JOURNAL OF VIROLOGY, Feb. 2002, p. 1971–1979 Vol. 76, No. 4
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.4.1971–1979.2002

Phylogenetic Analysis of Varicella-Zoster Virus: Evidence of Intercontinental Spread of Genotypes and Recombination

Winsome Barrett Muir, Richard Nichols, and Judith Breuer*

Schools of Medicine and Biological Sciences, Queen Mary College, University of London, London E1 1BB, England

Received 13 September 2001/Accepted 12 November 2001

A heteroduplex mobility assay was used to identify variants of varicella-zoster virus circulating in the United Kingdom and elsewhere. Within the United Kingdom, 58 segregating sites were found out of the 23,266 examined (0.25%), and nucleotide diversity was estimated to be 0.00063. These are an order of magnitude smaller than comparable estimates from herpes simplex virus type 1. Sixteen substitutions were nonsynonymous, the majority of which were clustered within surface-expressed proteins. Extensive genetic correlation between widely spaced sites indicated that recombination has been rare. Phylogenetic analysis of varicella-zoster viruses from four continents distinguished at least three major genetic clades. Most geographical regions contained only one of these three strains, apart from the United Kingdom and Brazil, where two or more strains were found. There was minimal genetic differentiation (one or fewer substitutions in 1,895 bases surveyed) between the samples collected from Africa (Guinea Bissau, Zambia) and the Indian subcontinent (Bangladesh, South India), suggesting recent rapid spread and/or low mutation rates. The geographic pattern of strain distribution would favor a major influence of the former. The genetic uniformity of most virus populations makes recombination difficult to detect. However, at least one probable recombinant between two of the major strains was found among the samples originating from Brazil, where mixtures of genotypes co-occur.

Genetic variation among strains of human herpesviruses has been used to distinguish viral genotypes and to conduct epidemiological studies (7, 14, 33, 40, 45–51). Variation results predominantly from single nucleotide polymorphisms (SNPs) or from alteration in the composition and number of repeat elements present either within the internal and terminal repeat regions or within tandem direct reiterations scattered throughout the genome (7, 8, 27, 35). Genetic variation in varicella-zoster virus (VZV) has been defined by the presence or absence of restriction sites, such as a *Pst*I site in gene 38 (25) and a *Bgl*I site in gene 54 (1), or differences in the number of repeat elements within the five repeat regions in the VZV genome (5, 20, 24, 25, 26, 28, 43, 44). These methods were found to distinguish viruses from different geographic regions and have been used to differentiate between the live attenuated Japanese Oka vaccine strain and wild-type viruses circulating in the United States and United Kingdom (17, 21, 23, 29, 47). Distinctive restriction enzyme patterns have also allowed differentiation between epidemiologically unrelated viruses, while viruses occurring in a single outbreak have been shown to be identical (20, 41). The establishment of latency by the virus in the dorsal root ganglion does not appear to affect the genotype, as evidenced by the identical restriction enzyme profiles of the infecting and reactivating strain from a single individual (42). Restriction enzyme profiles have also been shown to remain stable on serial passage in tissue culture (20).

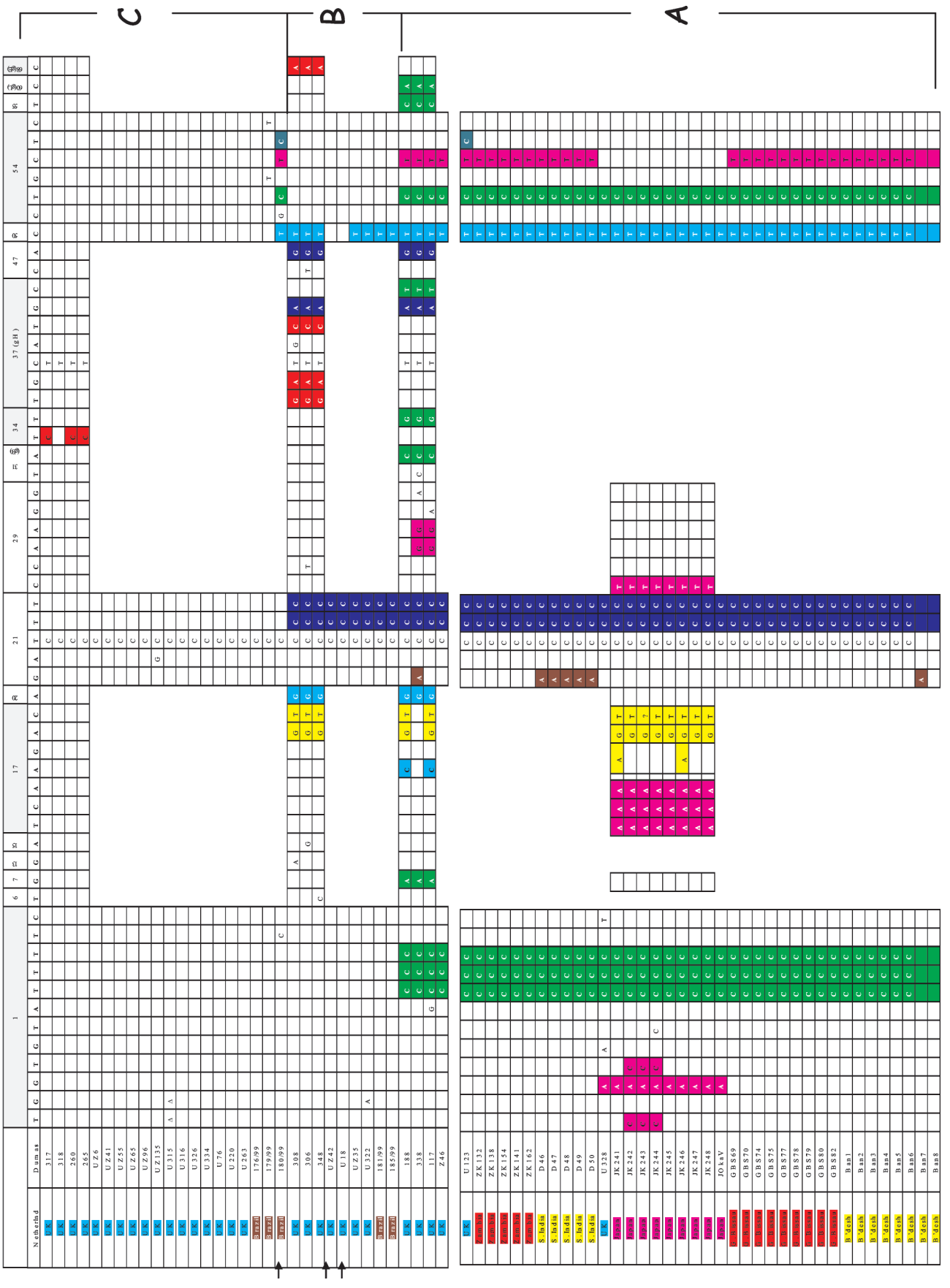
Using these established methods, we have previously shown a mixture of genetically distinct strains to be circulating in the United Kingdom, most particularly in the East End of London, where the proportion of VZV strains carrying the *Bgl*I restriction endonuclease site in gene 38 increased from 10% in the early 1980s to more than 30% in the 1990s (22). Sampling of viruses from outside the United Kingdom showed a 90 to 100% prevalence of *Bgl*I-positive viruses in Asian and African countries, while the United States was similar to the United Kingdom (4). More recently several groups, including ours, have identified additional SNPs which can be used to genotype VZV strains (3, 12; V. N. Loparev and D. S. Schmid, Abstr. 26th Int. Herpesvirus Workshop, abstr. 3.15., 2001). Using the SNP map we have generated, estimates of VZV variation have been derived. In this report we use phylogenetic and population genetic analysis, which has enabled us, for the first time, to examine the evolutionary relationships between genotypes and to analyze the putative mechanisms by which VZV evolution and spread are likely to occur in the future.

MATERIALS AND METHODS

Samples. The polymorphic map was developed using 10 VZV isolates randomly selected from around the United Kingdom. These data were used to estimate the time since the common ancestor of the major clades (see below). This calculation does not require a large sample size, because of the redundancy of information from similar sequences of the same clade, but it does benefit from the extensive sequence survey. The most informative sites were selected for further analysis (Fig. 1). We proceeded to test further 10 United Kingdom samples for the whole set of primers to verify this choice and to establish that these were SNPs and not rare variants or PCR/sequencing artifacts. A larger sample, made up by an additional 67 strains typed at the phylogenetically informative loci, was used to survey the global distribution of genetic diversity. These comprised an additional 25 United Kingdom samples, collected from patients with varicella and zoster in East London, and 42 non-United Kingdom samples (Fig. 1).

DNA extraction and amplification. DNA was extracted from 200 μl of each vesicle fluid sample using the QIAamp Blood Mini Kit (Qiagen Ltd., Crawley, United Kingdom). Viral DNA from each sample was initially genotyped at four loci by methods previously established (3). The 56 sets of primers shown in Table

* Corresponding author. Mailing address: Department of Medical Microbiology, 25-29 Ashfield St., London E1 1BB, England. Phone: 44 207 377 7141. Fax: 44 207 375 0518. E-mail: j.breuer@mds.qmw.ac.uk.


C
B
A

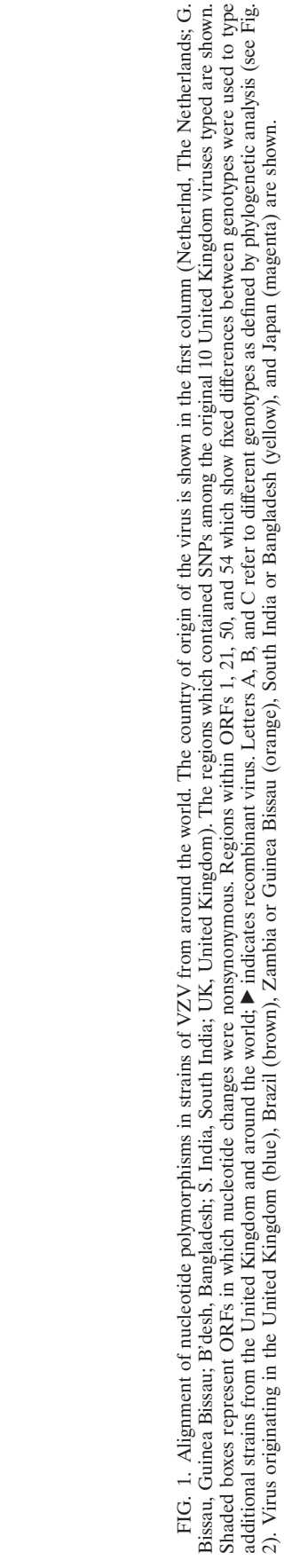


FIG. 1. Alignment of nucleotide polymorphisms in strains of VZV from around the world. The country of origin of the virus is shown in the first column (Netherlnd, The Netherlands; G. Bissau, Guinea Bissau; B'desh, Bangladesh; S. India, South India; UK, United Kingdom). The regions which contained SNPs among the original 10 United Kingdom viruses typed are shown. Shaded boxes represent ORFs in which nucleotide changes were nonsynonymous. Regions within ORFs 1, 21, 50, and 54 which show fixed differences between genotypes were used to type additional strains from the United Kingdom and around the world; ▶ indicates recombinant virus. Letters A, B, and C refer to different genotypes as defined by phylogenetic analysis (see Fig. 2). Virus originating in the United Kingdom (blue), Brazil (brown), Zambia or Guinea Bissau (orange), South India or Bangladesh (yellow), and Japan (magenta) are shown.

1 were designed to amplify 500-bp regions at 3,000-bp intervals (3). DNA amplification reactions were performed in 100-μl reaction volumes using 1 U of AmpliTaq Gold DNA polymerase enzyme (Perkin-Elmer). The mixture comprised PCR buffer II (Perkin-Elmer), 200 μM concentrations of each deoxynucleotide triphosphate, each primer at a concentration of 0.2 μM, 1 μl of test or control DNA extract, and an optimal $MgCl_2$ concentration. The optimum magnesium concentration was determined for each primer pair by titration from 1 to 4 mM $MgCl_2$ using the positive control DNA extract (3). Thermal cycling included an initial hot start at 95°C for 12 min, followed by 30 to 40 cycles of 94°C for 1 min, the annealing temperature for 1 min, and then 72°C for 1 min, with a final extension at 72°C for 10 min. PCR products were visualized by gel electrophoresis in agarose containing ethidium bromide, alongside a 100-bp ladder (Gibco BRL).

HMA. The heteroduplex mobility assay (HMA) was optimized using PCR products spanning a *Bgl*I site in gene 54 from samples with and without the restriction site polymorphism. The method used has been previously described (3). Briefly, 2 μl of each PCR product was mixed with 2 μl of gel loading buffer and denatured by heating at 98°C for 5 min, followed by 68°C for 30 min, and then held at 4°C. The samples were then mixed with 7.5 μl of gel loading buffer (45% formamide, 30% ethylene glycol, 10 mM EDTA, 5% Ficoll, and 0.05% bromophenol blue and xylene cyanol). Samples were electrophoresed on a 6% polyacrylamide gel, with 0.6 g of piperazine/liter in Tris-taurine-EDTA (TTE) buffer containing 10% ethanediol and 13.75% formamide. After electrophoresis, the gel was fixed and stained by immersing it sequentially in 0.1% silver nitrate for 15 min, followed by 1.5% sodium hydroxide–0.3% formaldehyde solution for approximately 5 min until bands were just visible. The gel was fixed in 5 to 6% glacial acetic acid for 5 min and removed from the silane-coated glass plates by immersion in 1.5% sodium hydroxide for 20 min.

DNA sequencing. Polymorphisms identified by HMA were confirmed and characterized by nucleotide sequencing of PCR products using the relevant PCR primers in an ABI Prism dRhodamine terminator cycle sequencing reaction mix (Perkin-Elmer Applied Biosystems) according to the manufacturer's instructions. Cycle sequencing reactions were electrophoresed on an ABI 377 Analyser (Applied Biosystems Inc.). Sequences generated were analyzed using Sequence Navigator (Applied Biosystems Inc.) and compared to that of the published VZV (strain Dumas) sequence (9, 11) in order to identify sequence polymorphisms.

Phylogenetic analysis. Aligned SNPs identified in 10 United Kingdom VZV strains were analyzed by the DNAPARS program within the Phyllip package (13), using the Dumas strain sequence as an outgroup (3). The primer pairs found to flank informative SNPs were used to analyze additional strains from the United Kingdom and other countries (Fig 1). The analysis was conducted using the PAUP 4.0b8 program. A neighbor-joining tree was calculated to illustrate the distribution of intergenotype differences. The support for the branching pattern was assessed by bootstrapping, using maximum parsimony as the criterion. Evidence of linkage disequilibrium was evaluated using Fisher's exact tests to compare the frequency of allelic combinations at pairs of loci with the frequencies expected under free recombination. The data were also used to calculate the nucleotide diversities within populations and between populations and the net difference between populations (π_X, π_{XY}, π_A) using the methods of Nei (34).

Time since branching of the major clades. The estimate of the time since the splitting of the major clades depends on the mutation rate, μ. The maximum likelihood value for the time can be calculated from the number of substitutions ($x = 12$) using the Poisson likelihood function, $L(x|\mu,t) \propto \lambda^x e^{\lambda}$, where λ is the expected number of synonymous substitutions in b bases ($\lambda = \mu b 2\, t$). Similarly the range of plausible values could be calculated as a likelihood curve. The synonymous substitutions and their rate were used, as this varies less between genes and taxa.

The 95% limits (the posterior 95% equiprobable interval) were calculated from this equation by using the mutation rate of $\mu = 1.0 \times 10^{-7}$ substitutions per nucleotide per year, obtained from a study of related alphaherpesviruses (31). The calculation assumed a uniform prior on the time (t). The distribution could therefore be obtained directly by numerical integration of the distribution for a range of mutation rates. These were normally distributed around this mean for μ, with a standard error of 20% to allow for uncertainty in the estimate and in extrapolation between viruses (but see Discussion).

Estimate of mutation rate. An alternative method of analysis is to assume that the major clades diversified at or before the major human migrations 100,000 years ago (100 kya), as has been suggested for alphaherpesviruses. An estimate of μ can be obtained by substituting rates in the likelihood function and examining the relative support for times exceeding 100 kya. We use the criterion that 5% of the distribution should exceed 100 ky (equivalent to a 5% probability under a uniform prior probability distribution).

TABLE 1. Primers used to amplify and sequence viruses

Primer name	Position (5′ → 3′)	Size (bp)	Sequence (5′ → 3′)
VZV-1 F	427–450	506	TCAGCTGGCTTTTCTAAGAATTCG
VZV-1R	932–911	506	TATTTTTGGGATCCGCAATGTC
VZV-4 F	3270–3291	458	GCAGACTCCAACGCTTCAATCA
VZV-4 R	3727–3706	458	AATCGAAGACGGCGTCCTACAA
VZV-6 F	6066–6087	455	CAACAAATCCCCGTTCCAGCAA
VZV-6 R	6520–6500	455	CATGCGAGCATTTCGTCATGG
VZV-7 F	8723–8743	487	AGAGGCTTTAACCGCTGTGAA
VZV-7 R	9208–9189	487	CGCCTCCGATTTAATAGGTGA
VZV-10 F	12423–12447	493	CTGGAACGAGGATTTATTCTCATGT
VZV-10 R	12915–12894	493	TGCAGATTGACTGGCGTACAAA
VZV-11 F	15898–15919	500	TGAACTTCCGTTTTGTGGGATT
VZV-12 R	16397–16378	500	CGGGCGTTGGAATAAGACAT
VZV-13 F	19017–19039	476	AACGGTGAATTATCCTGCCAAGT
VZV-13 R	19492–19470	476	TCCTTGGGACATTGGGTCTTATC
VZV-14a F gC	19434–19454	514	TGAACAGCAACGGATGCATAG
VZV-14a R	19947–19925	514	AATTCCACAGCAAAAACAAATGG
VZV-14b F gC	19925–19947	523	CCATTTGTTTTTGCTGTGGAATT
VZV-14b R	20447–20425	523	TGGTTTACGCGTCACCTTATAGA
VZV-15 F	21709–21730	522	CGTTTAAGGGTCCGGGAACTTT
VZV-15 R	22230–22211	522	TTCGGGGTTATGGCCAACAT
VZV-17 F	24402–24424	487	GTTGTACAGACGGCCCATTATCA
VZV-17 R	24888–24865	487	TCACAGCCCATGAAGAGTAAATCA
VZV-19 F	27209–27232	497	TCCATGCTTGTTTTCATTGTCTCA
VZV-19 R	27705–27682	497	TCTTCTGTTCCCATTCAAGATATC
VZV-20 F	30027–30049	496	TCAGATGGATGACGAAGGATTAA
VZV-20 R	30522–30500	496	GATAGTACCACGTCACGATTGCA
VZV-21 FOR	33497–33518	503	TAATGAATTGAGGCGCGGTTTA
VZV-21 R	33999–33976	503	CACGTGTAGCTCCAAAAACCTAGG
VZV-22a F	37007–37027	509	GGTAACCGGTTTTGTGGGACT
VZV-22a R	37515–37489	509	GTAATTCAGTCAACTTATGGGTATTCG
VZV22b F	40507–40528	492	AGGCCAATGTCGATGCAGTTAC
VZV-22b R	40998–40973	492	ATATACCCGAATCTGTAGCCATATGC
VZV-24 F	43985–44006	511	TCCTGCGTTCACTCCGTACATA
VZV-25 R	44495–44476	511	AGGGCCCGCGATTTAATTAC
VZV-28 F	47506–47525	509	GGGGCCACAATAACATAAGG
VZV-28 R	48014–47992	509	GCACATATAATTTCAACGGCTCT
VZV-29a F	51043–51060	507	GCGGTGGTTTCTGGAGCA
VZV-29aR	51549–51526	507	GGTGTTATAAAAGGGTGTTGGGTA
VZV-29 F	52795–52814	542	TTCGAAGCCACCCATTCAAT
VZV-29 R	53336–53315	542	CGAGGAAATAACAGGCCGTGAA
VZV-30 F	56314–56335	513	CGGCCTTTGATGCACATAGATT
VZV-30 R	56826–56804	513	GGGGACATGTTTCTTCGTACGTT
VZV-31a F gB	56988–57008	513	GCTTTTTATGCATATTTTCTA
VZV-31a R	57500–57480	513	CTTCAACTTTGTGGTTATTTC
VZV-31b F gB	57480–57500	507	GAAATAACCACAAAGTTGAAG
VZV-31b R	57986–57967	507	GAAAATTGTGTGCATACTCA
VZV-31c F gB	57967–57986	514	TGAGTATGCACACAATTTTC
VZV-31c R	58480–58463	514	AAATAGTCCGCTCCAAAG
VZV-31d F gB	58463–58482	515	CTTTGGAGCGGACTATTTCC
VZV-31d R	58977–58959	515	CGGCGTTGAATTTCACTGT
VZV-31e F gB	58935–58956	355	TGCGGGATACAGGATTACTAGA
VZV-31e R	59289–59268	355	AGTGGATATAATGCCTTCATCG
VZV-31f F gB	59268–59289	413	CGATGAAGGCATTATATCCACT
VZV-31f R	59680–59657	413	AATGGAGACACATGAGTAACGTAA
VZV-32 F	59798–59819	503	ACCGTCGTCTATCGCACATCAT
VZV-32 R	60300–60278	503	ATTCATAAAAGGCGTGTTTGGTT
VZV-34 F	63291–63312	509	TTCATAAACGTACGGGAAATGC
VZV-34 R	63799–63780	509	GCTCCCGCGTTTGAGTTAAC
VZV-37 F	66787–66806	516	CGGTTCGGTGCTTCTCACAA
VZV-37 R	67302–67277	516	ATCCGCATTAAAATGTAATTAAGGAA
VZV-37a F gH	66036–66055	508	AACGTTGCGGTGATATTGTA
VZV-37a R	66543–66526	508	AGGGGGTTTGGTGGGAAC
VZV-37b F gH	66527–66544	511	TTCCCACCAAACCCCCTT
VZV-37b R	67037–67014	511	CATGCTTTGAAAAATTCATATCCA
VZV-37c F gH	67007–67028	514	CGTGGGTTGGATATGAATTTTT
VZV-37c R	67520–67502	514	GCCTGGCGAGAGAAAGTGC
VZV-37d F gH	67440–67460	430	ATGAAGCCCGTGATCAACTAA
VZV-37d R	67869–67851	430	CGTGTTGGTATTGCCGACA

Continued on facing page

TABLE 1—*Continued*

Primer name	Position (5′ → 3′)	Size (bp)	Sequence (5′ → 3′)[a]
VZV-37e F gH	67851–67869	438	TGTCGGCAATACCAACACG
VZV-37e R	68288–68268	438	CCCCCGTGGTTAGATACCTAA
VZV-37f F gH	68252–68272	327	TGCGGAAGTGTTTTTCTTAGG
VZV37f R	68578–68558	327	ATATTCTCGAAGGCGGGAATT
VZV-38 F	70295–70320	490	TTAAAGCGTTTATCGGAACATTAACC
VZV-39 R	70784–70761	490	TCACCCCCATACGATAACATAACC
VZV-40 F	73771–73793	512	GGATGACACGTTTATTGCACCTA
VZV-40 R	74282–74260	512	CTTGTAGTGTGAGCAATGCATCA
VZV-42 F	77260–77283	500	AACTGTTGTGCTATAGCCAACGAT
VZV-42 R	77759–77741	500	AAGCCGGTTTTCATCACAA
VZV-44 F	80747–80765	513	AAGAACCCGATCGCGTACC
VZV-44 R	81259–81240	513	TGCGGTAATGCGAGGATTTT
VZV-47 F	84247–84266	513	GATTGATCTTTATGCATTGG
VZV-48 R	84759–84741	513	TTTTACAAAAGGGTGCTGT
VZV-50 F	87736–87756	515	CGCACCCAAAGTGAACATCAT
VZV-51 R	88250–88227	515	TCTCGGATGTCAAATATGTTACGA
VZV-52 F	91226–91247	495	GCTTGGTTCTTAGTGGCAATCG
VZV-52 R	91720–91699	495	CAGGCTGTTTGAAGACGGTAGA
VZV-54 F	94172–94194	505	GGGTGGCCCGTAATATGAATGAA
VZV-54 R	94676–94654	505	TCCGGAACAACAAACGTTCAAAG
VZV-54BgII F+	95005–95027	497	CGTAATGCATAACAGGCCAACAC
VZV-54BgII R+	95501–95479	497	GAAACCTGGCGTCAAACATTACA
VZV-55 F	98505–98523	496	CCATGTCGCGTACGGTCTCTC
VZV-56 R	99000–98982	496	TTGCGTGGGGTTTGTCCTG
VZV-60 F gL	101156–101176	522	CCGATAGTTTCATTTCATTGG
VZV-60 R	1011677–101656	522	GTCGTAGTGAAGGGAAAACACA
VZV-62a F	105497–105514	502	CCGCCTGGGTTTCTGACG
VZV-62a R	105998–105982	502	ACCCGCGTCCCCCTGTC
VZV-62 F	107299–107320	527	CGGTCGTATCGACGGCTCATAG
VZV-62 R	107825–107804	527	GGGCGCCAGAGACAGAAATCAT
VZV63 F-O*	110512–110529	515	TTCGTCCGATTCATAACG
VZV63 RI*	111026–111008	515	CTATGCAAAGGAACATTCG
VZV64 F	111512–111531	661	GCTGGTTTGAACGTCTCCAT
VZV64 R	112172–112151	661	CAGTACGCTTTTATCCGGTGTA
VZV-67NC F gI	114480–114502	508	GCCTCATTTAATCGCGATGTTTT
VZV-67 R	114987–114963	508	ATTTACACCAAGAATGAAACCATCG
VZV-67b F gl	114930–114954	522	ACTTCTTGTTCGGTTAGACCATAGC
VZV-67b R	115451–115430	522	TGTATGCCCCTTCTTGTTTTTG
VZV-68NC F gE	115798–115819	515	CGCCTGTAATATGGGGACAGTT
VZV-68 R	116312–116292	515	GGCCTTGGGGTTTTGGATTAA
VZV-68b F gE	116245–116268	511	TGTAAATGTGGACCAACGTCAATA
VZV-68b R	116755–116735	511	CAAAAACGTGGCGTAGGTAGA
VZV-68c F gE	116717–116732	507	CGCGGCTCCGATGGTA
VZV-68c R	117223–117199	507	CAAACTCTCGGGTGTATCTACAAAC
VZV-68d F gE	117174–117191	522	TACACGACGGGGGCACCA
VZV-68d R	117695–117680	522	TCGCCCGGTTCGGTGA

RESULTS

The HMA does not identify polymorphisms within the terminal 50 bases of the PCR product (15). Allowing for this, approximately 23,266 nucleotides (18.6% of the genome) were screened in 10 United Kingdom viruses. Fifty-eight polymorphic positions were identified (0.25%), and nucleotide (π) diversity was estimated to be 0.00063 (standard error, 0.00018). This agrees with previous estimates obtained by restriction fragment length polymorphism analysis (44). Sixteen of the nucleotide differences coded for an amino acid change (3), and the open reading frames (ORFs) in which they were located are shown in Fig. 1.

There were highly significant nonrandom associations (linkage disequilibrium) between alleles at widely separated loci. Of the 351 pairwise comparisons between loci, 73 showed associations that were significant at the 1% level. Under the null hypothesis of linkage equilibrium, only 3.51 comparisons would be expected to show this level of significance (Table 2). This result suggested that recombination had little effect on the evolution of the genotypes, and so phylogenetic methods were used to represent the relationships between the genotypes. The United Kingdom genotypes fell into three major clades (Fig. 2), designated A, B, and C.

To map the worldwide distribution of genotypes, we used a subset of informative markers. Four regions, located in ORFs 1, 21, 50, and 54, which contained fixed differences between the major clades were selected manually and used to analyze a further 25 United Kingdom viruses (Fig. 1). This restricted marker set gave distinct genotypes, consistent with the previously identified clades. Genotyping additional informative

TABLE 2. Nucleotide diversities estimated from the four regions of 1.9 kb typed worldwide[a]

Origin of virus	Nucleotide diversity in comparison with populations from:						
	UK	Brazil	G. Bissau	Zambia	Bangladesh	S. India	Japan
UK	**0.0005**	0.0001	0.0007	0.0007	0.0007	0.0008	0.0008
Brasil	0.0006	**0.0005**	0.0008	0.0007	0.0008	0.0009	0.0009
G. Bissau	0.0009	0.0011	**0.0000**	0.0000	0.0000	0.0001	0.0003
Zambia	0.0009	0.0011	0.0000	**0.0000**	0.0000	0.0001	0.0003
Bangladesh	0.0009	0.0011	0.0000	0.0000	**0.0000**	0.0001	0.0003
S. India	0.0011	0.0012	0.0001	0.0001	0.0001	**0.0000**	0.0005
Japan	0.0011	0.0012	0.0004	0.0004	0.0004	0.0006	**0.0002**

[a] Diagonal (bold) values, within-population diversities (px); above diagonal values, net nucleotide diversity between populations (pA); below diagonal values, mean proportion of pairwise differences between populations (pxy). The 1.9-kb subset had higher than average diversity, so the values in Table 2 are adjusted accordingly (by a factor of 0.277). UK, United Kingdom; G. Bissau, Guinea Bissau; S. India, South India.

markers located in ORFs 55, 20, and 47 did not reveal further genotypes (data not shown). Additional genotyping of 42 viruses from Africa, Asia, Europe, and South America (Fig. 1 and 2) using the four informative regions was therefore undertaken. Phylogenetic analysis of all of the viruses at the above four regions, representing 1,895 bp, is shown in Fig. 2. A bootstrap analysis using the maximum parsimony criterion was used to assess the support for the major clades. The number of inter- and intrapopulation nucleotide differences at these sites ranged from 0 to 13 and 1 to 12, respectively. These values were standardized as gross and net nucleotide diversities, using the methods set out by Nei (34), and are shown in Table 2.

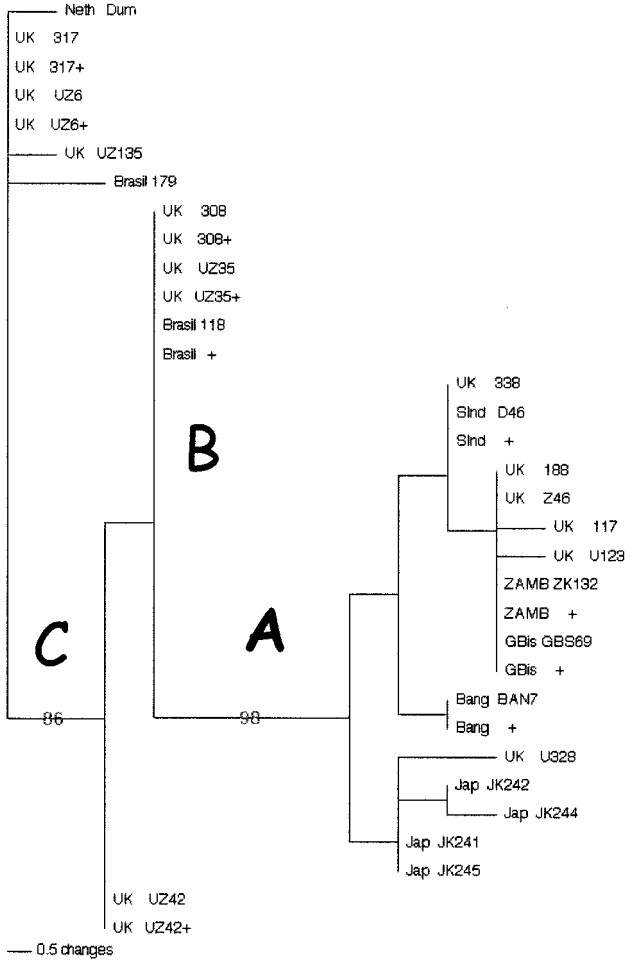


FIG. 2. Phylogenetic tree of varicella-zoster virus strains by the neighbor-joining method. The countries from which specimens were obtained are shown: Bang, Bangladesh; Brasil, Brazil; Jap, Japan; GB, Guinea Bissau; Neth, The Netherlands; Sind, South India; ZAMB, Zambia. Bootstrap values separating clades A, B, and C are shown. + denotes additional viruses with identical sequence.

The estimated time since the common ancestor of the major clades was 3,000 to 19,000 years, assuming the mutation rate derived from other alphaherpesviruses. The upper limit on the mutation rate that remains consistent with a divergence time of 100,000 years was one-seventh of the value for herpes simplex virus (HSV) type 1 and alphaherpesviruses (1.4×10^{-8}).

DISCUSSION

Evolutionary inference from the phylogenetic data. Our estimate of genetic diversity of VZV is 0.00063, which is 10 times lower than that obtained for HSV and 40 times less than estimated for cytomegalovirus (CMV) (36, 51). This difference could be due to a lower mutation rate, a more recent common ancestry, or a combination of the two. Most striking is the genetic uniformity within and between samples from Guinea Bissau, Zambia, Bangladesh, and Southern India. In pairwise comparisons between genotypes from these localities, a maximum of one substitution was found in the 1.9 kb surveyed. Although small samples were obtained from each locality, the broad geographical scope precludes a sampling artifact. The human populations in these countries are not thought to have a particularly recent common ancestry (6), so the low differentiation of VZV indicates possible spread subsequent to the initial colonization of these areas. The contrasting higher level of diversity in the United Kingdom and Brazil may be explained by the history of recent human migration. There is clear evidence from the United Kingdom in support of the hypothesis of viral spread associated with migration. In East London, the frequency of clade A viruses among cases of varicella has increased with time, during a period when immigrants from the Indian and African subcontinents have come to make up more than 25% of the population (22). Clade A viruses are characterized by a *Bgl*I restriction site in gene 54 and are ubiquitous in Asia, Africa, and the Far East but generally make up fewer than 10% of British strains (22).

One way of interpreting the differentiation between populations is to make use of it to estimate the time of divergence.

This estimate is conventionally obtained from the net accumulation of nucleotide diversity (π_a) after subtracting an estimate of the diversity within the ancestral populations, which is obtained from the current diversity (34). Unfortunately, this calculation is not appropriate for VZV, as we have evidence that the high diversity of current populations may be a consequence of mixing, which has not yet reached equilibrium. We instead obtain an alternative estimate of the time (t) since the splitting of the major clades rather than populations, from the number of synonymous substitutions that have accumulated between them. This calculation requires an estimate of the mutation rate. One possible approach is to make use of estimates of mutation rates obtained from related viruses. For example, a synonymous mutation rate of 10^{-7} (per year) has been obtained from differentiation between HSV type 1 strains from different parts of the world. It would imply that differentiation was initiated at the time of human dispersal out of Africa around 100 kya and is also consistent with rates estimated from other alphaherpesviruses (31, 32, 36). When this rate is applied to the VZV data, it leads to a much more recent estimate, implying that the currently circulating strains of VZV have spread in the past 3,000 to 19,000 years.

This interpretation is somewhat at odds with established estimates of varicella-zoster dispersal (19, 31, 32). Although the calculation allowed for some error on the estimate of the mutation rate (standard deviation, 20%), an alternative explanation is that the mutation rate of VZV is substantially lower than that of other alphaherpesviruses. Our calculations showed that if the rate were one-seventh of that of other alphaherpesviruses, then the different clades could date back 100,000 years. Is such a low rate plausible? Given the similar biology of the alphaherpesviruses, a lower mutation rate per division seems unlikely.

Is it then possible that VZV has a smaller number of divisions per year than other herpesviruses? The natural history of varicella is of primary infection in childhood following which the virus remains quiescent before reactivating once in adulthood, although recent evidence has confirmed observations that subclinical reactivation of VZV is not uncommon (30). The number of replication cycles before latency is established could be as low as 20 (19), and viral shedding is limited to the duration of lesions (approximately 5 to 7 days). By contrast, herpes simplex virus is known to have multiple reactivation episodes, which give rise to both symptomatic and asymptomatic viral shedding. Similarly CMV is 40 times more variable than VZV (51). This virus appears to be more distantly related to VZV than the alphaherpesviruses (32), but it is not sufficiently different that it is likely to be subject to a higher mutation rate per division. Instead, the same two explanations that we discussed for HSV may explain its higher diversity: a more ancient ancestry or a higher number of divisions since the common ancestor. The more ancient ancestry could be again explained by the mode of infection, which typically involves close contact. Furthermore, although CMV may be latent in mononuclear cells, there are long periods of asymptomatic shedding which suggest a high rate of division per unit time.

While the activity of VZV between varicella and zoster is not quantified, it is instructive to calculate the effect on the number of divisions per year under the assumption that this is an essentially quiescent period. The relevant transmission dynamics can be assessed using the parameters estimated by Schuette and Hethcote (39), which indicate that the majority of infections along any evolutionary lineage will be from varicella case to varicella case. The proportion of zoster episodes in the evolutionary history will be lower, first because of the lower relative infectivity (0.07) and second because only around 15% of cases reactivate as zoster. The latent period for varicella is only around 14 days, whereas the mean quiescent period before zoster activation is around 40 years. The ratio of time spent quiescent along an evolutionary lineage is given by $(n_z v t_z p_z + n_v t_v p_z)/(n_z v t_z + n_v t_v)$, where n_z and n_v are the proportion of the population with zoster and varicella, respectively, t_z and t_v are the periods between initial infection and transmission, and p_z and p_v are the proportion of time spent quiescent between infection and transmission ($p_z \approx 1$ and $p_v \approx 0$). The infectivity of zoster (relative to varicella) is given by v. The values of n depend on the population demographics and viral epidemiology of the population, but we estimate that the proportion of time spent quiescent is of the order of 96%. HSV may spend less time quiescent, and since the infectivity and interval of each episode are relatively constant, the proportion of quiescent time along an evolutionary lineage will correspond to that within an individual. It is plausible that the proportion of time quiescent could be several times smaller than for VZV. Although this explanation of low diversity in VZV may be credible, the issue is most likely to be resolved by a more accurate calibration of the difference in mutation rate between VZV and other herpesviruses.

Although rapid spread and a low mutation rate could both explain the low level of genetic diversity of VZV, the geographical distribution and relatedness of strains does suggest that the spread of VZV subsequent to the major human migrations has made a strong contribution to its geographic epidemiology. The validity of this hypothesis is underlined by the recent spread in London of clade A strains, which appears to coincide with the major period of immigration from the Indian subcontinent and Africa to the United Kingdom. Such rapid spread is possible because VZV, uniquely among herpesviruses, is transmitted by aerosolization of virus, resulting in airborne epidemic infection. Geographic spread may be particularly rapid in warmer regions, where varicella occurs on average 10 years later in life (16). In this scenario, the disease is transmitted between individuals who are more mobile and who mix more freely than the children who account for the majority of infectious cases in more temperate climates.

Significant associations (linkage disequilibrium) between SNPs that are widely spaced within the genome suggest low recombination rates. Recombination rates, however, cannot be calculated from population genetic data because there is evidence of population subdivision, so that the different haplotypes may not come into contact. This study revealed that two geographic areas, London in the United Kingdom and Rio de Janeiro in Brazil, are populated with mixtures of strains. One haplotype originating with Brazil (Fig. 1) is most parsimoniously explained as a recombinant between A and C strains which coexist in the region. At least two other strains, UZ 42 and UZ 42+ (Fig. 1 and 2) from London, are also likely to be recombinants. Although in vitro recombination of the vaccine vOka strain with a wild-type U.S. strain has been described

(10), this is the first evidence of recombination between strains occurring in vivo.

Recombination of wild-type strains may have consequences for virulence and selection. For example, the MSP VZV strain, in which a single nucleotide substitution results in an amino acid change from aspartate to asparagine in glycoprotein E, appears to be more virulent in the skin SCID-hu mouse model (37, 38). Epidemic spread of VZV affords the opportunity for repeated and widespread exposure of immune and naïve populations to circulating virus, and reinfection is well described (2, 18). In these circumstances, cocirculation of distinct wild-type genotypes, such as occurs in East London, will allow measurement of the rate of recombination and assessment of its contribution to VZV evolution. Such data may prove useful in understanding the spread and pathogenesis of VZV, especially against a background of mass immunization with the attenuated Oka vaccine.

ACKNOWLEDGMENTS

This work was supported by grants from North Thames Health Authority Research & Development and the Special Trustees of St. Bartholomew's Hospital.

We thank colleagues who provided samples for this study, including Yamima Talukdar (Bangladesh), Marilda Siqueira (Brazil), Peter Aaby (Guinea Bissau), Koichi Yamanashi (Japan), Jacob John (South India), and Patrick Mtondo (Zambia).

REFERENCES

1. **Adams, S. G., D. E. Dohner, and L. D. Gelb.** 1989. Restriction fragment differences between the genomes of the Oka varicella vaccine virus and American wild-type varicella-zoster virus. J. Med. Virol. **29:**38–45.
2. **Arvin, A. M., C. M. Koropchak, and A. E. Wittek.** 1983. Immunologic evidence of reinfection with varicella-zoster virus. J. Infect. Dis. **148:**200–205.
3. **Barrett-Muir, W., K. Hawrami, J. Clarke, and J. Breuer.** 2001. Investigation of varicella-zoster virus variation by heteroduplex mobility assay. Arch. Virol. **17**(Suppl. 1)**:**17–25.
4. **Breuer, J.** 1997. Oka vaccine for VZV. Herpes **4:**62–67.
5. **Casey, T. A., W. T. Ruyechan, M. N. Flora, W. Reinhold, S. E. Straus, and J. Hay.** 1985. Fine mapping and sequencing of a variable segment in the inverted repeat region of varicella-zoster virus DNA. J. Virol. **54:**639–642.
6. **Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza.** 1994. The history and geography of human genes. Princeton University Press, Princeton, N.J.
7. **Chou, S., and K. M. Dennison.** 1991. Analysis of interstrain variation in cytomegalovirus glycoprotein B sequences encoding neutralizing epitopes. J. Infect. Dis. **163:**1229–1234.
8. **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.
9. **Davison, A. J., and J. E. Scott.** 1986. The complete DNA sequence of varicella-zoster virus. J. Gen. Virol. **67:**1759–1816.
10. **Dohner, D. E., S. G. Adams, and L. D. Gelb.** 1988. Recombination in tissue culture between varicella-zoster virus strains. J. Med. Virol. **24:**329–341.
11. **Dumas, A. M., J. L. Geelen, M. W. Weststrate, P. Wertheim, and J. van der Noordaa.** 1981. *Xba*I, *Pst*I, and *Bgl*I restriction enzyme maps of the two orientations of the varicella-zoster virus genome. J. Virol. **39:**390–400.
12. **Faga, B., W. Maury, D. A. Bruckner, and C. Grose.** 2001. Identification and mapping of single nucleotide polymorphisms in the varicella-zoster virus genome. Virology **280:**1–6.
13. **Felsenstein, J.** 1993. PHYLIP inference package, version 3.5. Department of Genetics, University of Washington, Seattle.
14. **Franti, M., J.-T. Aubin, L. Poirel, A. Gautheret-Dejean, D. Candotti, J.-M. Huraux, and H. Agut.** 1998. Definition and distribution analysis of glycoprotein B gene alleles of human herpesvirus 7. J. Virol. **72:**8725–8730.
15. **Ganguly, A., M. J. Rock, and D. J. Prockop.** 1993. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **90:**10325–10329.
16. **Garnett, G. P., M. J. Cox, D. A. P. Bundy, J. M. Didier, and J. St. Catharine.** 1993. The age of infection with varicella-zoster virus in St-Lucia, West-Indies. Epidemiol. Infect. **110:**361–372.
17. **Gelb, L. D., D. E. Dohner, A. A. Gershon, et al.** 1987. Molecular epidemiology of live, attenuated varicella virus vaccine in children and in normal adults. J. Infect. Dis. **155:**633–640.
18. **Gershon, A. A., S. P. Steinberg, and L. Gelb.** 1984. Clinical reinfection with varicella-zoster virus. J. Infect. Dis. **149:**137–142.
19. **Grose, C.** 1999. Varicella-zoster virus: less immutable than once thought. Pediatrics **103:**1027–1028.
20. **Hawrami, K., D. Harper, and J. Breuer.** 1996. Typing of varicella zoster virus by amplification of DNA polymorphisms. J. Virol. Methods **57:**169–174.
21. **Hawrami, K., and J. Breuer.** 1997. Analysis of United Kingdom wild-type strains of varicella-zoster virus: differentiation from the Oka vaccine strain. J. Med. Virol. **53:**60–62.
22. **Hawrami, K., I. J. Hart, F. Pereira, S. Argent, B. Bannister, B. Bovill, D. Carrington, M. Ogilvie, S. Rawstorne, Y. Tryhorn, and J. Breuer.** 1997. Molecular epidemiology of varicella-zoster virus in East London, England, between 1971 and 1995. J. Clin. Microbiol. **35:**2807–2809.
23. **Hayakawa, Y., S. Torigoe, K. Shiraki, K. Yamanishi, and M. Takahashi.** 1984. Biologic and biophysical markers of a live varicella vaccine strain (Oka): identification of clinical isolates from vaccine recipients. J. Infect. Dis. **149:**956–963.
24. **Hondo, R., and Y. Yogo.** 1988. Strain variation of R5 direct repeats in the right-hand portion of the long unique segment of varicella-zoster virus DNA. J. Virol. **62:**2916–2921.
25. **Hondo, R., Y. Yogo, M. Yoshida, A. Fujima, and S. Itoh.** 1989. Distribution of varicella-zoster virus strains carrying a PstI-site-less mutation in Japan and DNA change responsible for the mutation. Jpn. J. Exp. Med. **59:**233–237.
26. **Kinchington, P. R., J. Remenick, J. M. Ostrove, S. Straus, W. D. Ruyechan, and J. Hay.** 1986. Putative glycoprotein gene of varicella-zoster virus with variable copy numbers of a 42-base-pair repeat sequence has homology to herpes simplex virus glycoprotein C. J. Virol. **59:**660–668.
27. **Kinchington, P. R., and S. E. Turse.** 1995. Molecular basis for a geographic variation of varicella-zoster virus recognized by a peptide antibody. Neurology **45**(Suppl. 8)**:**S13-S14.
28. **Kinoshita, H., R. Hondo, F. Taguchi, and Y. Yogo.** 1988. Variation of R1 repeated sequence present in open reading frame 11 of varicella-zoster virus strains. J. Virol. **62:**1097–1100.
29. **La Russa, P., O. Lungu, I. Hardy, A. Gershon, S. P. Steinberg, and S. Silverstein.** 1992. Restriction fragment length polymorphism of polymerase chain reaction products from vaccine and wild-type varicella-zoster virus isolates. J. Virol. **66:**1016–1020.
30. **Mainka, C., B. Fuss, H. Geiger, H. Hofelmayr, and M. H. Wolff.** 1998. Characterization of viremia at different stages of varicella-zoster virus infection. J. Med. Virol. **56:**91–98.
31. **McGeoch, D. J., and S. Cook.** 1994. Molecular phylogeny of the Alphaherpesvirinae subfamily and a proposed evolutionary timescale. J. Mol. Biol. **238:**9–22.
32. **McGeoch, D. J., S. Cook, A. Dolan, F. E. Jamieson, and E. A. Telford.** 1995. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. J. Mol. Biol. **247:**443–458.
33. **Meng, Y. X., T. J. Spira, G. J. Bhat, C. J. Birch, J. D. Druce, B. R. Edlin, R. Edwards, C. Gunthel, R. Newton, F. R. Stamey, C. Wood, and P. E. Pellett.** 1999. Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. Virology **261:**106–119.
34. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, N.Y.
35. **Rixon, F. J., and D. J. McGeoch.** 1985. Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. Nucleic Acids Res. **13:**953–973.
36. **Sakaoka, H., K. Kurita, Y. Lida, S. Takada, K. Umene, Y. Tae Kim, C. Shang 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.
37. **Santos, R. A., J. A. Padilla, C. Hatfield, and C. Grose.** 1998. Antigenic variation of varicella zoster virus Fc receptor gE: loss of a major B cell epitope in the ectodomain. Virology **249:**21–31.
38. **Santos, R. A., C. C. Hatfield, N. L. Cole, J. A. Padilla, J. F. Moffat, A. M. Arvin, W. T. Ruyechan, J. Hay, and C. Grose.** 2000. Varicella-zoster virus gE escape mutant VZV-MSP exhibits an accelerated cell-to-cell spread phenotype in both infected cell cultures and SCID-hu mice. Virology **275:**306–317.
39. **Schuette, M. C., and H. W. Hethcote.** 1999. Modelling the effects of varicella vaccination programs on the incidence of chickenpox and shingles. Bull. Math. Biol. **61:**1031–1064.
40. **Shepp, D. H., M. Match, S. M. Lipson, and R. G. Pergolizzi.** 1998. A fifth human cytomegalovirus glycoprotein B genotype. Res. Virol. **149:**109–114.
41. **Straus, S. E., J. Hay, H. Smith, and J. Owens.** 1983. Genome differences among varicella-zoster virus isolates. J. Gen. Virol. **64:**1031–1041.
42. **Straus, S. E., W. Reinhold, H. A. Smith, W. T. Ruyechan, D. K. Henderson, R. M. Blaese, and J. Hay.** 1984. Endonuclease analysis of viral DNA from varicella and subsequent zoster infections in the same patient. N. Engl. J. Med. **311:**1362–1364.
43. **Takada, M., T. Suzutani, I. Yoshida, M. Matoba, and M. Azuma.** 1995. Identification of varicella-zoster virus strains by PCR analysis of three repeat elements and a PstI-site-less region. J. Clin. Microbiol. **33:**658–660.
44. **Takayama, M., N. Takayama, N. Inoue, and Y. Kameoka.** 1996. Application

of long PCR method to identification of variations in nucleotide sequences among varicella-zoster isolates. J. Clin. Microbiol. **34:**2869–2874.
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. **Umene, K., and H. Sakaoka.** 1997. Populations of two eastern countries of Japan and Korea and with a related history share a predominant genotype of herpes simplex virus type 1. Arch. Virol. **142:**1953–1961.
47. **Williams, D. L., A. Gershon, L. D. Gelb, M. K. Spraker, S. Steinberg, and A. H. Ragab.** 1985. Herpes zoster following varicella vaccine in a child with acute lymphocytic leukaemia. J. Paediatr. **106:**259–261.
48. **Zimber, U., H. K. Adldinger, G. M. Lenoir, M. Vuillaume, M. V. Knebel-Doeberitz, G. Laux, C. Desgranges, P. Wittman, U. K. Freese, U. Schneider, and G. Bornkamm.** 1986. Geographical prevalence of two types of Epstein-Barr virus. Virology **154:**56–66.
49. **Zipeto, D., C. Hong, G. Gerna, M. Zavattoni, D. Katzenstein, T. C. Merigan, and L. Rasmussen.** 1998. Geographic and demographic differences in the frequency of human cytomegalovirus gB genotypes 1–4 in immunocompromised patients. AIDS Res. Hum. Retrovir. **14:**533–536.
50. **Zong, J.-C., D. M. Ciufo, D. J. Alcendor, X. Wan, J. Nicholas, P. J. Browning, P. L. Rady, S. K. Tyring, J. M. Orenstein, C. S. Rabkin, I.-J. Su, K. F. Powell, M. Croxson, K. E. Foreman, B. J. Nickoloff, S. Alkan, and G. S. Hayward.** 1999. High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major subtypes and multiple variants or clades in different human populations. J. Virol. **73:**4156–4170.
51. **Zweygberg, L., B. Wirgart, M. Brytting, A. Linde, B. Wahren, and L. Grillner.** 1998. Sequence variation within three important cytomegalovirus gene regions in isolates from four different patient populations. J. Clin. Microbiol. **36:**3662–3669.