Supplementary Tables:

Table S1. Summary statistics of viral load in different body fluids at time of seroconversion

Source	Total # samples (Placebo/gBMF59)	Mean $\log_{10}$ VL	Median $\log_{10}$ VL	Log_{10} VL IQR
Oral	24 (14/10)	2.95	2.12	<2 – 2.74
Urine	30 (20/10)	3.06	<2	<2 – 2.56
Vaginal	25 (16/9)	3.86	3.25	<2 – 3.87
Whole blood	21 (14/7)	2.21	<2	<2 – 2.30

Table S2. Trends in percent detectable ($\log_{10}$ viral load ≥ 2.0 copies/ml) over time by source

	Oral	**Urine**	**Vaginal**	**Whole Blood**
Visit*	% Detectable (n)	% Detectable (n)	% Detectable (n)	% Detectable (n)
0	55.0 (20)	42.7 (24)	73.7 (19)	26.9 (26)
1 month	75.0 (4)	42.9 (7)	83.3 (6)	25.0 (4)
2-36 months	20.0 (20)	23.8 (21)	22.2 (18)	5.6 (18)

*Visit 0 is around the time of seroconversion, visit 1 is one month later, etc.

Table S3. Primer sequences

Primer	Sequence
fullgB_PCR1_F	5’–ACACGCAAGAGACCACGACG–3’
fullgB_PCR1_R	5–TTGAAAAACATAGCGGACCG–3’
fullgB_PCR2_F	5’–ATGGAATCCAGGATCTGGTG–3’
fullgB_PCR2_R	5’–TCAGACGTTCTCTTCTTCGT–3’
gBamp_PCR2_F	5’–Illumina_overhang–GAAAACAAAACCATGCAATT–3’
gBamp_PCR2_R	5’–Illumina_overhang–GTCGGACATGTTCACTTCTT–3’
UL130_PCR1_F	5’–TGGGATGGGTGCAGAAGGT–3’
UL130_PCR1_R	5’–GGCTTCTGCTTCGTCACCAC–3’
UL130_PCR2_F	5’–Illumina_overhang–ATCGCACCTGAAAAGACACG–3’
UL130_PCR2_R	5’–Illumina_overhang–CCCCGCCATGGTCTAAACTG–3’

Supplementary Figures:

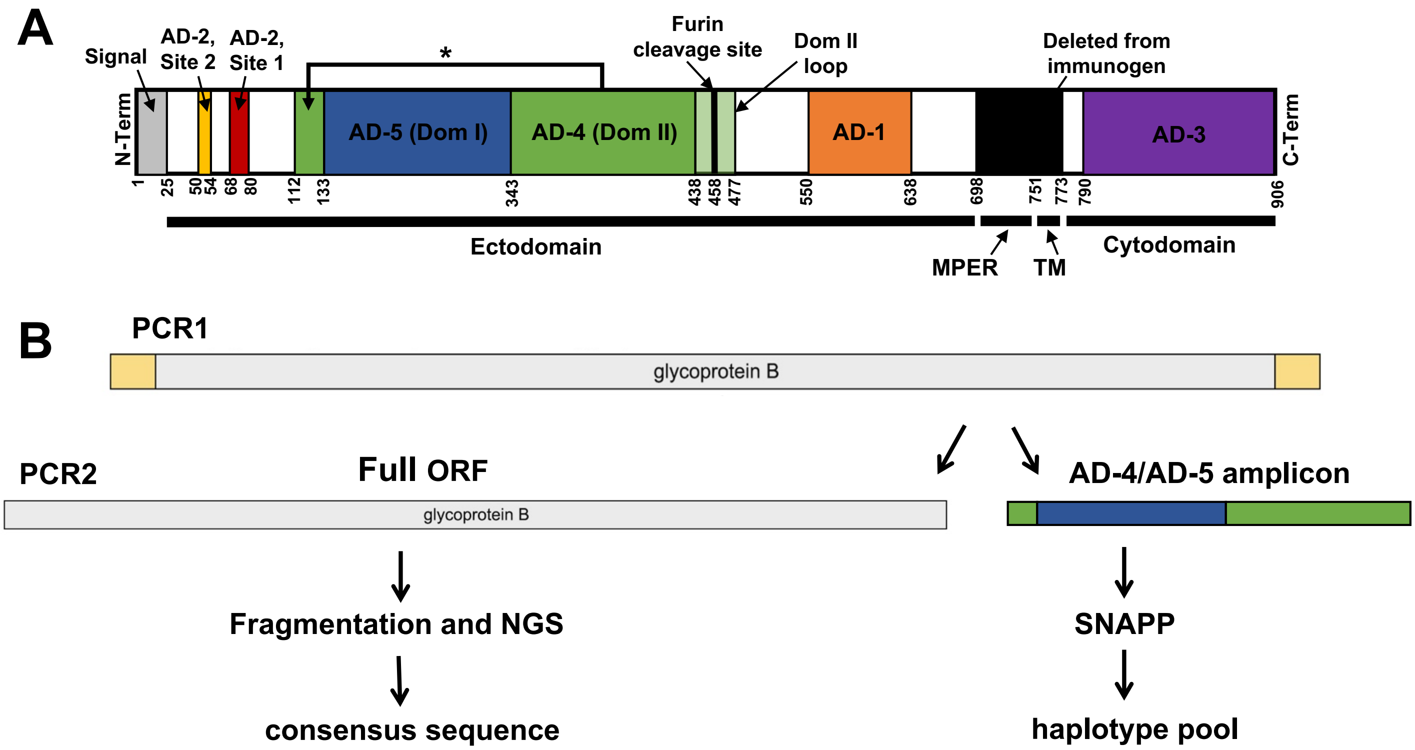

Figure S1. Linear structure of gB and PCR amplification/next-generation sequencing strategy. (A) The full gB HCMV open reading frame (ORF) is shown, from N-terminus on the left to C-terminus on the right. The four distinct regions of the gB structure are indicated by black bars at the base of the figure, including the ectodomain, membrane proximal external region (MPER), transmembrane domain (TM), and the cytodomain. Major antigenic regions indicated include AD-1 (orange), AD-2 site 1 (red), AD-2 site 2 (yellow), AD-3 (purple), AD-4 (Domain II) (green), and AD-5 (Domain I) (blue). Numbers indicate approximate amino acid residues dividing each region of interest. The gB immunogen employed in this clinical trial contained the full gB ORF with the furin cleavage site mutated and excluding a region from amino acid residue 698 to 773 (containing MPER and TM regions) to facilitate protein secretion during production. Diagram was adapted from Burke et al., *Plos Pathogens*, 2015 and Hebner et al., *Nature Communications*, 2015. (B) PCR amplification strategy consists of an initial PCR1 step with primers external to the gB ORF, followed by PCR2 amplification of the full gB ORF or an amplicon containing AD-4 and AD-5. Full gB ORF was NGS sequenced to generate a consensus sequence, while gB amplicons were sequenced directly and raw reads used to infer unique viral haplotypes. *indicates discontinuous nature of AD-4.

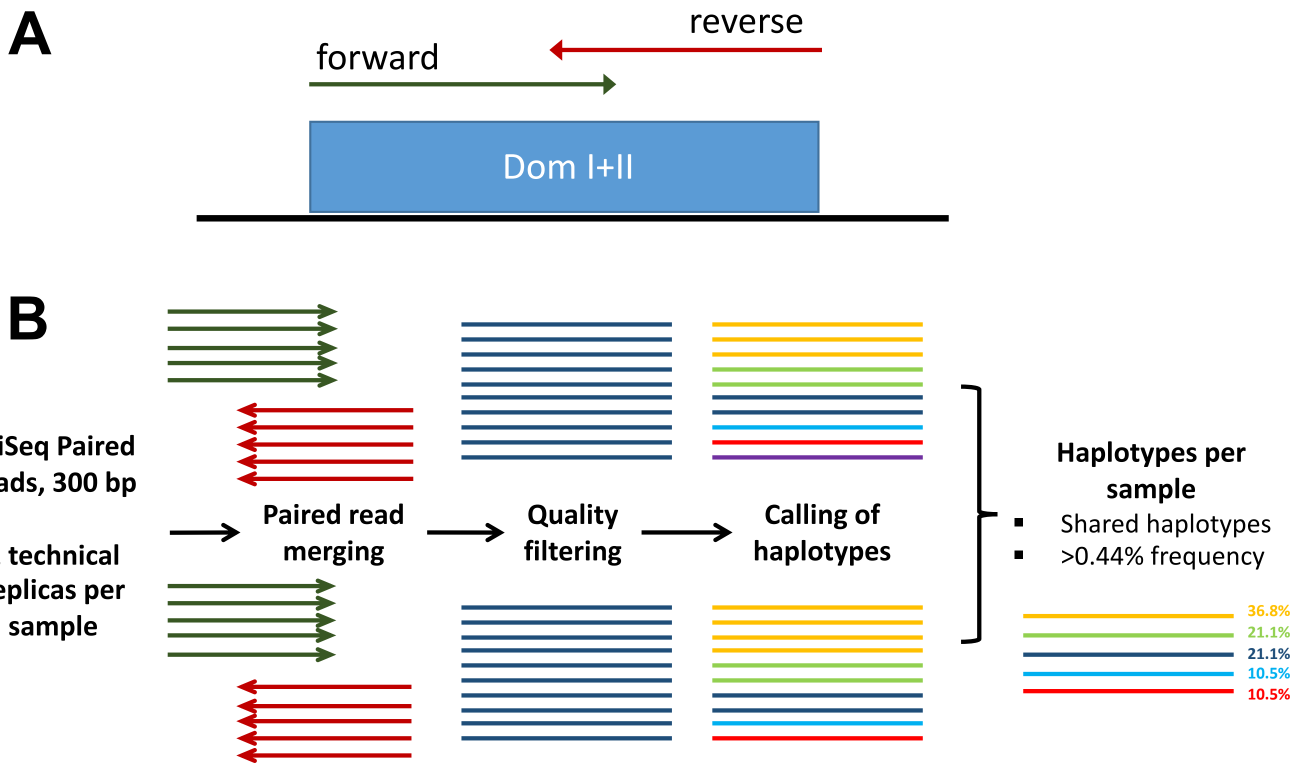

Figure S2. SNAPP analysis pipeline using SeekDeep. (A) Paired-end reads (forward=green, reverse=red) were obtained for an approximately 550 base-pair amplicon on an Illumina Miseq platform. (B) Paired-end reads were merged, filtered for read quality, then clustered into unique haplotypes. Haplotypes identified in both technical replicates at a frequency above the determined 0.44% cutoff were included for subsequent analysis.

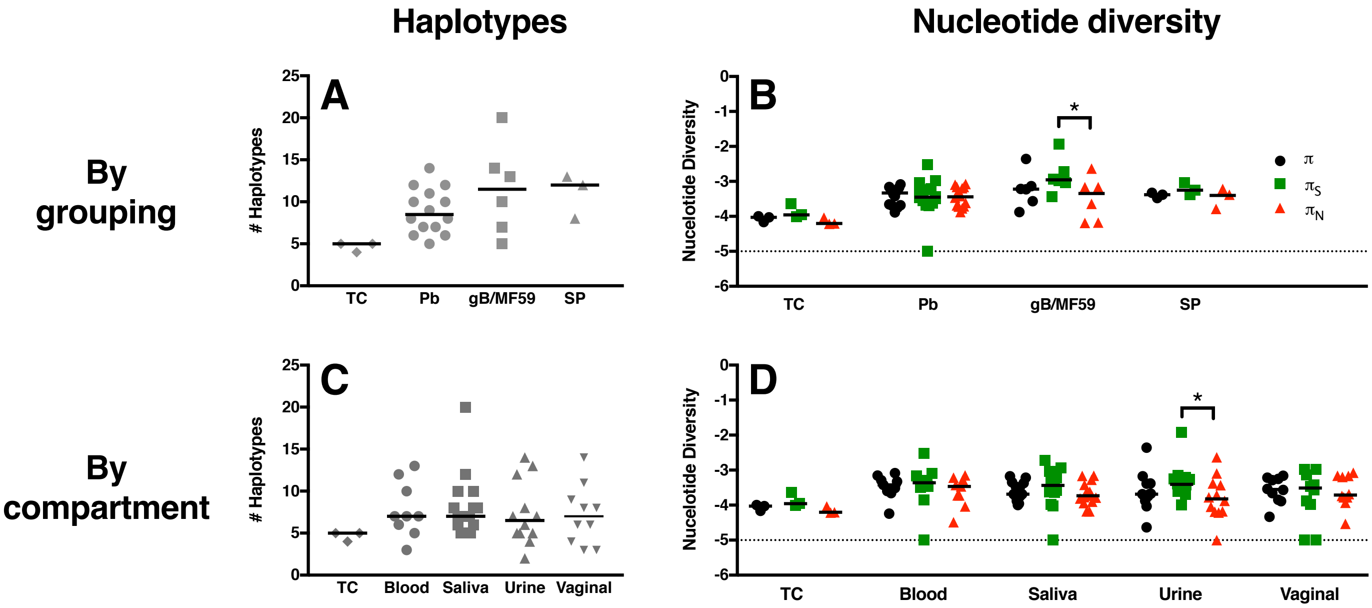

Figure S3. UL130 unique viral variants and peak nucleotide diversity is similar both between gB vaccine and placebo groups and between anatomic compartments. The peak number of unique viral haplotypes as well as peak nucleotide diversity (π) were assessed for viral DNA amplified at the UL130 locus between treatment groups (A,B) as well as between physiologic compartments (C,D). In total samples were analyzed from 14 placebo recipients, 6 gB/MF59 vaccinees, 4 seropositive individuals and 3 tissue culture viruses (TC virus). Additionally, peak viral load by compartment was identified for 9 whole blood, 15 saliva, 12 urine, and 10 vaginal fluid samples. (G) The magnitude of nucleotide diversity resulting in synonymous (π_S) vs. nonsynonymous (π_N) changes was compared. Horizontal bars indicate the median values for each group. *p<0.05, viral load = Friedman test + post hoc Pairwise Wilcoxon Signed Rank test, haplotypes & π = Kruskal-Wallis test + post hoc Exact Wilcoxon Rank Sum test, π_S vs. π_N = Wilcoxon Signed Rank test.

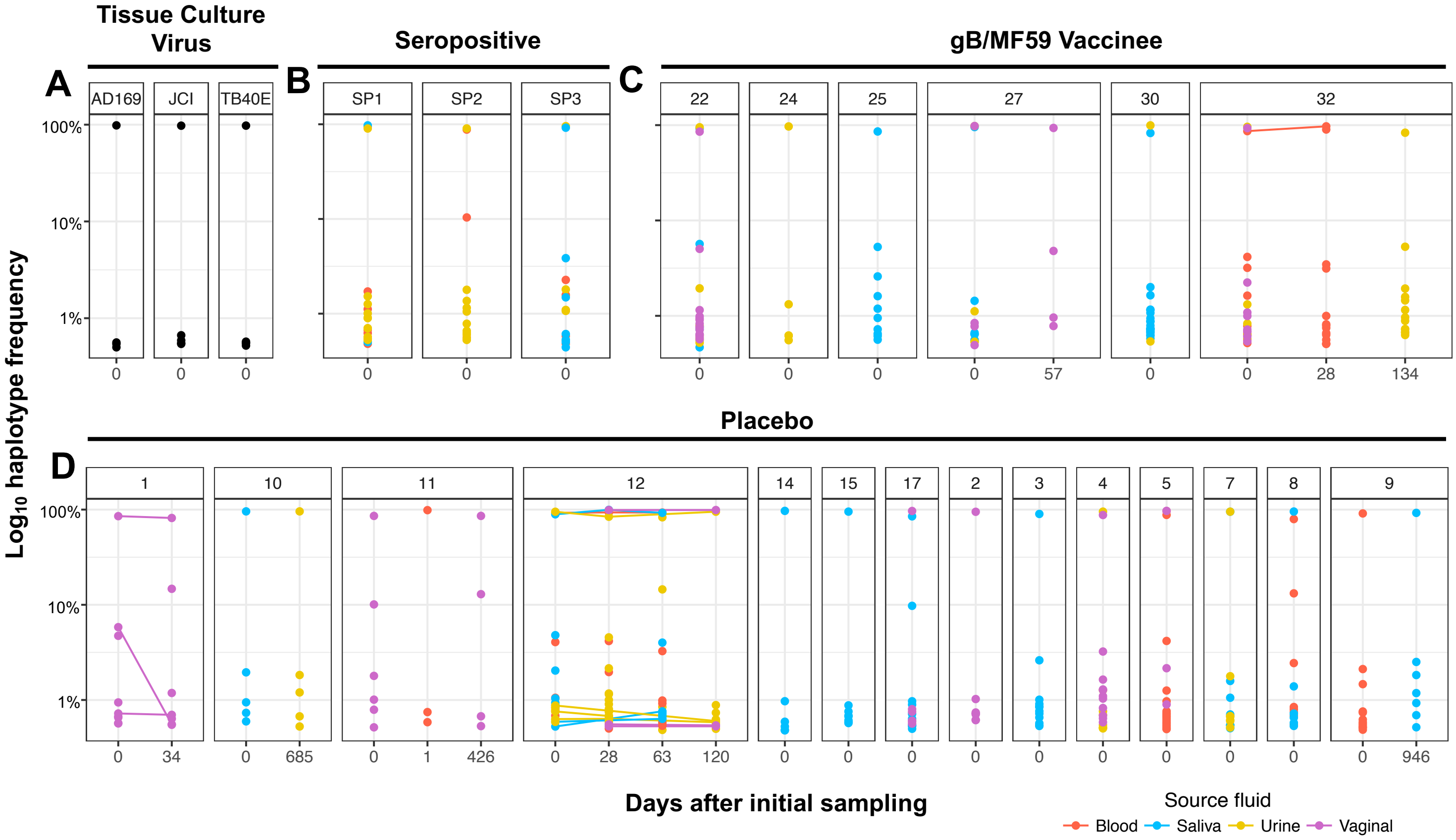

Figure S4. Low-frequency viral variants detectable at UL130 locus in both primary HCMV-infected and chronically-infected individuals. The relative frequency of each unique UL130 haplotype identified by SNAPP is displayed by individual patient and time point of sample collection. In primary HCMV-infected placebo recipients (A) and gB vaccinees (B) as well as chronically HCMV-infected women (C), there are typically one or more high-frequency haplotypes representing the dominant viral variants comprising the population accompanied by haploptyes at very low frequency representing minor viral variants (<1% of viral haplotype prevalence). Tissue culture viruses (D) exhibited reduced population complexity by comparison. Color indicates source fluid: red=blood, blue=saliva, yellow=urine, and pink=vaginal fluid. All haplotypes displayed exceed the 0.44% threshold of PCR and sequencing error established for the SNAPP method (see materials and methods for detail).

Group	PTID	# Comp	F_{ST}	S_{nn}	SM	AI	r	r_b	Result
Placebo	4	2	**0**	**0**	**0.001**	**0.218**	**0.001**	**0.001**	**Yes**
	5	2	UP	UP	1.00	0.912	0.442	0.880	No
	7	2	UP	UP	0.704	0.799	0.357	0.635	No
	8	2	UP	UP	0.388	0.534	0.505	0.515	No
	9	2	**0**	**0**	**0**	**0.014**	**0.001**	**0.001**	**Yes**
	10	2	**0.023**	**0**	1	0.609	0.917	0.985	No
	11	2	UP	UP	1	0.503	0.111	0.831	No
	12	4	0.241	0.43	0.306	0.966	0.771	1	No
	17	2	0.448	0.65	**0.034**	0.753	0.185	0.517	No
gB/MF59	22	3	0.505	0.134	0.152	1.056	0.646	0.791	No
	27	3	**0.044**	0.1	**0.015**	0.458	0.499	0.158	No
	30	2	UP	UP	1	0.838	0.349	0.956	No
	32	4	**0**	**0**	**0.001**	0.879	**0.001**	**0.001**	**Yes**
Seropos	SP1	3	**0**	**0**	**0.001**	0.679	**0.001**	**0.001**	**Yes**
	SP3	2	0.819	0.761	0.373	0.963	0.376	0.814	No
	SP4	3	**0**	**0**	**0**	**0.192**	**0.001**	**0.001**	**Yes**

Figure S5. Lack of genetic compartmentalization of anatomic compartment-specific viral variants detected at UL130 locus in vaccinees. Table indicating the results of 6 distinct tests for genetic compartmentalization performed on the pool of unique UL130 haplotypes identified per patient, including Wright's measure of population subdivision (F_{ST}), the nearest-neighbor statistic (S_{nn}), the Slatkin-Maddison test (SM), the Simmonds association index (AI), and correlation coefficients based on distance between sequences (r) or number of phylogenetic tree branches (r_b). For each test, >1000 permutations were simulated. Significant test results suggesting genetic compartmentalization are shown in gray with bold text. Values for F_{ST}, S_{nn}, SM, r, and r_b represent uncorrected p-values, with $p<0.05$ considered significant. An AI<0.3 was considered a significant result. Three or more positive tests per patient was considered strong evidence for genetic compartmentalization, indicated in green.

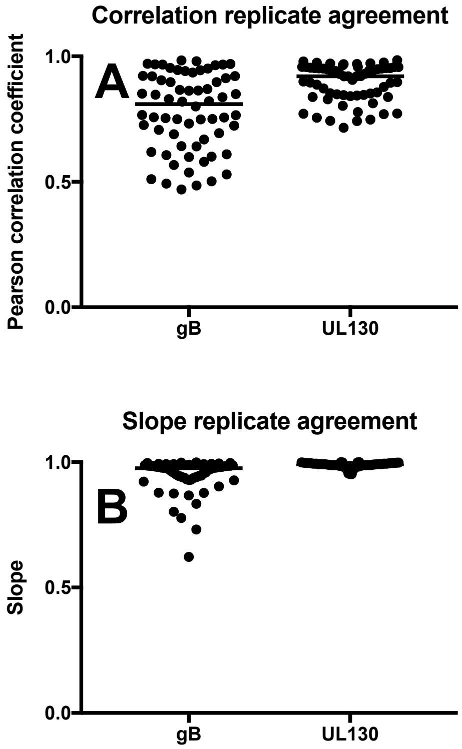

Figure S6. High degree of concordance in haplotype identity and frequency between sequencing replicates. Haplotype identity and frequency were calculated for two technical replicates. The correlation (A) and slope (B) of the haplotype frequencies was compared between technical replicates for both gB and UL130 amplicons, and indicate a high degree of agreement between replicates.

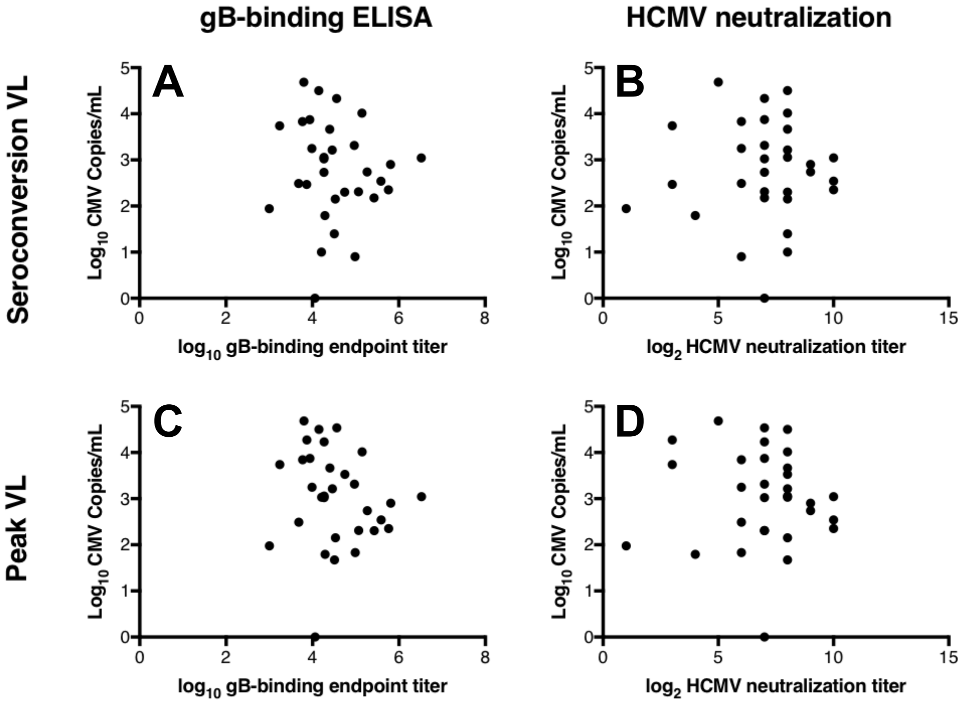

Figure S7. Viral load is not correlated with gB antibody-binding or HCMV neutralization. Viral load at seroconversion does not correlate with the magnitude of gB-binding (R^2=0.0428) (A) nor with HCMV-neutralization titer (R^2=0.0035) (B). Furthermore, peak viral load neither correlates with gB-binding (R^2=0.0035) (C) nor neutralization titer (R^2=0.0009) (D).

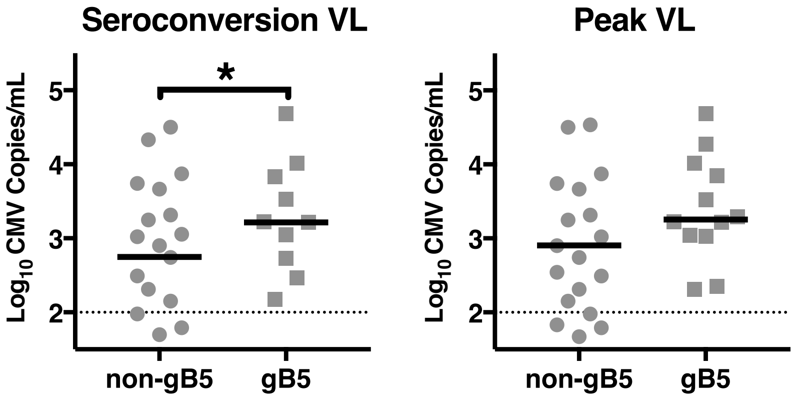

Figure S8. Higher viral load among women that acquired gB5 genotype viruses. A linear regression analysis of log$_{10}$ viral load on genotype was performed at time of seroconversion (A) as well as peak viral load (B). At time of seroconversion, the viral load among women who acquired a gB5 genotype virus was 3.44 time greater than that of women shedding non-gB5 genotype virus (95% CI 1.13-10.51, p=0.031). Solid line for each grouping indicates mean value, whereas dotted black line indicates threshold of qPCR detection (100 copies/mL).

A

gB locus

	VL (/mL)	Haplotypes	π	π_S	π_N
VL		-0.017	-0.115	-0.062	-0.049
Haplotypes			**0.452**	**0.321**	**0.419**
π				**0.757**	**0.777**
π_S					**0.419**
π_N					

B

UL130 locus

	VL (/mL)	Haplotypes	π	π_S	π_N
VL		0.103	-0.068	-0.174	0.039
Haplotypes			**0.537**	**0.323**	**0.506**
π				**0.341**	**0.717**
π_S					0.049
π_N					

Figure S9. Correlations between viral load, number of unique variants, and nucleotide diversity. Kendall Tau correlation coefficients are shown for viral load (VL), number of haplotypes, nucleotide diversity (π), as well as synonymous nucleotide diversity (π_S), and nonsynonymous nucleotide diversity (π_N). Bold values indicate a significant correlation (uncorrected $p<0.05$).