33)). The evolutionary conservatism may reflect either the peculiarities of distribution of these DNA repeats or their functional significance.

Many VNTR have been isolated from gene regions during the analysis of genes. Presumably, such a predominant localization is not occasional. Thus, we have isolated a GVR, containing VNTR, as a random fragment from the total human DNA library without its preliminary testing for encoding functions. In this case, too, VNTR seems to be surrounded by regions typical for encoding genes. Evidence for this assumption can be derived from the presence of G/C enriched region in the GVR, GGGGCGG hexanucleotides detected by SpI- polymerase as well as from a high number of CG-sites, long open reading frames of translation (300–550 b.p.), a region (97% b.p.), bearing a 67% homology with the calf β -crystalline gene and hybridization of GVR probe with genomic DNA animals from frog to primates (in press). The simple tandem element clusters of genes can be the cause of genomic instability and possibly may thus be responsible for the populational heterogeneity with regard to the activity of certain genes.

Not long ago in eucaryotic genome there were discovered genetic elements, similar almost to all main components of retroviral genome: LTR, gag-, pol-genes. Such elements can be represented by some coding cell sequences (34) or long dispersed repeats (LI) of animals (35). Besides endogenous proviruses in eucaryote's genome no sequences similar to env genes have been known, env genes being the most variable components of retroviral genomes. In this connection one should point a direct similarity of nucleotide sequences of apo-repeats with the Rauscher retroviral env-gene region which is responsible for coding of the proline rich part of protein. It is suggested that some short clustered DNA repeats in the human genome as well as long disperse (LI) ones can originate from the retroviral genomes.

In animal genomes the groups of multicopy complex pericentromeric tandem satellites are quite common. These satellites from different organisms are usually unhomologous to each other, but can have common ancestors. Thus, for 170–340 b.p. α -DNA-repeats, occurring only in primate genome, there has been found an internal repeated motif as a possible ancestor: TGAAAAA. The comparison (37) shows a high degree of this element with a simple ancestor sequence of a complex (234 b.p.) satellite of mouse: TGAAAAATG (38). Another, well known group of DNA tandem repeats with a simple structure (minisatellites) is concentrated in other genome regions close to telomers (39). Minisatellites are widely variable in their sequence and length of the repeated units (9–60 b.p.) in different clusters even of one genome, but also have a simple common 'core' site (10–15 b.p.): $C_2T C_2T GC(C)_2 AC_2 TC_2$ (8). We have established similarity of the 'core' site as well as the similarity of the RecBC- dependent Chi-site of $(C_2A C_2AGC)$ with the site apo-consensus $(C_2CC_2 AGC(C)_3 TC_2 TC_2)$ and with the consensus of hypervariable tandem repeats of M13 protein III gene $(C_2 AC_2 A(C)_3 TC)$ (31).

These data suggested that some simple nucleotide sequences emerging either from genome sites or from 'traces' of viral sequences stimulate self reproduction and the reproduction of the adjacent areas, corresponding to the concept of the 'selfish' DNA.

Hence, apo- and tau-simple repeats are characterized by a number of features, such as the tight linkage of apo- and tau-like clusters in different regions of genome, similarity of sequences to retroviral sites, ability to form direct moderately long repeats at the ends of short retroposons as well as tandem clusters.

Different groups of tandem repeats are appropriate for genotyping: for identification of the specific human (40), sex (41) and individual belonging of the DNA (DNA fingerprinting) (9) and linkage analysis (13). DNA probes containing apo- and tau-repeats as well as other VNTR, can be used for the analysis of mechanisms of VNTR families evolution and investigation of hypervariable regions for mapping of the human genome.

ACKNOWLEDGEMENTS

The author thanks Dr. I.I.Chumakov, Dr. F.B.Berdichevsky, Dr. B.B.Kapitonov and Professor M.E.Vartanyan for helpful in the present work and discussion.

REFERENCES

- Wyman, A. and White, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6754–6758.
- Ulrich, A., Dull, T.J., Gray, A., Philips, J.A. and Peter, S. (1982) *Nucl. Acids*

- Res. 10, 2225–2240.
3. Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) *Nature* **295**, 31–35.
4. Jarman, A.P. *et al.* (1986) *EMBO J.* **5**, 1851–1863.
5. Capon, D.J. *et al.* (1983) *Nature* **302**, 33–37.
6. Stoker, N.J. *et al.* (1985) *Nucl. Acids Res.* **13**, 4613–4622.
7. Knott, T.J. *et al.* (1986) *Nucl. Acids Res.* **14**, 9215–9216.
8. Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) *Nature* **314**, 67–73.
9. Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) *Nature* **316**, 76–79.
10. Rogaev, E.I. in *Genetics and Biochemistry of Microorganisms: Biotechnology*, 81 (Moscow, 1986) (in Russian).
11. Rogaev, E.I. and Shapiro, Yu.A. (1987) *Bull. Exp. Biol. Med. (USSR)* **1**, 57–58.
12. Rogaev, E.I. (1989) *Nucl. Acids Res.* **17**, 1246.
13. Nakamura *et al.* (1987) *Science* **235**, 1616–1622.
14. Kunkel, L.M. *et al.* (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1245–1249.
15. Blin, N. and Stafford, D.W. (1976) *Nucl. Acids Res.* **3**, 2303–2308.
16. Singh, L. and Jones, K.W. (1987) *Nucl. Acids Res.* **12**, 5627–5638.
17. Hong, G.F.J. (1982) *J. Mol. Biol.* **158**, 539–542.
18. Zaitsev, I.Z. and Rogaev, E.I. (1986) *Mol. Biologia (USSR)* **20**, 663–673.
19. Wey, C.F. *et al.* (1985) *J. Biol. Chem.* **260**, 15211–15221.
20. Lowe, D.G. *et al.* (1987) *Proc. Natl. Acad. Sci. USA* **48**, 137–146.
21. Bestwick, R.K., Boswell, B.A. and Kabat, D. (1984) *J. Virol.* **51**, 695–705.
22. KiPaik-Yuonng *et al.* (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3445–3449.
23. Taurtz, D. and Renz, M.J. (1984) *Mol. Biol.* **172**, 229–235.
24. Horikawa, S. *et al.* (1983) *Nature* **306**, 611–614.
25. Takeya, T. *et al.* (1979) *Nucl. Acids Res.* **6**, 1831–1841.
26. Dhar, R. *et al.* (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
27. Ou, C.J., Boon, L.R. and Yang, W.K.A. (1983) *Nucl. Acids Res.* **11**, 5603–5620.
28. Rather, L. *et al.* (1985) *Nature* **313**, 227–284.
29. Foio, S.W., Law, S.W. and Brewer, B.J. (1987) *FEBS Lett.* **213**, 221–226.
30. Jeffreys, A.J. and Morton, D.B. (1987) *Anim. Genet.* **18**, 1–16.
31. Vassart, G. *et al.* (1987) *Science* **235**, 683–684.
32. Georges, M. *et al.* (1988) *Cytogenet. Cell Genet.* **47**, 127–131.
33. Ryskov, A.P. *et al.* (1988) *Genetics (USSR)* **2**, 227–238.
34. Toh, H., Ono, M. and Miyata, T. (1985) *Nature* **318**, 388–389.
35. Hattori, M., Kuhara, S. and Takenaka, O. (1986) *Nature* **321**, 625–628.
36. Wirth, Th., Glogler, K. and Baumruker, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3327–3330.
37. Zaitsev, I.Z. and Rogaev, E.I. (1986) *Mol. Biologia (USSR)* **20**, 674–682.
38. Hoz, W. and Altenburger, W. (1981) *Nucl. Acids Res.* **9**, 683–696.
39. Royle, N.J. *et al.* (1988) *Genomics* **3**, 352–360.
40. Rogaev, E.I., Vetchinkina, A.A. and Yurov, Yu.B. (1989) *Molecular Genetics Microbiol. Virol. (USSR)* **4**, 6–10.
41. Cooke, H.J. (1976) *Nature* **262**, 182–186.