

Characterization of Human Cytomegalovirus Strains by Analysis of Short Tandem Repeat Polymorphisms

ADRIANNE WALKER,¹ SUSAN J. PETHERAM,² LINDA BALLARD,¹ JODY R. MURPH,³
GAIL J. DEMMLER,⁴ AND JAMES F. BALE, JR.^{2,5*}

Departments of Pediatrics² and Neurology,⁵ University of Utah School of Medicine,¹ University of Utah, Salt Lake City, Utah; Department of Pediatrics, University of Iowa, Iowa City, Iowa³; and Department of Pediatrics, Baylor College of Medicine, Houston, Texas⁴

Received 28 November 2000/Returned for modification 16 January 2001/Accepted 12 March 2001

Human cytomegalovirus (HCMV) strains display genetic polymorphisms, and these polymorphisms can be analyzed to study viral transmission and pathogenesis. Recently, short tandem repeat (STR) length polymorphisms have been identified in the HCMV genome. We assessed the utility of STRs in characterizing HCMV strains and found that a multiplexed PCR assay using primers based upon these STRs accurately maps HCMV strains. Using primers for 10 microsatellite regions, the STR profiles of 44 wild-type and 2 laboratory strains of HCMV were characterized. The results of STR analysis were compared with those for strain characterization using nucleotide sequencing and restriction fragment length polymorphism analysis. In each instance, STR analysis accurately and specifically identified strains that were indistinguishable or distinct by conventional molecular analysis. Analysis of short tandem repeats also detected polymorphisms that supported simultaneous excretion of two HCMV strains. These results indicate that STR analysis allows rapid, precise molecular characterization of HCMV strains.

Human cytomegalovirus (HCMV), a betaherpesvirus, has a large, complex genome consisting of approximately 240,000 nucleotide base pairs (bp) and over 200 open reading frames (20). Wild-type and laboratory strains of HCMV display nucleotide polymorphisms in several regions of the genome, especially in the *a* sequence, major immediate early (MIE), glycoprotein B (gB; UL55), and UL144 gene regions (7, 8, 12, 16–18; J. F. Bale, S. J. Petheram, M. Robertson, J. R. Murph, and G. Demmler, submitted for publication). Although the effects of these polymorphisms on the biology of HCMV infections are largely unknown, molecular variations can be analyzed to study the epidemiology and pathogenesis of HCMV infections (4, 5, 16).

Mapping the molecular profiles of HCMV has enabled investigators to determine the source and transmission patterns of HCMV infections. Studies of HCMV epidemiology have mapped strains by using analysis of the restriction fragment length polymorphisms (RFLPs) of the entire HCMV genome (1, 10), PCR-based analysis of *a* sequence variations (2, 22), DNA sequence analysis of polymorphic regions (11, 19; Bale et al., submitted), and stepwise analysis of the RFLPs of gene regions by using PCR-based methods (3, 7, 15, 18). Our laboratory has previously used a PCR-based algorithm that compares the size and RFLPs of the *a* sequence amplicon, the gB genotype as described by Chou and Dennison (8), and the RFLPs of an amplicon derived by amplification of the MIE gene region (18). Although these molecular approaches can

characterize HCMV strains accurately, they are time-consuming and available only in research laboratories.

Recently, Davis and colleagues reported that the HCMV genome contains numerous short tandem repeats (STRs) (9). The microsatellite regions consist of iterated motifs of one to six bases that occur in the genomes of eukaryotes and some prokaryotes and represent potential sites of mutation (14, 21). Davis and colleagues found at least 24 regions in the HCMV genome that exhibit length or sequence polymorphisms and suggested that analysis of certain polymorphic regions could enable laboratories to compare HCMV strains (9).

In this report, we confirm that analysis of STR length polymorphisms can be used to characterize HCMV strains. Forty-six HCMV strains, comprising 44 wild-type strains and 2 laboratory strains, Towne and AD169, were compared by PCR-based STR analysis using primers for 10 microsatellite regions. In each instance, STR analysis was concordant with the results of conventional analyses of strains based on DNA sequences and RFLPs. These results indicate that STR analysis provides a novel, effective means to compare the molecular profiles of HCMV strains.

MATERIALS AND METHODS

Virus strains. Forty-six HCMV strains, consisting of 2 laboratory strains, Towne and AD169, and 44 wild-type strains, isolated from 28 children and the mother of one congenitally infected child, were analyzed. Wild-type strains were collected in Iowa and Texas between 1986 and 1995 and were isolated on human foreskin fibroblast cells grown in shell vials or 24-well plates using methods described previously (3, 4, 16). Collection of HCMV strains was approved by the institutional review boards of the respective institutions, the University of Iowa and Baylor College of Medicine.

DNAs from wild-type strains were extracted from low-passage (<4) human foreskin fibroblast cell cultures according to methods published previously (3, 4; Bale et al., submitted) and were stored at 4°C in distilled water until analyzed. DNA from the Towne strain of HCMV was provided by Mark Stinski, University

* Corresponding author. Mailing address: Division of Pediatric Neurology, Suite 2700, Primary Children's Medical Center, 100 N. Medical Dr., Salt Lake City, UT 84113. Phone: (801) 588-3385. Fax: (801) 588-3392. E-mail: pcjbale@ihc.com.

TABLE 1. PCR primers used for STR analysis

Region ^a	Primer sequence ^b
UL69C-F6FAM-GCTTAACTTGATGACGCCGTCGCC
UL69C-RGTTTCTCGTAATCGGGCGTCTGGCAGCG
UL123I1N-F6FAM-CCTCCACCGTTAACAGCACC
UL123I1N-RGTTTCTGACTCATGGTCGCTCGGCAG
UL50C-F6FAM-GCGGATGACGGTGATGGGTCCG
UL50C-RGTTTCTGCGGAAAGCGGTCCTCTCGG
UL68C-F6FAM-AGGGAGAGGACGTGGGGCTCG
UL68C-RGTTTCTGCCAGACGCCGATTACGAGG
UL46C-FHEX-CTCGCGTTTGGCTTTGAGC
UL46C-RGTTTCTCCCGCAGCTGCTCTATCAAC
TRL8C-FHEX-CCCAGACGAGAGCCACCC
TRL8C-RGTTTCTATCCTCGCCTTCGGACGCC
TRAN3C-FHEX-CCTCTCCTCCAGTGTAGTCGTG
TRAN3C-RGTTTCTCCTCCTCTCCCCATCATCTTCTCC
UL38N-FHEX-GCAGCATGTGGGGCTAATAGG
UL38N-RGTTTCTGCTGCGAGCCAATTCGTTGG
UL23C-FTET-GTCGCGACGTCTTAGGAGG
UL23C-RGTTTCTTCCCGTTTACTCGTGC
UL122N-FTET-GGCTGAGAACAGTGATCAGG
UL122N-RGTTTCTGCGTGACACGTTTATTGAGTAGG
UL95C-FTET-GCTGAGCTGTACGTCTTTGTTGG
UL95C-RGTTTCTGCGAGTCTTGGATAACGATGGGG
UL111C-FTET-CCATGATCAATTAAGCCCACCACC
UL111C-RGTTTCTCGTAATCCTCTGGACGACACTGCG

^a F, forward; R, reverse; UL, unique long; TR, terminal repeat.

^b 6FAM, 6-carboxyfluorescein; HEX, hexachloro-6-carboxyfluorescein; TET, tetrachloro-6-carboxyfluorescein.

of Iowa College of Medicine, Iowa City. The HCMV AD169 DNA was purchased from Advanced Biotechnologies, Inc., Columbia, Md.

Wild-type strains were selected from strains stored in the laboratories of J. F. Bale and G. Demmler. The selection allowed comparison of (i) strains from epidemiologically unrelated individuals; (ii) strains from individuals with known horizontal or vertical transmission (for example, children from the same child care center and a mother-infant pair); (iii) sequential isolations from an individual known to have shed a single strain; (iv) sequential isolations from individuals known to have shed strains that were distinct (i.e., reinfections); and (v) duplicate strains from the same individual.

PCR primers. Twelve STR markers (Table 1; Fig. 1) were chosen from the group of 24 locations known to be polymorphic in HCMV (9). The primers were

selected because they appeared to identify the most informative length polymorphisms and could be multiplexed into a single gel lane.

All primers were synthesized by Integrated DNA Technologies, Inc., Coralville, Iowa. The forward primers were modified from the published sequences by the 5' addition of a fluorescent tag (6-carboxyfluorescein, hexachloro-6-carboxyfluorescein, or tetrachloro-6-carboxyfluorescein). The reverse primers were modified by the 5' addition of a 6-bp nonspecific tail, as described previously (6, 12). Additional modifications were made on some primers to lower the percentage of GC content or to modify the annealing temperature.

PCR amplification and fragment analysis. Each locus was amplified in a 20- μ l PCR mixture using 0.5 μ l of unquantified target HCMV DNA. Reaction mixtures included 200 μ M concentrations of each deoxynucleoside triphosphate, 10 mM Tris-HCl (pH 8.3), 40 mM NaCl, 1.5 mM MgCl₂, 0.2 mM spermidine-HCl, 0.25 U of *Taq* Platinum (Life Technologies, Gaithersburg, Md.), and 0.5 μ M concentrations of each primer. Reaction mixtures for UL50C, UL38N, and UL95C also included 5% dimethyl sulfoxide.

Thermocycling was performed on a PTC-200 or -225 instrument (MJ Research, Inc., Watertown, Mass.). For most markers, samples were denatured for 5 min at 94°C followed by 30 cycles at 94°C for 20 s, 62°C for 20 s, and 72°C for 40 s and a final extension at 72°C for 10 min. The PCR annealing temperatures were 55°C for UL95C, 68°C for UL50C and UL38N, and 62°C for the remaining primer sets.

All samples were electrophoresed on a 5% Long Ranger acrylamide gel (BMA, Rockland, Maine) using the ABI Prism 377 DNA sequencer with ABI Prism 377-96 Collection version 2.6 software (Applied Biosystems, Foster City, Calif.). A 20-bp size standard was used (Genotype TAMRA 50-500 DNA Ladder; Life Technologies). Fragment sizing was performed with GeneScan version 3.1 (Applied Biosystems). Two previously sized controls were run on each gel to ensure consistent scoring of amplicons. STR variants were assigned a numerical score based upon the relative sizes of the repeats for each region.

Conventional genotyping and sequence analysis. HCMV strains studied by STR were analyzed by PCR-based methods and primers for the *a* sequence, gB, MIE, and UL144 HCMV gene regions using methods described previously (3, 15, 18). Strains were considered distinct if they displayed differences in the *a* sequence amplicon size, gB genotype, or MIE RFLP, or had <95% homology by UL144 or *a*-sequence nucleotide DNA sequences. Conversely, strains were considered indistinguishable if they displayed identical patterns for *a* sequence amplicon size, gB genotype, and MIE RFLP or had identical gB genotypes and \geq 95% nucleotide sequence homology for *a* sequence and UL144 gene regions.

Analysis of STR results and comparison with conventional molecular strain analysis. Strains were coded, and STR analysis was conducted by persons unaware of the relationships of HCMV strains. Upon completion of STR analysis, strains were identified as distinct or indistinguishable by the STR method, and these results were compared with the results of conventional molecular analysis and the epidemiologic information regarding each strain.

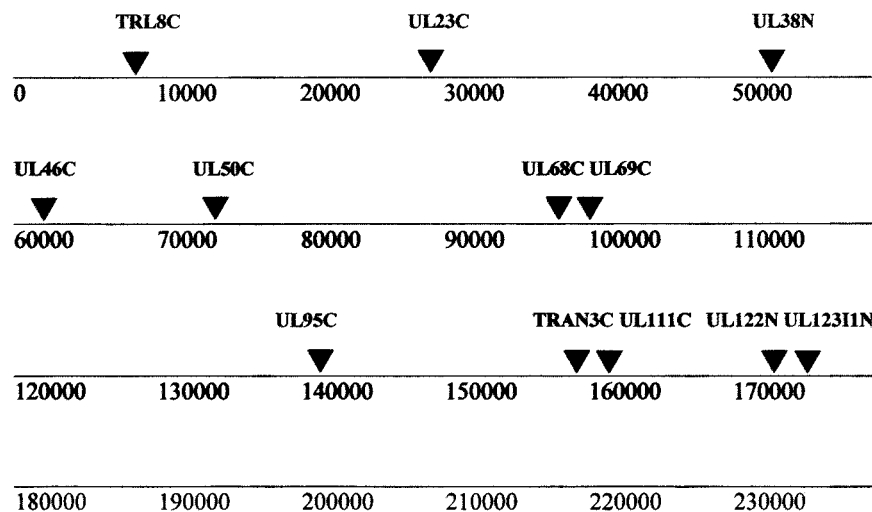


FIG. 1. Linear map of the HCMV genome showing the relative sites of the STR loci analyzed in this study; adapted from data of Davis et al. (9).

TABLE 2. Summary of STR loci

HCMV STR locus	Repeat sequence	Reported size variants	Observed size variants	Fragment size range (bp)
UL69C	(GGT)5(GGC)4(TGC)3	5	6	191–206
UL123I1N	(TGG)4	3	3	262–268
UL50C	(CTC)4(GGC)4(GAG)4	3	5	351–369
UL68C	(T)11	7	10	402–418
UL46C	(CG)5	2	1	144
TRL8C	(GGC)4	2	2	252–255
TRAN3C	(TGC)4(T)10(TAT)4	4	5	347–355
UL38N	(A)10	3	Ambiguous	402–407
UL23C	(A)10	4	4	232–235
UL122N	(AT)5	3	6	323–345
UL95C	(GGT)5	4	11	385–447
UL111C	(GGT)4(A)10(A)10(CGA)4	4	12	435–451

RESULTS

STR polymorphisms. Of the 12 regions identified as showing potential length polymorphisms for the STRs of HCMV, 10 yielded results that were useful for strain comparison (Table 2). One region (UL46C) showed no polymorphisms among the 46 strains analyzed, and an additional region (UL38N) proved difficult to amplify and exhibited anomalous fragment mobility. The methods used in this analysis had high sensitivity, enabling detection of 1 to 3 bp differences between HCMV strains.

Among the remaining regions, PCR analysis detected between 2 and 12 size polymorphisms, depending upon the region (Table 2; Fig. 2). For each region, we observed at least the numbers of length polymorphisms identified by Davis et al. (9), and in two regions (UL95C and UL111C) several additional size polymorphisms were detected (Tables 2 and 3). Each region had a dominant allelic pattern (Table 3).

Comparison of HCMV strains. All strains yielded sufficient PCR product to allow comparison of STR patterns. Three

strains failed to amplify for one region each, but the data for the remaining regions were adequate to allow valid characterization of each of these strains. Distinct STR patterns typically differed between strains at 5 to 7 sites of potential polymorphism. The AD169 and Towne strains showed distinct STR patterns and differed from each other as well as each of the wild-type strains. AD169 and isolate TX02 displayed the closest patterns of any distinct isolates, differing in 3 of 10 polymorphic regions.

We first compared the STR patterns of strains that were known or predicted to be indistinguishable (Table 4). These included (i) serial isolates shed by individual children shown by conventional molecular analysis to be the same strain (isolates 311A and 311B); (ii) duplicate isolates from the same individual [isolates 501 and 501A, and TX11(1) and TX11(2)]; (iii) several isolates from a large Cedar Rapids, Iowa, child care center (isolates 902, 916A, 90003, 90022B, and 90064); and (iv) isolates from a congenitally infected infant and the mother

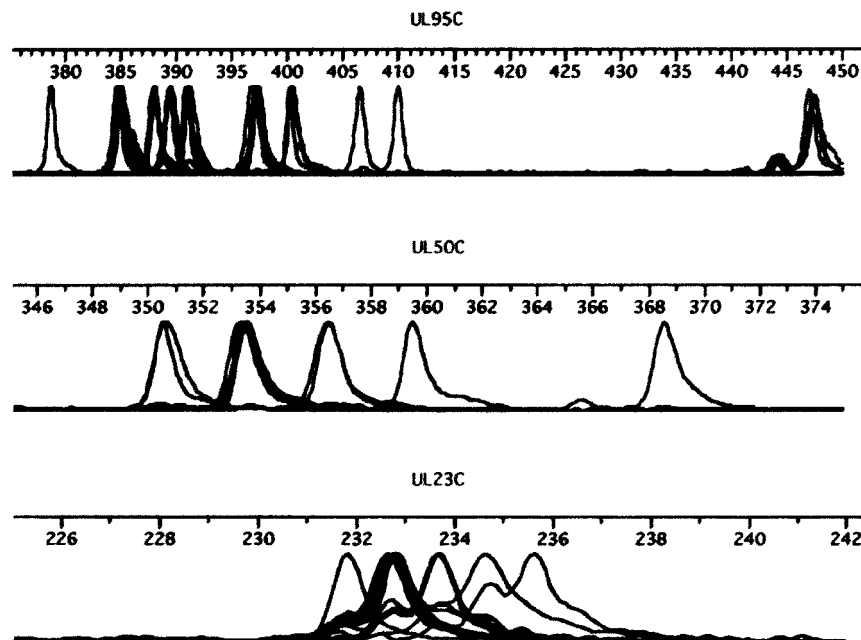


FIG. 2. Stacked electropherograms showing the size polymorphisms among HCMV strains for three of the polymorphic STR regions (UL95C, UL50C, and UL23C), illustrating the data used to generate an HCMV STR map.

TABLE 3. Summary of STR alleles at the 10 loci used to compare strains^a

Allele pattern no.	% of isolates with allele at locus:									
	TRL8C	UL23C	UL122N	UL111C	UL69C	UL12311N	UL50C	UL68C	TRAN3C	UL95C
1	76	27	48	24	24	53	23	22	22	2
2	24	47	15	4	48	36	55	24	4	2
3		7	2	18	15	11	17	4	67	9
4		18	22	2	2		2	7	2	22
5		2	7	11	2		2	20	4	2
6			7	2	9			2		9
7				7				9		4
8				13				9		9
9				2				2		37
10				9				2		2
11				2						2
12				4						

^a Strains were assigned a numerical designation based upon the relative size of the repeats, i.e., the STR pattern, as they were identified. The first column (allele pattern) indicates the numerical designation for alleles. The numbers in remaining columns indicate the percentage of isolates that possessed the allele designation at the respective locus. For example, the TRL8C locus displayed two allelic variations, and 76% of the isolates displayed pattern 1 and the remainder displayed pattern 2. Similarly, UL111C had 12 allelic variants, and 24% of the isolates conformed to pattern 1, the most common pattern for this locus, and 11 other patterns were observed.

(isolates 40060000 and 40060066). In each instance, the STR patterns were identical among the strains, as predicted, but the STR patterns allowed differentiation of all groups of strains from each other group. Isolate 501A did not amplify at the UL122N locus but was identical to isolate 501 at the other nine loci.

Data for isolates 311A and 311B illustrated the potential power of STR analysis (Tables 4, 5, and 6). These strains were isolated 16 months apart from a single child attending an Iowa City child care center. By previous analyses in three regions, the *a* sequence amplicon size, gB genotype, and MIE RFLP, these strains had been deemed indistinguishable. The present STR analysis showed identical polymorphisms for these strains in each of the 10 regions studied, confirming that these isolates were indeed the same strain. These results also indicated that the STR polymorphisms remained stable during prolonged excretion by an individual.

We next compared isolates known to be genetically or epi-

demologically distinct (Tables 5 and 7). These isolates included (i) serial isolates from individual children known to be reinfected with new HCMV strains (191A, 191B, and 191C; 916A and 916B; 90022A and 90022B; and TX11 and TX12 [Table 7] and isolates 311A, 311B, and 311C [Table 5]); and (ii) isolates from children attending different child care centers in the Iowa City-Cedar Rapids region (isolates coded with 300s [an Iowa City center] versus isolates coded with 900s and 90000s [a Cedar Rapids Center]). The latter groups of isolates have been studied extensively and found to conform to distinct clades by sequence analysis of the *a* sequence and UL144 gene regions (Bale et al., submitted). STR analysis correctly classified each isolate, without exception.

Data for strain 501, isolated from a child attending an Iowa City child care center, illustrated further the sensitivity of the STR method (Fig. 3). This isolate possessed 100% nucleotide sequence homology for the *a* sequence region compared with isolates from a different Iowa City child care center but dif-

TABLE 4. HCMV isolates predicted to be indistinguishable

Isolate	Date	Allelic pattern no. displayed at locus:									
		TRL8C	UL23C	UL122N	UL111C	UL69C	UL12311N	UL50C	UL68C	TRAN3C	UL95C
311A	5/13/86	1	4	6	7	2	2	3	1	3	9
311B	9/15/87	1	4	6	7	2	2	3	1	3	9
500	7/2/93	2	2	5	10	3	2	2	8	3	3
501	1/13/92	2	2	5	10	3	2	2	8	3	3
501A	1/13/92	2	2	0 ^a	10	3	2	2	8	3	3
TX11(1)	10/16/91	1	2	4	1	2	1	2	2	3	9
TX11(2)	10/16/91	1	2	4	1	2	1	2	2	3	9
902	5/31/91	1	1	1	3	1	2	1	5	3	4
916A	2/21/91	1	1	1	3	1	2	1	5	3	4
90003	1/19/93	1	1	1	3	1	2	1	5	3	4
90022B	4/20/94	1	1	1	3	1	2	1	5	3	4
90064	4/21/94	1	1	1	3	1	2	1	5	3	4
40060000		1	1	1	3	1	2	1	5	3	4
40060066		1	1	1	3	1	2	1	5	3	4
40032800	4/26/91	1	1	1	3	1	2	2, 1 ^b	5	3 ^c	4

^a Failure to amplify is indicated by a zero.

^b Locus with two numbers indicates detection of additional alleles.

^c Possible second allele.

TABLE 5. STR analysis results for three isolates from a child attending group child care^a

Isolate	Allelic pattern no. displayed at locus:									
	TRL8C	UL23C	UL122N	UL111C	UL69C	UL123I1N	UL50C	UL68C	TRAN3C	UL95C
311A	1	4	6	7	2	2	3	1	3	9
311B	1	4	6	7	2	2	3	1	3	9
311C	1	2	4	1	2	1	2	2	3	9

^a Isolates 311A and 311B, isolated from the same child 16 months apart, possessed identical STR allelic patterns at all loci. By contrast, 311C, isolated from the same child 22 months after 311B, differed from A and B at 6 loci, indicating that 311C was a distinct strain. The 311 C STR pattern was observed in an isolate from another child attending the same center.

ferred from these strains by a 4-nucleotide deletion in the UL144 sequence (Bale et al., submitted). When compared using STR analysis, strain 501 differed substantially from the strains in the other center but possessed an STR profile identical to that of a strain isolated from another child (isolate 500) attending the same center (Table 4).

We next compared isolates that were predicted to be similar based on geographic proximity (Table 4). These consisted of isolates from children attending a Cedar Rapids center (isolates 902, 916A, 90003, 90022B, and 90064) and isolates from congenitally infected children (isolates 4006000 and 4003280); the latter isolates were suspected by prior molecular analysis to be linked to the Cedar Rapids center (15). Each of the congenitally infected infants was born in Cedar Rapids and had siblings who attended the Cedar Rapids center during their mother's pregnancy. STR analysis showed that infant 40060000 shed an isolate that was indistinguishable from those in the Cedar Rapids center, and isolate 40032800 differed from the center strains at only one locus.

Finally, we compared the STR patterns of the Iowa and Texas isolates (Table 8). This analysis indicated that several Texas isolates had patterns that were distinct and different from all other Texas or Iowa isolates (isolates TX02, TX12, TX17, and TX23). In contrast, certain Texas isolates (TX08, TX11, and TX25) possessed STR patterns that were indistinguishable from a large Iowa City cluster (isolates 130, 216, 20009, 311C, and 375). We were not aware of any direct contact between these children or their parents.

Detection of simultaneous excretion of HCMV strains. Prior conventional analysis of strains 191A, 191B, and 191C yielded unusual results, suggesting genetic recombination or simultaneous shedding of multiple HCMV strains (3). In the present study these strains yielded STR patterns that suggested simultaneous excretion of two strains. Strains 191A and 191B each showed two STR polymorphisms at one or more loci, including

regions UL122N, UL123I1N, and UL95C (Table 7). The most plausible explanation for the appearance of two alleles at each of these loci is the presence of two genetic variants, i.e., simultaneous shedding of two strains.

DISCUSSION

This report demonstrates that STR analysis accurately characterizes HCMV strain patterns. PCR analysis using primers for 10 microsatellite regions in the HCMV genome correctly identified HCMV strains shown to be distinct or indistinguishable by conventional molecular analysis. When allelic patterns were compared, distinct HCMV strains showed differences at three or more STR loci. With only one exception, indistinguishable strains had identical allelic patterns at all loci. The sensitivity of this method detects differences as small as 1 bp in STR patterns for each gene region. By analyzing multiple regions throughout the genome, STR analysis allows comprehensive comparisons of HCMV strains.

STR analysis has distinct advantages over prior molecular approaches to HCMV strain characterization. Current strategies, such as the one routinely used in our laboratory (18), require stepwise analysis of several gene regions for definitive comparisons of HCMV strains. Even though certain regions, e.g., the *a* sequence and UL144, are hypervariable, these regions have dominant genotypes, indicating that identical patterns can occur by chance if limited numbers of HCMV gene regions are examined. In contrast, multiplexed STR analysis allows simultaneous evaluation of several polymorphic regions, thus increasing efficiency of the analysis and reducing substantially the probability of chance associations.

STR also appears to facilitate detection of simultaneous excretion of distinct HCMV strains. Prior methods, such as genotyping of strains by analyzing the polymorphisms observed at the gB cleavage site (8, 13), have enabled investigators to identify mixed genotypes, which suggest excretion of two strains. Detecting multiple alleles at STR loci, however, provides convincing evidence of genetic allelism, a phenomenon possible only when two or more genetic variants, i.e., distinct HCMV strains, are present in a single sample.

A remarkable finding of this study was the observation that several Texas strains appeared to be identical to strains from Iowa City, Iowa. This observation was confirmed by nucleotide sequencing of *a* sequence and UL144 regions and by RFLP analysis of gB (Bale et al., submitted). We are unaware of any point source acquisition for these strains or any direct contact between the Iowa children and the parents of the Texas infants. This observation implies that the identical HCMV

TABLE 6. Conventional molecular analysis results for three isolates from a child attending group child care^a

Isolate	<i>a</i> sequence size	gB genotype	MIE RFLP
311A	J	1	XII
311B	J	1	XII
311C	C	1	X

^a Isolates 311A and 311B were indistinguishable by conventional molecular analysis. The isolates possessed the same *a* sequence amplicon size, identical gB genotype, and identical RFLPs for the MIE amplicon. By contrast, isolate 311C differed from A and B in both the *a* sequence size and the MIE RFLP, implying that 311C was a distinct strain.

TABLE 7. Genetically and epidemiologically distinct HCMV isolates

Isolate	Date	Allelic pattern no. displayed at locus:									
		TRL8C	UL23C	UL122N	UL111C	UL69C	UL123IIN	UL50C	UL68C	TRAN3C	UL95C
191A	6/12/86	1	2	2	9	6	3,1 ^a	4	7	5	9,1 ^a
191B	7/30/86	1	4	6,1 ^a	8	2	3 ^b	2 ^b	1	1 ^b	9
191C	1/9/87	1	4	1	8	2	3	2	1	1	9
916A	2/21/91	1	1	1	3	1	2	1	5	3	4
916B	9/30/91	1	2	1	4	5	2	2	1	4	2
90022A	1/26/94	1	4	1	7	2	1	3	7	3	8
90022B	4/20/94	1	1	1	3	1	2	1	5	3	4
TX11(1)	10/16/91	1	2	4	1	2	1	2	2	3	9
TX12	7/13/94	1	2	1	6	4	1	2	1	1	8
375	2/20/91	1	2	4	1	2	1	2	2	3	9
311C	6/14/89	1	2	4	1	2	1	2	2	3	9
902	5/31/91	1	1	1	3	1	2	1	5	3	4
916A	2/21/91	1	1	1	3	1	2	1	5	3	4
90003	1/19/93	1	1	1	3	1	2	1	5	3	4
90022B	4/20/94	1	1	1	3	1	2	1	5	3	4
90064	4/21/94	1	1	1	3	1	2	1	5	3	4
AD169		1	2	1	5	1	1	3	7	3	6
TOWNE		2	5	1	2	2	1	2	2	1	11

^a Loci with two numbers indicate detection of additional alleles.

^b Possible second allele.

strains had a common ancestral origin. This conclusion requires that the STR patterns do not mutate during HCMV excretion and after transmission between individuals.

Although the STRs of eukaryotes have relatively high rates of mutation (20), the present data suggest that the HCMV STRs remain reasonably stable. This conclusion is supported by the identification of identical STR polymorphisms in (i) strains isolated 16 months apart from the same individual, (ii) strains isolated from children who had epidemiologic evidence

for horizontal transmission (e.g., the isolates from the Cedar Rapids child care center), and (iii) a vertically transmitted strain from a mother-infant pair.

We selected the 12 primers according to the number of predicted size polymorphisms and the size of the predicted amplicons (9). By carefully modifying primers and selecting fluorescent tags, we were able to create an efficient multiplex of amplicons separated by size and fluorescent tags. One of the selected regions (UL46C) yielded no polymorphisms, and an

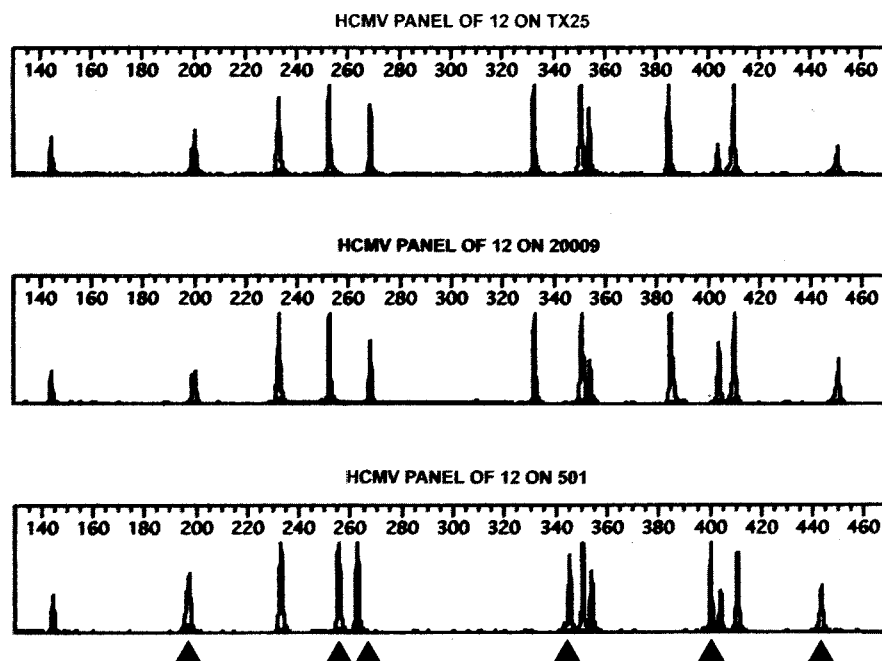


FIG. 3. STR polymorphisms for three strains (TX25, 20009, and 501). Bands represent the size polymorphism for each of the STR loci, as described in Materials and Methods. The numbers 140 to 460 indicate the base pair size of PCR amplicons. Strains TX25 and 20009 display identical length polymorphisms for all loci. By contrast, strain 501 differs from the above strains at several loci (arrowheads).

TABLE 8. Comparison of HCMV isolates from Iowa and Texas

Isolate	Date	Allelic pattern no. displayed at locus:									
		TRL8C	UL23C	UL122N	UL111C	UL69C	UL12311N	UL50C	UL68C	TRAN3C	UL95C
TX02	9/13/90	1	2	1	5	3	1	3	6	3	5
TX12	7/13/94	1	2	1	6	4	1	2	1	1	8
TX17	10/30/92	2	1	1	5	6	1	2	1	3	6
TX23	6/8/94	0	2	1	12	2	1	5	9	1	8
TX08	1/6/94	1	2	4	1	2	1	2	2	3	9
TX11	10/16/91	1	2	4	1	2	1	2	2	3	9
TX25	11/14/94	1	2	4	1	2	1	2	2	3	9
130	2/11/86	1	2	4	1	2	1	2	2	3	9
216	5/29/91	1	2	4	1	2	1	2	2	3	9
20009	1/11/94	1	2	4	1	2	1	2	2	3	9
311C	6/14/89	1	2	4	1	2	1	2	2	3	9
375	2/20/91	1	2	4	1	2	1	2	2	3	9

additional region gave ambiguous results. These regions were dropped from the final analysis without affecting the ability to discriminate HCMV strains.

Because the amplified loci include mononucleotide repeats as well as di- and trinucleotide repeats, it was important to be able to repeatably score single base pair differences in amplicons exceeding 300 bp. To ensure crisp peaks and avoid ambiguity in markers with mono- or dinucleotide repeat polymorphisms, the reverse primers were synthesized with a 6-bp tail known to promote nontemplated nucleotide addition by *Taq* DNA polymerase (6, 12).

To promote accuracy and repeatability in DNA fragment sizing, we used an internal size standard with a 20-bp ladder. We had previously determined that a 10-bp size standard does not improve the accuracy of size determination or reproducibility compared with the 20-bp ladder (data not shown). Since the repeat regions were typically short (3 to 11 repeats), we observed few or no stutter bands of descending size, as can be seen in polymorphic repeat regions of greater length.

From a time and cost-management perspective, STR analysis is very efficient. The multiplex of 10 amplicons pooled for electrophoresis results in a turnaround time of less than 24 h, including PCR, gel electrophoresis, and scoring of alleles. The assay requires a specialized fluorescence detection platform, either slab gel or capillary, which is increasingly available in resource facilities providing DNA sequencing or DNA fragment analysis services. Although there is a relatively high initial cost in purchasing fluorescent-tagged primers (approximately \$100 per pair), the amortized cost per assay is low since a single 100-nmol synthesis should be sufficient for approximately 8,000 assays. The Genomics Facility of the University of Utah calculates its costs at \$1.20/genotype or \$12.00 per HCMV strain.

In summary, these results confirm that wild-type HCMV strains possess STRs and that analysis of STR size polymorphisms accurately characterizes HCMV strains. The STR length polymorphisms appear to remain stable during prolonged excretion by an individual or after transmission to other individuals, a finding that has important implications for studies of the transmission and evolution of HCMV strains. STR analysis also allows identification of simultaneous excretion of multiple HCMV strains. Consequently, PCR-based analysis of STR size polymorphisms provides a robust method for comparing HCMV strains.

ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health grant HD22136 and the Primary Children's Medical Center Foundation.

REFERENCES

- Adler, S. P. 1988. Molecular epidemiology of cytomegalovirus: viral transmission among children attending a day care center, their parents, and caretakers. *J. Pediatr.* **112**:366-372.
- Bale, J. F., M. E. O'Neill, S. S. Fowler, and J. R. Murph. 1993. Analysis of acquired cytomegalovirus infections by polymerase chain reaction. *J. Clin. Microbiol.* **31**:2433-2438.
- Bale, J. F., Jr., S. J. Petheram, I. E. Souza, and J. R. Murph. 1996. Cytomegalovirus reinfection in young children. *J. Pediatr.* **128**:347-352.
- Bale, J. F., Jr., J. R. Murph, G. J. Demmler, J. Dawson, J. E. Miller, and S. J. Petheram. 2000. Intrauterine cytomegalovirus infection and glycoprotein B genotypes. *J. Infect. Dis.* **182**:933-936.
- Bongarts, A., D. von Laer, C. Vogelberg, K. Ebert, J. van Lunzen, J. Garep, P. Vaith, F. T. Hufert, O. Haller, and U. Meyer-Konig. 1996. Glycoprotein B genotypes of human cytomegalovirus: distribution in HIV-infected patients. *Scand. J. Infect. Dis.* **28**:447-449.
- Brownstein, M. J., J. D. Carpten, and J. R. Smith. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**:1004-1010.
- Chou, S. 1990. Differentiation of cytomegalovirus strains by restriction analysis of DNA sequences amplified from clinical specimens. *J. Infect. Dis.* **162**:738-742.
- Chou, S., and K. M. Dennison. 1991. Analysis of interstrain variation in cytomegalovirus glycoprotein B sequences encoding neutralization related epitopes. *J. Infect. Dis.* **163**:1220-1234.
- Davis, C. L., D. Field, D. Metzgar, R. Saiz, P. A. Morin, I. L. Smith, S. A. Spector, and C. Wills. 1999. Numerous length polymorphisms at short tandem repeats in human cytomegalovirus. *J. Virol.* **73**:6265-6270.
- Huang, E. S., C. A. Alford, D. W. Reynolds, S. Stagno, and R. Pass. 1980. Molecular epidemiology of cytomegalovirus infections in women and their infants. *N. Engl. J. Med.* **303**:958-962.
- Lurain, N. S., K. S. Kapell, D. D. Huang, J. A. Short, J. Paintsil, E. Winkfield, C. A. Benedict, C. F. Ware, and J. W. Bremer. 1999. Human cytomegalovirus UL144 open reading frame: sequence hypervariability in low-passage clinical isolates. *J. Virol.* **73**:10040-10050.
- Magnuson, V. L., D. S. Ally, S. J. Nylund, Z. E. Karanjawala, J. B. Rayman, J. I. Knapp, A. L. Lowe, S. Ghosh, and F. S. Collins. 1996. Substrate nucleotide-determined non-templated addition of adenine by *Taq* DNA polymerase: implications for PCR-based genotyping and cloning. *BioTechniques* **21**:700-709.
- Meyer-Konig, U., M. Haberland, D. von Laer, O. Haller, and F. T. Hufert. 1998. Intra-genetic variability of human cytomegalovirus glycoprotein B in clinical strains. *J. Infect. Dis.* **177**:1162-1169.
- Moxon, R., and C. Willis. 1999. DNA microsatellites: agents of evolution? *Sci. Am.* **280**:94-99.
- Murph, J. R., I. E. Souza, J. D. Dawson, P. Benson, S. J. Petheram, D. Pfab, A. Gregg, M. E. O'Neill, B. Zimmerman, and J. F. Bale, Jr. 1998. Epidemiology of congenital cytomegalovirus infection: maternal risk factors and molecular analysis of cytomegalovirus strains. *Am. J. Epidemiol.* **147**:940-947.
- Murph, J. R., J. C. Baron, K. C. Brown, C. L. Ebelhack, and J. F. Bale, Jr. 1991. The occupational risk of cytomegalovirus infection among day-care workers. *JAMA* **265**:603-608.

17. **Rasmussen, L., C. Hong, Zipeto, S. Morris, D. Sherman, S. Chou, R. Miner, W. L. Drew, R. Wolitz, A. Dowling, A. Warford, and T. C. Merigan.** 1997. Cytomegalovirus gB genotype distribution differs in human immunodeficiency virus infected patients and immunocompromised allograft recipients. *J. Infect. Dis.* **175**:179–184.
18. **Souza, I. E., A. Gregg, D. Pfab, J. D. Dawson, P. Benson, M. E. O'Neill, J. R. Murph, S. J. Petheram, and J. F. Bale, Jr.** 1997. Cytomegalovirus infection in newborns and their family members: polymerase chain reaction analysis of isolates. *Infection* **25**:144–149.
19. **Spaete, R. R., and E. S. Mocarski.** 1985. The *a* sequence of the cytomegalovirus genome functions as a cleavage/packaging signal for herpes simplex virus defective genomes. *J. Virol.* **54**:817–824.
20. **Stinski, M. F.** 1990. Cytomegalovirus and its replication, p. 1959–1980. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, et al. (ed.), *Virology*. Raven Press, Ltd., New York, N.Y.
21. **Van Belkum, A.** 1999. The role of short sequence repeats in epidemiologic typing. *Curr. Opin. Microbiol.* **2**:306–311.
22. **Zaia, J. A., G. Gallez-Hawkins, M. A. Churchill, A. Morton-Blackshire, H. Pande, S. P. Adler, G. M. Schmidt, and S. J. Forman.** 1990. Comparative analysis of cytomegalovirus *a* sequence in multiple clinical isolates by using the polymerase chain reaction and restriction fragment length polymorphism assay. *J. Clin. Microbiol.* **28**:2602–2607.