

## Numerous Length Polymorphisms at Short Tandem Repeats in Human Cytomegalovirus

CLARA L. DAVIS,<sup>1\*</sup> DAWN FIELD,<sup>1</sup> DAVID METZGAR,<sup>1</sup> ROBERT SAIZ,<sup>2</sup> PHILLIP A. MORIN,<sup>2</sup>  
IRENE L. SMITH,<sup>3</sup> STEPHEN A. SPECTOR,<sup>3,4,5</sup> AND CHRISTOPHER WILLS<sup>1,5\*</sup>

*Department of Biology,<sup>1</sup> Department of Pediatrics,<sup>3</sup> Center for AIDS Research,<sup>4</sup> and Center for Molecular Genetics,<sup>5</sup> University of California, San Diego, La Jolla, California 92093-0116, and Axyx Pharmaceuticals, La Jolla, California 92037<sup>2</sup>*

Received 6 January 1999/Accepted 12 April 1999

**We show the presence of numerous short tandem repeats in the human cytomegalovirus (HCMV) genome and assess their usefulness as molecular markers. The genome is shown to contain at least 24 microsatellite regions that exhibit length polymorphisms. Insertion-deletion polymorphisms at these short tandem repeats are common (80% of repeats examined are polymorphic among two laboratory strains and 10 clinical isolates). This is the first report of widespread microsatellite length polymorphism in a viral genome. Some regions are highly polymorphic: one was revealed by DNA sequencing to contain length variants at five closely linked sites, which combined resulted in 10 variants for this region among the 12 strains and isolates examined. This study not only provides a new molecular marker system for this virus but also extends our understanding of microsatellite polymorphism in two important ways. First, variable-length repeats in HCMV can be considerably shorter than polymorphic repeats previously found in other organisms. Second, highly variable microsatellite repeats are not confined to prokaryotes and eukaryotes, as previously assumed. This variation provides a useful marker system for distinguishing viral isolates, and similar markers are also likely to be found in other large-genome DNA viruses.**

Short tandem repeats, or microsatellites, defined as iterations of short motifs made up of one to six bases, are known to be hot spots of length mutation in higher (42) and lower (14) eukaryotes. Even prokaryotic species have short iterations of short tandem repeats (14, 15) that can vary in length (16, 33), as well as longer hypermutable repeats that regulate virulence factors (28). Insertions and deletions at microsatellite loci are generally considered to be due to replication slippage errors (36, 42). Polymorphisms that result from these slippage errors are valuable in pathogenic species because of their use as epidemiological markers (13, 26, 27).

Microsatellites are unusual for the following reasons. First, because of replication slippage, they undergo a much higher rate of both insertion and deletion mutations than nonrepetitive regions. Second, unlike point mutations and insertions-deletions in nonrepetitive DNA, microsatellite mutations are highly reversible, with back mutation rates similar to forward mutation rates (8, 25, 46, 48). In some cases, microsatellites can provide pathogenic microorganisms with functional variation adaptive in rapidly changing environments (28, 29).

It has been suspected that viral genomes may also be prone to the accumulation of replication slippage errors (23). Variation has been found at the minisatellite level (defined as repeats of a motif of 11 to 60 bp [30]) in herpes simplex virus, although no clinical correlates of this variation have been discovered (44). Such large, minisatellite repeats have also been found within Epstein-Barr virus (4, 20, 21, 24, 46) and have proven useful in epidemiological studies of this virus (12, 37). These mutations affect very long repetitive regions, but

other mutations that affect repeats of very short motifs have been found.

Several examples of functional microsatellite tracts have been found among different classes of viruses. Functional mononucleotide repeats have been found in mengovirus (10, 22), vesicular stomatitis virus (VSV) (2), hepatitis C virus (50), and human respiratory syncytial virus (18). Changes in the length of trinucleotide and hexanucleotide repeats at the hemagglutinin cleavage site in avian influenza virus have been associated with increased virulence (31, 49). While these microsatellite tracts function in different ways within each virus, their presence in a variety of lengths in several types of viral populations demonstrates that mutational processes leading to expansion and contraction of nucleotide repeats do occur in viruses.

Although some effects of repetitive DNA sequences within viruses have been demonstrated, the full extent of naturally occurring variation at a variety of repeat loci, particularly at short repeats, has not been investigated for any virus. Studies in humans have shown that repeats less than five units long are typically monomorphic, and it has been generally assumed that extremely short iterations do not undergo replication slippage (47). Long repeats have been shown experimentally to be more mutable than short ones in yeast (43). Extremely short repeats, however, have not been well investigated in any organism. In this study, we have investigated the prevalence of length polymorphisms at short tandem repeats in human cytomegalovirus (HCMV). HCMV is a large double-stranded DNA virus of clinical importance. HCMV infection is considered endemic throughout the world, with seroprevalence in adults ranging from 50 to 90%, and infections are associated with a diverse range of clinical syndromes (39). We have examined the potential of insertion-deletion events at short tandem repeats to differentiate clinical viral isolates. It might seem unlikely that the short repeats in this genome would be sites of length

\* Corresponding author. Mailing address: Department of Biology 0116, University of California, San Diego, La Jolla, CA 92093. Phone: (619) 534-4113. Fax: (619) 534-7108. E-mail: cldavis@ucsd.edu and cwills@ucsd.edu.

polymorphisms. However, abundant length polymorphisms were discovered at many of these repeats.

#### MATERIALS AND METHODS

**Computer survey of the genome.** A computer survey of the HCMV AD169 genome (7) (GenBank entry HEHCMVCG) was carried out by using a program written in True Basic for the Macintosh and available from D.F. on request (12a). The program searches for all possible mono- through hexanucleotide repeats and has been previously described (14). Expected repeat distributions are derived from the mean distribution of repeat sequences in randomly generated genomes of identical length and base content (the HCMV genome is 57% G+C). We targeted our attention on sites that had at least four trinucleotide repeat units or five dinucleotide repeats or where the mononucleotide tract was nine or more bases long. This resulted in 85 potentially polymorphic repeat sites.

**Source of strains.** A total of 12 isolates were used to test for polymorphisms at short repeats. HCMV laboratory strains AD169 (34) and Towne (32) were obtained from the American Type Culture Collection (Rockville, Md.). Of the 10 clinical isolates examined, 8 were cultured from five urine specimens and three blood specimens of seven individuals diagnosed with AIDS-related CMV retinitis. Two isolates were cultured from bronchial alveolar lavage of bone marrow transplant recipients who were receiving treatment for CMV pneumonitis. CMV cultures were expanded on human foreskin fibroblasts maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U of penicillin per ml, 0.01% streptomycin, and 20 µg of amphotericin B per ml. DNA was purified from fully cytopathic CMV-infected human foreskin fibroblasts by phenol extraction and ethanol precipitation as described previously (38).

**PCR and analysis.** Primer pairs were designed to amplify 42 of the 85 longest mono-, di-, and trinucleotide repeats in the HCMV genome. A total of 30 primer pairs were manually designed, each pair amplifying a product containing from one to four target repeats (Table 1) and purchased from Research Genetics (Huntsville, Ala.). PCR was performed as previously described (13), with the addition of 0.1 µl of fluorescence-labeled dCTP from Perkin-Elmer (Foster City, Calif.). A Perkin-Elmer 2400 thermocycler was used for the reactions, with a PCR program of 40 cycles of 1-min denaturation at 94°C, 1-min annealing at various temperatures depending on primer (Table 1), and 1-min extension at 72°C. All products were analyzed on an ABI 373 automated DNA sequencer (Applied Biosystems, Inc.) using labeled dCTP incorporation. Genescan software was used to search for polymorphisms by length. All polymorphisms were confirmed by analysis of independently amplified PCR products. Sequencing to confirm length changes was carried out on an ABI 373 sequencer, and all reported sequences were determined twice in each direction. Before sequencing, all reaction products were cleaned by using a QiaQuick PCR Clean-up kit from Qiagen (Santa Clarita, Calif.).

#### RESULTS

The computer survey revealed that the HCMV genome contains only repeats with very short iterations (two to five units), and there is a slight excess of the longer mono-, di-, and trinucleotide repeats compared to randomized genomes of equivalent nucleotide composition and size (14). Observed and expected repeat distributions are compared in Table 2. For all repeat classes, longer repeats are more common than expected; however, in the HCMV genome the maximum number of repeat units of a di- or trinucleotide motif is only five, and the longest mononucleotide tract is 13 bases.

We surveyed 42 of the longest tandem repeats in 12 HCMV isolates, using 30 PCR primer pairs. Despite the short lengths of the repeats analyzed, 24 of 30 PCR primers (80%) amplified multiple length forms (Table 1), demonstrating repeat polymorphism, in this small group. All 12 isolates could be uniquely identified by using the 24 polymorphic regions. Figure 1 is a diagram of the HCMV AD169 genome showing positions of the polymorphic and monomorphic microsatellites that have been identified in this study.

Variability in the number of repeat units at some of the targeted short tandem repeats was confirmed by sequencing, as shown in Fig. 2. In most cases, for the di- and trinucleotide repeats, more variants were revealed by sequencing than by size separation. For example, separation by length at the TRAN3 locus showed 4 length variants, while sequencing revealed a total of 10 different variants among the 12 strains. The

microsatellites in UL69 and TRAN3 are both strikingly complex, with polymorphisms in more than one contiguous or neighboring repeat. Further, some of the polymorphic repeats in these clustered microsatellites (13) have as few as two repeat units, similar to the pattern seen in short clustered microsatellites in *Candida albicans* (27).

#### DISCUSSION

The 12 HCMV isolates examined were distinguished from one another by variability at 24 repeat regions. Therefore, this study provides a substantial panel of molecular markers with which to investigate evolutionary and epidemiological patterns in this important pathogen. Because the genome of HCMV is smaller than those of the eukaryotes and prokaryotes in which microsatellite variation has been previously utilized, it will be possible to survey virtually all microsatellite variation in HCMV populations. Our results suggest that this will also be a useful approach to the development of molecular markers in similar viruses.

The microsatellites screened here do not yet constitute an exhaustive survey of the HCMV genome; we surveyed only 49% of the total repeats longer than the chosen cutoff (tracts of four trinucleotide, five dinucleotide, or nine mononucleotide repeats), and these excluded the regions of the Toledo genome that are not found in Towne or AD169 (6). Therefore, a variety of additional length polymorphisms may be present throughout the rest of the genome. Examination of these additional repetitive regions will increase the likelihood of finding specific polymorphic markers that are linked to biological or pathological phenotypes.

A single set of primers often flanks multiple short polymorphic interactions, and sequencing reveals that length variants from different isolates with the same overall length are often made up of more than one sequence (Fig. 2). This phenomenon, in which apparently equivalent length variants can be shown to have different mutational histories, is an example of homoplasy (40), and many cases of homoplasy in microsatellite length variants can be resolved by sequencing (11, 19, 27). Thus, variation in length offers only a first order of discrimination. Sequencing of alleles or viral length variants adds a second level and is necessary to obtain maximum resolution between viral isolates.

Sequencing to confirm length variation at tandem repeats also revealed a variety of point mutations. It has been shown that there are more point mutations in regions around microsatellites than in nonrepetitive regions (9). Targeting of microsatellites therefore permits the detection of more point mutations than would be found if random segments of the genome of the same length were sequenced.

Many of the HCMV repeats are found in noncoding regions and are associated with a high local level of base pair polymorphism. This suggests that these polymorphisms are evolutionarily neutral or nearly so. They are therefore likely to provide appropriate markers for epidemiology and strain identification. Other repeats are found in known or putative coding regions, promoters, operators, and other functional DNA. These may not be neutral and may therefore be limited in utility as markers. These microsatellites, however, may provide adaptive variation important to viral evolution and genetic variability, perhaps similar to the functionally important mononucleotide runs found in VSV (2) and respiratory syncytial virus (18). If changes in these microsatellites are found to directly influence the expression of a gene or function of a gene

TABLE 1. Primer pairs found to be polymorphic in this survey

Name	Sequence <sup>a</sup>	Expected size (bp) <sup>b</sup>	Target repeat(s) <sup>c</sup>	Position in genome <sup>d</sup>	Temp (°C) <sup>e</sup>	No. of variants <sup>f</sup>
HCMVTRL8C	F, cccccgagacgagaccacc R, ccatacctcgcttcggaccccc	252	<b>(GGC)4</b>	7287	70	2
HCMVUL23C	F, ggtcgcgactcttaggagg R, cttcccgtttgactcgcgtgc	229	<b>(A)10</b>	28762	70	4
HCMVUL30C	F, cctccacacgctcagccgc R, gcgccgtaacacgctgcc	282	<b>(TG)5</b>	37254	70	2
HCMVUL37C	F, ggctcgtatgggttcacctcgggcg R, ggagaccgagcggctcccgtggactgg	157	<b>(TTC)4</b>	52405	60	2
HCMVUL38N	F, gcagcatgtgggctaataagg R, gctcgcgaccaattcgttgg	400	<b>(A)10</b>	51083	70	3
HCMVUL46C	F, cctcgcggtttgctttgagc R, cccgcagctcctatcaac	138	<b>(CG)5</b>	60374	68	2
HCMVUL50C	F, ggccgatgacggtgatgggtcg R, ggccggaaagcggctcctcctcgg	348	<b>(CTC)4...(GGC)4...(GAG)4</b>	72429	70	3
HCMVUL67AN	F, gcctttggtgctcgttgagcc R, gtggggacgggaatcgtatgc	280	<b>(T)10</b>	96977	63	2
HCMVUL67BC	F, gacatcattcccgtccccc R, gaggacgagagacgacgacc	303	<b>(GCG)4</b>	97476	65	2
HCMVUL68C	F, gggagggagagacgtggggctcg R, cgctgccacgcccattacgagg	414	<b>(T)11</b>	97928	60	7
HCMVUL69C	F, gcttaactgatgaccccgtcgc R, cctcgtaatcggcgctcggcagcg	209	<b>(GGT)5(GGC)4(TGC)3</b>	98290	65	5/9
HCMVUL95C	F, ggctgagctgtacgtctttgttgg R, ggccagctcttgataacgatgggg	388	<b>(GGT)5</b>	139531	60	4/4
HCMVUL96N	F, cggtttcggcagctcgcatttgg R, gcgcagcttcgcgcatccactgg	297	<b>(A)10</b>	140406	70	3
HCMVUL102C	F, cctccaccgctccaccacc R, cccccagcgaacgcacc	167	<b>(TTC)4</b>	148933	60	2
HCMVTRAN1N	F, cccgtaaggacttacc R, gctgtaagtctcacctagc	397	<b>(A)11</b>	156864	55	3
HCMVTRAN2N	F, ccatacctctagtgttttcg R, ccaccgtgataagaagacc	307	<b>(TGTA)4</b>	157332	55	2
HCMVTRAN3C	F, cctctcctcagtggtatcgtg R, cctcctccccatcatcttctcc	342	<b>(TGC)4...(T)10...(TAT)4</b>	157835	60	4/10
HCMVUL111C	F, ccatacctctcggcagcactgcg R, cgtaatcctcggcagcactgcg	431	<b>(GGT)4...(A)10...(A)10...(CGA)4</b>	159451	62	4/5
HCMVUL112AC	F, ggaggacagcaacggcagagg R, ggactaccgctgttctcggagg	233	<b>(GGT)3</b>	161192	63	2
HCMVUL122N	F, ggctgagaacagtatcagg R, gcgtgacacgtttattgagtagg	330	<b>(AT)5</b>	170988	60	3/5
HCMVUL123I2N	F, gcacatgggtcacaacagagg R, ggtgctgttaacggtggagg	295	<b>(A)9</b>	172660	60	—/6
HCMVUL123I1N	F, ccctccaccgtaacagcacc R, gactcatggtcctcggcag	265	<b>(TGG)4</b>	173066	70	3
HCMVUL132C	F, ggaatctgccagctggtgcttcc R, gccggatttgtcctcagccc	307	<b>(CCT)4</b>	177227	65	2/2
HCMVUS35C	F, gacgaagatcccgatgtgac R, cgtagcgtacgcccaacggc	325	<b>(G)12</b>	225448	65	—/6

<sup>a</sup> F, forward; R, reverse.

<sup>b</sup> Expected length of the PCR fragment in the AD169 genome (GenBank entry HEHCMVCG).

<sup>c</sup> Ellipses indicate repeats separated by a stretch of nonrepetitive sequence; contiguous repeats are not separated by ellipses. Repeats in bold are those found within open reading frames according to the GenBank annotation.

<sup>d</sup> Nucleotide number of the first repeat within the primer from the same reference sequence.

<sup>e</sup> Annealing temperature used for PCR.

<sup>f</sup> The number of variant alleles as determined by length analysis and sequencing (number of length variants/number of sequence variants).

product, studies of these polymorphisms may provide novel insight into gene regulation in the virus.

While microsatellites are frequently used as markers in mapping, gene function, and population studies of higher eukaryotes (1, 5, 17), the full extent of such repeats has not been determined in any virus. Other classes of repeats have been examined. Many studies have used polymorphic restriction enzyme sites as epidemiological markers. Herpesvirus restriction enzyme polymorphisms have been examined in several studies,

some of which have found relatively little variation (3, 45). Some restriction enzyme variants have been of use in distinguishing among different strains of VZV (41) and even in detecting different patterns of herpesvirus evolution in different parts of the world (35). Microsatellites are likely to evolve more quickly than restriction sites. This will make them useful for following changes in pathogen populations over short time scales.

None of the repeats examined in this study are long enough

TABLE 2. Observed and expected distributions of mono-, di-, and trinucleotide repeats in the HCMV AD169 genome<sup>a</sup>

No. of repeat units	Mononucleotides			Dinucleotides			Trinucleotides		
	No. observed	No. expected	Deviation	No. observed	No. expected	Deviation	No. observed	No. expected	Deviation
2	31,203	32,112 ± 30.8	-	8,209	7,394 ± 13.2	+	4,307	2,518 ± 8.8	+
3	7,152	8,241 ± 18.9	-	644	486 ± 5.8	+	275	43.7 ± 1.3	+
4	2021	2,153 ± 7.3	-	70	33 ± 1.1	+	<b>27</b>	1.15 ± 0.3	+
5	659	572 ± 5.3	+	<b>5</b>	1.9 ± 0.4	+	<b>2</b>		
6	199	154 ± 2.3	+						
7	69	42 ± 1.6	+						
8	24	11 ± 0.6	+						
9	<b>27</b>	3 ± 0.4	+						
10	<b>17</b>								
11	<b>2</b>								
12	<b>3</b>								
13	<b>2</b>								

<sup>a</sup> Target repeats for study are in boldface. The expected number of repeats is the mean of 20 genomes with standard error. All observed values were significantly different from expectation (in all cases,  $P \leq 0.001$ ). "Deviation" indicates whether the observed number was significantly more (+) or less (-) than expectation. Tetra-, penta-, and hexanucleotide repeats are extremely rare and match random expectations and are therefore not shown.

to have been included in a similar survey of a eukaryotic genome for microsatellite length variants (48). Despite previous assumptions that such polymorphisms should be extremely rare, this study demonstrates the potential of very short repeats to provide polymorphic molecular markers. It also reveals an important role for insertion-deletion mutations as a mecha-

nism for producing genetic variation in viral genomes, as previously demonstrated for eukaryotic and prokaryotic genomes. This study provides an important new type of molecular marker with which to investigate questions of epidemiology and evolution within this virus. While these short tandem repeats are shorter than those frequently examined for studies of

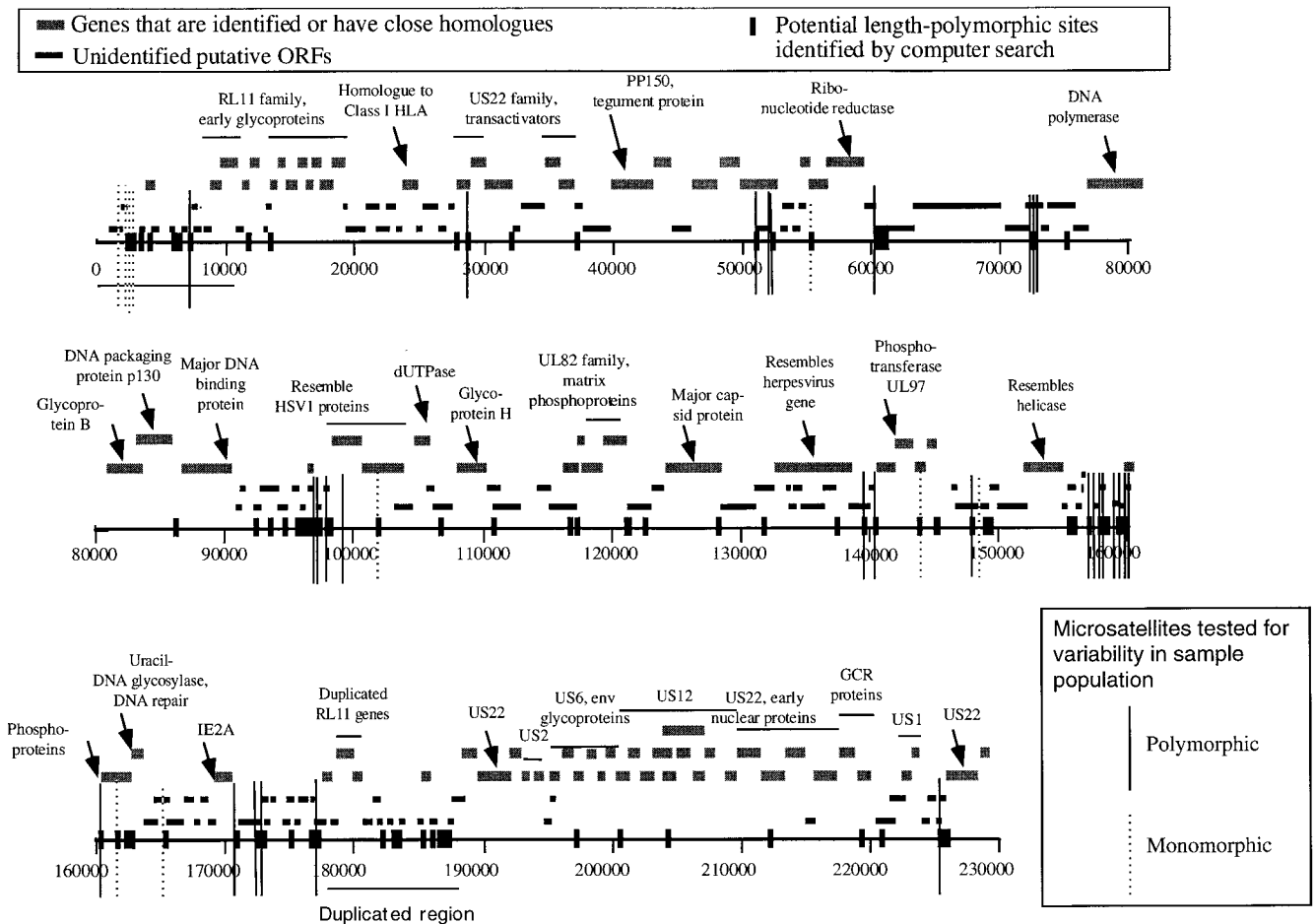


FIG. 1. Diagram of the HCMV AD169 genome showing positions of polymorphic and monomorphic satellites identified in this study. ORFs, open reading frames; HSV1, herpes simplex virus type 1.

I. HCMVUL69C	194	-----	(ggc)5(tgc)2	II. HCMVTRAN3C	339	(tgc)4...(t)10...(tat)3(cat)3
	194	(ggg)2(ggc)3(ttc)2			339	(tgc)3...(t)10...(tat)4(cat)2(tat)
a. 98290	197	(ggg)2(ggc)4(tgc)2		a. 157835	339	(tgc)4...(t)10...(tat)3(cat)2(tat)
	200	(ggg)2(ggc)4(tgc)3			339	(tgc)3...(t)10...(tat)4(cat)2(tat)
b. 5/9	200	(ggg)3(ggc)3(ttc)3		b. 4/10	342	(tgc)4...(t)10...(tat)4(cat)2(tat)*
	203	(ggg)2(ggc)5(tgc)3			342	(tgc)4...(t)10...(tat)4(cat)3
	203	(ggg)2(ggc)4(tgc)4			345	(tgc)4...(t)10...(tat)5(cat)2(tat)
	203	(ggg)3(ggc)4(tgc)3			345	(tgc)4...(t)10...(tat)4(cat)3(tat)
	209	(ggg)5(ggc)4(tgc)3*			345	(tgc)4...(t)10...(tat)5(cat)2(tat)
					346	(tgc)4...(t)11...(tat)3(cat)(tat)(cat)2(tat)
III. HCMVUL122N	328	.....	(tg)0(ta)4	IV. HCMVUL68C	409	(a)6
	330	.....	(tg)0(ta)5*		411	(a)8
a. 170989	330	(tgcg)(tg)0(ta)3		a. 97928	412	(a)9
	330	.... ..	(tg)2(ta)4		413	(a)10
b. 3/5	332	.... ..	(tg)0(ta)6	b. 7	414	(a)11*
					415	(a)12
					417	(a)14

FIG. 2. Length variants from all 12 strains were sequenced at four loci (I to IV), and the short tandem repeats contributing polymorphisms to each unique length variant found are shown. The number at the beginning of each sequence refers to PCR product length. a, location in genome (HEHCMVCG); b, number of variants by length/by sequence divergence. Variants from reference sequence AD169 are marked by asterisks.

microsatellite variation, the insertion-deletion events at these repeats clearly classifies them as microsatellites.

#### ACKNOWLEDGMENTS

This work was supported by grant 1-RO3 AI41370 from the National Institute of Allergy and Infectious Diseases (to C.W.). D.M. is an NSCORT graduate fellow, and D.F. is a Markey graduate fellow.

We are most grateful to Chris Lambros for making this study possible. We thank Frank Huang and Tamarah Westmoreland for assistance.

#### REFERENCES

- Ashley, M., and B. Dow. 1994. The use of microsatellite analysis in population biology: background, methods and potential applications. *EXS* **69**:185-201.
- Barr, J. N., S. P. J. Whelan, and G. W. Wertz. 1997. *cis*-acting signals involved in termination of vesicular stomatitis mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J. Virol.* **71**:8718-8725.
- Black, D. H., and R. Eberle. 1997. Detection and differentiation of primate alpha-herpesviruses by PCR. *J. Vet. Diagn. Investig.* **9**:225-231.
- Bornkamm, G., H. Delius, U. Zimmer, J. Hudewentz, and M. Epstein. 1980. Comparison of Epstein-Barr virus strains of different origin by analysis of the viral DNAs. *J. Virol.* **35**:603-618.
- Caskey, C., A. Pizzuti, Y. H. Fu, R. G. J. Fenwick, and D. L. Nelson. 1992. Triplet repeat mutations in human disease. *Science* **256**:784-789.
- Cha, T., E. Tom, G. Kemble, G. Duke, E. Mocarski, and R. Spaette. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **70**:78-83.
- Chee, M., A. Bankier, S. Beck, R. Bohni, C. Brown, R. Cerny, T. Horsnell, C. Hutchison, T. Kouzarides, J. Martignetti, et al. 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr. Top. Microbiol. Immunol.* **154**:125-169.
- Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin, and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**:3166-3170.
- Djian, P., J. Hancock, and H. Chana. 1996. Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. *Proc. Natl. Acad. Sci. USA* **93**:417-421.
- Duke, G. M., J. E. Osorio, and A. C. Palmberg. 1990. Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly(C) tract. *Nature* **343**:474-476.
- Estoup, A., C. Tailliez, J. M. Cornuet, and M. Solignac. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* **12**:1074-1084.
- Falk, K., J. Gratama, M. Rowe, J. Zou, F. Khanim, L. Young, and M. Oosterveer. 1995. The role of repetitive DNA sequences in the size variation of Epstein-Barr virus (EBV) nuclear antigens, and the identification of different EBV isolates using RFLP and PCR analysis. *J. Gen. Virol.* **76**:779-790.
- Field, D. dfield@ucsd.edu.
- Field, D., D. Metzgar, L. Eggert, R. Rose, and C. Wills. 1996. The use of short, clustered polymorphic microsatellite loci to distinguish strains of *Candida albicans*. *FEMS Lett.* **15**:73-79.
- Field, D., and C. Wills. 1998. Abundant microsatellite polymorphisms in *S. cerevisiae* and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae* results from strong mutation pressures and a variety of selection forces. *Proc. Natl. Acad. Sci. USA* **95**:1647-1652.
- Field, D., and C. Wills. 1996. Long, polymorphic microsatellites in simple organisms. *Proc. R. Soc. Lond.* **263**:209-215.
- Foster, P. L., and J. M. Trimarchi. 1994. Adaptive reversion of a frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. *Science* **265**:407-409.
- Freimer, N. B., and M. Slatkin. 1996. Microsatellites: evolution and mutational processes. Variation in the human genome. *Ciba Found. Symp.* **197**:51-72.
- Garcia-Barreno, B., T. Delgado, and J. A. Melero. 1994. Oligo(A) sequences of human respiratory syncytial virus G protein gene: assessment of their genetic stability in frameshift mutants. *J. Virol.* **68**:5460-5468.
- Garza, J. C., and N. B. Freimer. 1996. Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Res.* **6**:211-217.
- Gergely, L., L. Sternas, J. Dillner, and G. Klein. 1984. Molecular size variation of EBNA is determined by the EB viral genome. *Intervirology* **22**:85-96.
- Given, D., and E. Kieff. 1979. DNA of Epstein-Barr virus. VI. Mapping of the internal tandem reiteration. *J. Virol.* **31**:315-324.
- Hahn, H., and A. C. Palemborg. 1995. Encephalomyocarditis viruses with short poly(C) tracts are more virulent than their mengovirus counterparts. *J. Virol.* **69**:2697-2699.
- Hancock, J. M., W. Chaleeprom, W. Chaleeprom, J. Dale, and A. Gibbs. 1995. Replication slippage in the evolution of potyviruses. *J. Gen. Virol.* **76**:3229-3232.
- Hayward, S., L. Noguee, and G. Hayward. 1980. Organization of repeated regions within the Epstein-Barr virus DNA molecule. *J. Virol.* **33**:507-521.
- Kruglyak, S., R. Durrett, M. Schug, and C. Aquadro. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* **95**:10774-10778.
- Metzgar, D., A. van Belkum, D. Field, R. Haubrich, and C. Wills. 1998. Random amplification of polymorphic DNA and microsatellite genotyping of pre- and posttreatment isolates of *Candida* spp. from human immunodeficiency virus-infected patients on different fluconazole regimens. *J. Clin. Microbiol.* **36**:2308-2313.
- Metzgar, D., D. Field, R. Haubrich, and C. Wills. 1998. Sequence analysis of a compound coding-region microsatellite in *Candida albicans* resolves homoplasy and provides a high-resolution tool for genotyping. *FEMS Immunol. Med. Microbiol.* **20**:103-109.
- Moxon, E., P. Rainey, M. Nowak, and R. Lenski. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**:24-33.
- Moxon, R., and C. Wills. 1999. DNA microsatellites: agents of evolution? *Sci. Am.* **280**:94-99.
- Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616-1622.

31. **Perdue, M. L., M. Garcia, D. Senne, and M. Fraire.** 1997. Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res.* **49**:173–186.
32. **Plotkin, S. A., T. Farukama, N. Zygraich, and C. Huggelen.** 1975. Candidate cytomegalovirus strain for human vaccination. *Infect. Immun.* **12**:521–527.
33. **Rosenberg, S. M., S. Longerich, P. Gee, and R. S. Harris.** 1994. Adaptive mutation by deletions in small mononucleotide repeats. *Science* **265**:405–407.
34. **Rowe, E. A., J. W. Hartley, S. Waterman, H. C. Turner, and R. J. Huebner.** 1956. Cytopathic agent resembling common salivary gland virus recovered from tissue cultures of human adenoids. *Proc. Soc. Exp. Biol. Med.* **92**:418.
35. **Sakaoka, H., K. Kurita, Y. Iida, S. Takada, K. Umene, Y. T. Kim, C. S. Ren, and A. J. Nahmias.** 1994. Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. *J. Gen. Virol.* **75**:513–527.
36. **Schlotterer, C., and D. Tautz.** 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**:211–215.
37. **Sidagis, J., K. Ueno, M. Tokunaga, M. Ohyama, and Y. Eizuru.** 1997. Molecular epidemiology of Epstein-Barr virus (EBV) in EBV-related malignancies. *Int. J. Cancer* **72**:72–76.
38. **Smith, I. L., J. M. Cherrington, R. E. Jiles, M. D. Fuller, W. R. Freeman, and S. A. Spector.** 1997. High-level resistance of cytomegalovirus to ganciclovir is associated with alterations in both UL97 and DNA polymerase genes. *J. Infect. Dis.* **176**:69–77.
39. **Spector, S.** 1996. Spectrum and treatment of cytomegalovirus disease in persons with AIDS. *J. Int. Soc. Phys. AIDS Care* **2**:9–22.
40. **Stewart, C.-B.** 1993. The powers and pitfalls of parsimony. *Nature* **361**:603–617.
41. **Takayama, M., N. Takayama, N. Inoue, and Y. Kameoka.** 1996. Application of long PCR method of identification of variations in nucleotide sequences among varicella-zoster virus isolates. *J. Clin. Microbiol.* **34**:2869–2874.
42. **Tautz, D., and C. Schlotterer.** 1994. Simple sequences. *Curr. Opin. Genet. Dev.* **4**:832–837.
43. **Tran, H., J. Keen, J. Krickler, M. Resnick, and D. Gordenin.** 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell. Biol.* **17**:2859–2865.
44. **Umene, K., R. Watson, and L. Enquist.** 1984. Tandem repeated DNA in an intergenic region of *Herpes simplex* virus type 1 (Patton). *Gene* **30**:33–39.
45. **Umene, K., and M. Yoshida.** 1993. Genomic characterization of two predominant genotypes of herpes simplex virus type 1. *Arch. Virol.* **131**:29–46.
46. **Vaisanen, M., R. Haataja, and J. Leisti.** 1996. Decrease in the CGG<sub>n</sub> trinucleotide repeat mutation of the fragile X syndrome to normal size range during parental transmission. *Am. J. Hum. Genet.* **59**:540–546.
47. **Valdes, A., M. Slatkin, and N. Freimer.** 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**:737–749.
48. **Weber, J. L.** 1990. Informativeness of human (dC-dA)<sub>n</sub>(dG-dT)<sub>n</sub> polymorphisms. *Genomics* **524**:524–530.
49. **Webster, R. G., W. Bean, O. T. Gorman, and Y. Kawaoka.** 1992. Ecology of avian influenza viruses. *Microbiol. Rev.* **56**:151–179.
50. **Yamada, N., K. Tanihara, A. Takada, T. Yoriyuzi, M. Tsitsumi, H. Shimomura, T. Tsuji, and T. Date.** 1996. Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology* **223**:255–261.