Deep Sequencing of Viral Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans

Daniel P. Depledge,*[,1] Samit Kundu,[1] Nancy J. Jensen,[2] Eleanor R. Gray,[1] Meleri Jones,[1] Sharon Steinberg,[3] Anne Gershon,[3] Paul R. Kinchington,[4] D. Scott Schmid,[2] Francois Balloux,[5] Richard A. Nichols,[6] and Judith Breuer[1]

[1]MRC Centre for Medical Molecular Virology, Division of Infection and Immunity, London, United Kingdom

[2]National VZV Reference Laboratory, CDC, Atlanta, GA

[3]Division of Infectious Disease, Columbia University Medical Centre, New York

[4]Departments of Opthamology, Microbiology and Molecular Genetics, University of Pittsburgh

[5]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

[6]School of Biological and Chemical Sciences, Queen Mary, University of London, London, United Kingdom

***Corresponding author:** E-mail: d.depledge@ucl.ac.uk.

Associate editor: Helen Piontkivska

Abstract

Immunization with the vOka vaccine prevents varicella (chickenpox) in children and susceptible adults. The vOka vaccine strain comprises a mixture of genotypes and, despite attenuation, causes rashes in small numbers of recipients. Like wild-type virus, the vaccine establishes latency in neuronal tissue and can later reactivate to cause Herpes zoster (shingles). Using hybridization-based methodologies, we have purified and sequenced vOka directly from skin lesions. We show that alleles present in the vaccine can be recovered from the lesions and demonstrate the presence of a severe bottleneck between inoculation and lesion formation. Genotypes in any one lesion appear to be descended from one to three vaccine-genotypes with a low frequency of novel mutations. No single vOka haplotype and no novel mutations are consistently present in rashes, indicating that neither new mutations nor recombination with wild type are critical to the evolution of vOka rashes. Instead, alleles arising from attenuation (i.e., not derived from free-living virus) are present at lower frequencies in rash genotypes. We identify 11 loci at which the ancestral allele is selected for in vOka rash formation and show genotypes in rashes that have reactivated from latency cannot be distinguished from rashes occurring immediately after inoculation. We conclude that the vOka vaccine, although heterogeneous, has not evolved to form rashes through positive selection in the mode of a quasispecies, but rather alleles that were essentially neutral during the vaccine production have been selected against in the human subjects, allowing us to identify key loci for rash formation.

Key words: **alphaherpesvirus, viral evolution, pathogenesis.**

Article

Introduction

The live varicella vaccine is widely used for prevention of chickenpox (varicella), caused by the ubiquitous herpesvirus varicella zoster virus (VZV) (Sharrar et al. 2000; Rentier and Gershon 2004). The vaccine was derived by standard growth attenuation techniques of a clinical virus isolate Oka, in semi permissive culture prior to commercial development and use (Takahashi et al. 1975). The vaccine virus is heterogeneous and has accumulated more than 40 mutations during attenuation, although only three of these have become universally fixed (Gomi et al. 2002; Quinlivan et al. 2004, 2006; Loparev et al. 2007; Breuer and Schmid 2008; Levin et al. 2008). The vaccine therefore contains a mixture of related viruses, comprising a large number of unique haplotypes. In 2–3% of cases, vaccination is followed by clinically visible rashes or skin lesions that are frequently varicella-like, occurring 6–42 days after inoculation (Sharrar et al. 2000; Goulleret et al. 2010). Although vaccine varicella rashes are typically mild, these occasionally become severe, particularly in children with underlying disorders of innate or adaptive immunity (Kramer et al. 2001; Levy et al. 2003; Chaves et al. 2008; Banovic et al. 2011). Like wild-type VZV strains (albeit with lower frequency), the vOka strain can also establish latency in the dorsal root ganglia and is capable of reactivating months or years later to cause herpes zoster that is often indistinguishable from that caused by wild-type virus (Civen et al. 2009).

We exploit the unusual heterogeneity of this DNA virus vaccine to examine the influence of population bottlenecks after inoculation, on rash formation, and reactivation from latency. As the vOka vaccine strain is known to be attenuated for replication in skin (Moffat et al. 1998; Arvin 2001; Chen et al. 2003), rash formation following vaccination could represent reversion toward the wild phenotype or even

recombination with wild-type viruses. To enable whole genome sequencing of vOka virus from opportunistically collected clinical samples of vaccine-associated rashes, we have adapted a targeted enrichment methodology (shown previously to recover low copy number virus genomes) that, unlike methods requiring culture or polymerase chain reaction (PCR) amplification, does not alter viral population diversity (Depledge et al. 2011). In this article, we show that the virus present in vaccine rashes derives from haplotypes present in the original vaccine mixture and that haplotypes cannot be segregated by pathogenesis or neurotropism (i.e., those which enter latency and later reactivate). Moreover, we show that the frequency of variants in the original vaccine and associated rashes are preserved by the targeted enrichment method, unless the sample has previously been cultured. Finally, we identify both previously undetected alleles segregating in the vaccine and novel mutations that contribute to viral diversity in skin lesions. Thus, comparison of whole vOka genome sequences from varicella vaccine preparations, with sequences obtained from vaccine varicella and zoster lesions provides insight into VZV evolution in vaccinated individuals.

Results

Optimization of Targeted Enrichment

Archived diagnostic samples of vesicular fluid contain small amounts (20–150 ng) of extracted total DNA and required whole-genome amplification (WGA) using the Φ29 polymerase to generate sufficient starting material (2–3 μg) for the SureSelect (SS) custom target capture methodology (Depledge et al. 2011). Illumina sequencing was performed on 20 vaccine rash samples collected between 1988 and 2007 and three distinct batches of the VariVax (Merck) VZV vaccine from 2008, 2010, and 2012 (table 1). To determine whether any bias was associated with these methods, we prepared five sequence libraries (WGA only, SS only, an SS only replicate, WGA + SS, and direct sequencing without WGA or SS; supplementary table S1, Supplementary Material online) from a single batch of VZV vaccine. The consensus sequences for each sample were identical, while the correlation of allele frequencies between libraries was high (Pearson's product moment $R^2 > 0.93$) at sites where total read depth was over 50 reads per base and where the variant allele was present on two or more independent reads mapping to opposite strands (fig. 1). Deviations in the variant allele frequency of up to 3.2% were observed but these were limited to comparisons between enriched and nonenriched libraries (fig. 1). This is explained by the fact that read depths in nonenriched libraries are approximately 20-fold. A comparison of replicate samples that were independently enriched showed the standard deviation from the mean allele frequency, that is, the error of the method to be 0.654%.

Viral DNA was enriched from all vaccine and vaccine-rash samples with a median of 87.3% on-target reads (table 1). Read data aligned against the pOKA reference genome (Gomi et al. 2002) showed a bias against enrichment in GC-rich regions of the variable region R1–R5 and short region genomic repeated elements (supplementary fig. S1, Supplementary Material online). Mean read depths were consistently higher than 1,000 reads per base for all samples except the nonenriched vaccine. Genome coverage exceeded 97%, except in two cases; samples O27 (90.5%) and K48 (83.8%). The paired-end sequence reads were too short to adequately map the variable R1–R5 regions and for the majority of the rashes there was insufficient material to fill these gaps using PCR and Sanger methods. Sequencing of vaccine rash samples on the Illumina GAIIx platform gave significantly more low frequency polymorphic sites (defined as having a variant allele at <10%) in vaccine rash samples when compared with samples sequenced on an Illumina MiSeq (supplementary fig. S2, Supplementary Material online). Resequencing of three samples for which sufficient DNA was available (K11, L53, and N13) on both platforms showed the polymorphisms present in the MiSeq data set to be a subset of those present in the GAIIx data set with good correlation between like data sets ($K11 - r^2 = 0.897$, $N13 - r^2 = 0.908$, and $L53 - r^2 = 0.897$; supplementary fig. S3, Supplementary Material online). In all cases, the excess of polymorphic sites in data sets from the GAIIx did not alter the consensus sequences for these samples but do confirm that a higher sequencing error rate is associated with the GAIIx, a finding that has been previously reported (Minoche et al. 2011; Loman et al. 2012; Quail et al. 2012).

Consensus Sequence Analyses

Animal alphaherpesvirus vaccines have been shown to recombine with circulating wild type and other attenuated vaccine viruses with the potential for restoration of viral virulence (Lee et al. 2012). To exclude this in vOka-based viruses recovered from rashes, Bayesian (fig. 2) and Maximum-likelihood trees (not shown) of vaccine and rash consensus sequences were constructed (Barrett-Muir et al. 2003; Loparev et al. 2007; Sengupta et al. 2007). We found no significant evidence for the vaccine rashes representing recombination with wild-type strains. BaTs analysis of the consensus phylogeny by year of vaccination excluded clustering of rash sequences by, mode of sample preparation (AI $P = 0.36$, MS $P = 1.0$, MC [high passage in vitro culture] $P = 1.0$, MC [low passage culture] $P = 0.45$, and MC [no culture] $P = 1.0$) or disease presentation (zoster or varicella: AI $P = 0.85$, PS $P = 0.99$, MC [varicella] $P = 1.0$, and MC [zoster] $P = 0.64$). The last indicates that viruses that establish latency and reactivate are not limited to a subset of haplotypes within the population of vOka variants that cause rashes.

Conservation of Polymorphic Sites in Vaccine

All three sequenced vaccine batches were heterogeneous with between 235 and 336 polymorphic sites being identified. Of these, 224 were present in two or more vaccine preparations (supplementary table S2, Supplementary Material online) and 207 were shared between all three vaccine preparations, of which 36 had been previously identified by Sanger or pyrosequencing (Gomi et al. 2002; Loparev et al. 2007; Quinlivan et al. 2007; Tillieux et al. 2008; Kanda et al. 2011). These 25–120 single nucleotide polymorphisms (SNPs) were

Table 1. Samples used in this study.

Sample ID	Type	Platform	% OTR	% Genome Coverage at 50×	Country and Year of Vaccination	Days to Rash Formation
Varicella vaccine rash						
N13[a]	Uncultured	MiSeq	83.7	97.6	USA-1998	Unknown
VR1			89.4	99.9	Europe-2007	14
VR2			86.4	98.3	UK-2007	16
VR3			81.1	99.9	UK-2006	14
VR4		GAiix	71.0	99.9	UK-2010	Unknown
O27[a]			92.7	90.5	USA-1999	16
*N13**			92.3	96.8	USA-1998	Unknown
A182B[a]	High-passage	HiSeq	93.4	99.6	USA-1988	16
A185B[a]			97.2	99.5	USA-1988	21
Vaccine[b]						
VV10	NL13110	HiSeq	66.8	99.9	UK-2010	N/A
VV12	G007544	MiSeq	89.5	100.0	UK-2012	N/A
VVAG[a]	1526X[a]		88.1	99.9	USA-2009	N/A
Zoster vaccine rash						
K11	Uncultured	MiSeq	79.7	99.8	USA-1997	74
L53[a]			70.5	98.4	USA-1997	131
ZR1			91.0	99.3	UK-2006	330
T61		GAiix	94.8	99.0	USA-2001	90
K11			93.4	98.9	USA-1997	75
L53[a]			97.2	99.4	USA-1997	131
U14[a]			73.1	98.2	USA-1997	>1,500
K48[a]			18.1	83.8	USA-1995	603
R73[a]			9.6	98.0	USA-unknown	Unknown
R3[a]	Low-passage		57.3	98.0	USA-1999	244
R52[a]			94.8	98.7	USA-1999	490
T17[a]			95.7	99.5	USA-2000	310
T25[a]			60.4	99.4	USA-2000	547
v76			79.3	99.5	USA-1982	630

NOTE.—OTR, on-target reads (i.e., reads mapping to pOka reference genome). Samples in italics were sequenced on both MiSeq and GAiix platforms.
[a]Underwent WGA prior to SS.
[b]Note that all vaccine rashes are derived from VariVax batches used to prevent varicella.

limited to single vaccine batches of which 90% were low frequency (<10%). Five sites (12694 and 12779 [open reading frame {ORF} 10], 31732 [ORF 21], 82225 [ORF 42], and 106710 [ORF 62]) previously reported as polymorphic in Merck vaccine batches were fixed for the wild-type allele in all batches sequenced here (Gomi et al. 2002; Loparev et al. 2007; Quinlivan et al. 2007; Tillieux et al. 2008). The distribution of nonsynonymous, synonymous, and noncoding mutations is shown in figure 3. The vaccine allele frequencies at each of the 224 sites shared between vaccines were highly conserved (Pearson's product moment $R^2 > 0.98$), indicating minimal batch to batch variation at vaccine loci despite each having been issued in different years and in one case a different country (table 1; supplementary fig. S4*A*–*C*, Supplementary Material online). Therefore, we infer that the vaccines sequenced in this study are genetically compared with the batches used to inoculate the subjects whose rash viruses we analyze here. This proposal is also supported by the high correlation (Pearson's product moment $R^2 = 0.83$) between the mean allele frequencies in the vaccines and rash samples (supplementary fig. S4*D*, Supplementary Material online).

Vaccine Rashes Are Mono- or Oligomorphic

Because of the increased proportion of erroneous polymorphic sites introduced by GAIIx sequencing, only the variant data from rashes sequenced on the MiSeq were analyzed further (i.e., GAIIx data were only included in consensus level analyses). Between 32 and 112 polymorphisms per virus were identified in these six uncultured vaccine rashes, when compared with 235–336 in the vaccine batches. The lower genetic diversity of rashes can also be quantified by the proportion of polymorphic sites, being 0.19% for the vaccine, which is 2–4 times greater than rash diversity (0.03–0.09%). In vitro tissue culture further reduced rash virus diversity even at ≤3 passages (supplementary fig. S5, Supplementary Material online). Of the polymorphisms in the vaccine rashes, 37–93% (mean 55%) was at the sites of 224 vaccine SNPs (loci polymorphic in two or more vaccine samples). Vaccine-allele frequencies at the majority of these sites were lower than in the

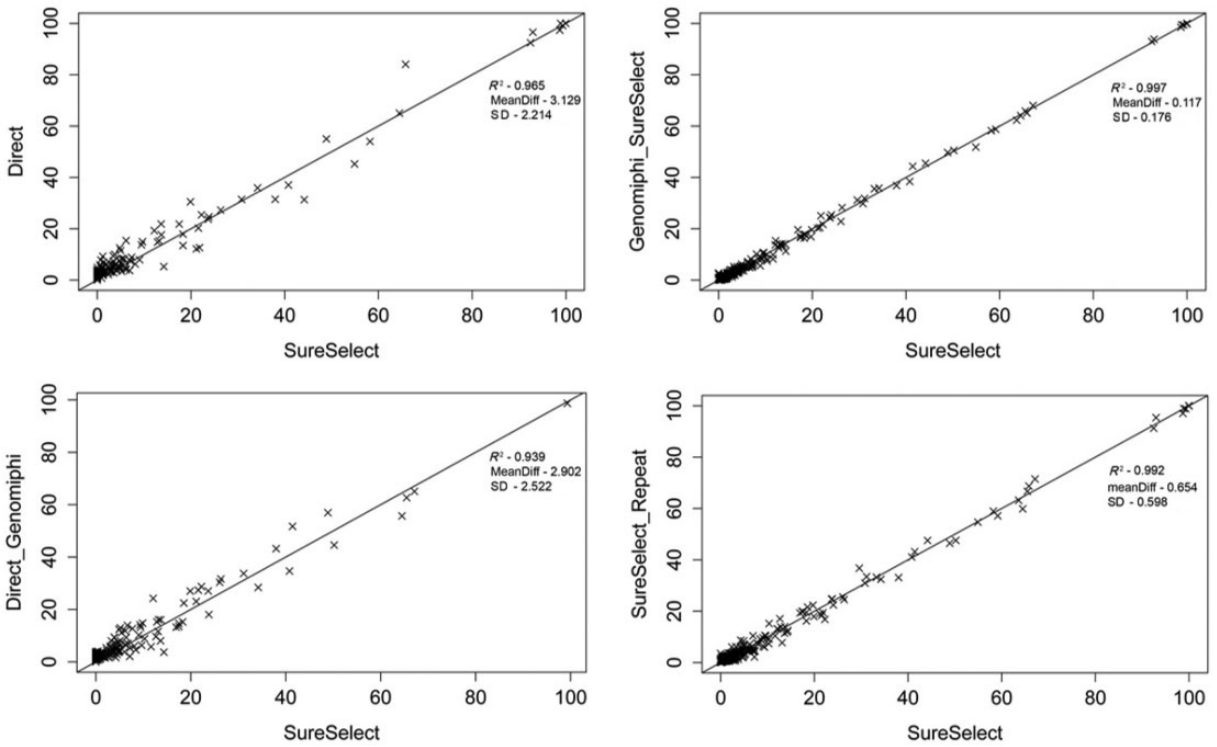

Fig. 1. WGA and target capture methodologies do not bias population structures. Comparative analyses of allele frequencies at polymorphic sites between sequencing samples prepared using different methodologies from the same source DNA show high correlation scores and minimal deviations in frequency. Comparisons are as follows: (*A*) Direct sequencing (no enrichment) versus SS (enrichment); (*B*) WGA and SS versus SS; (*C*) WGA followed by direct sequencing versus SS; and (*D*) SS versus SS (repeat).

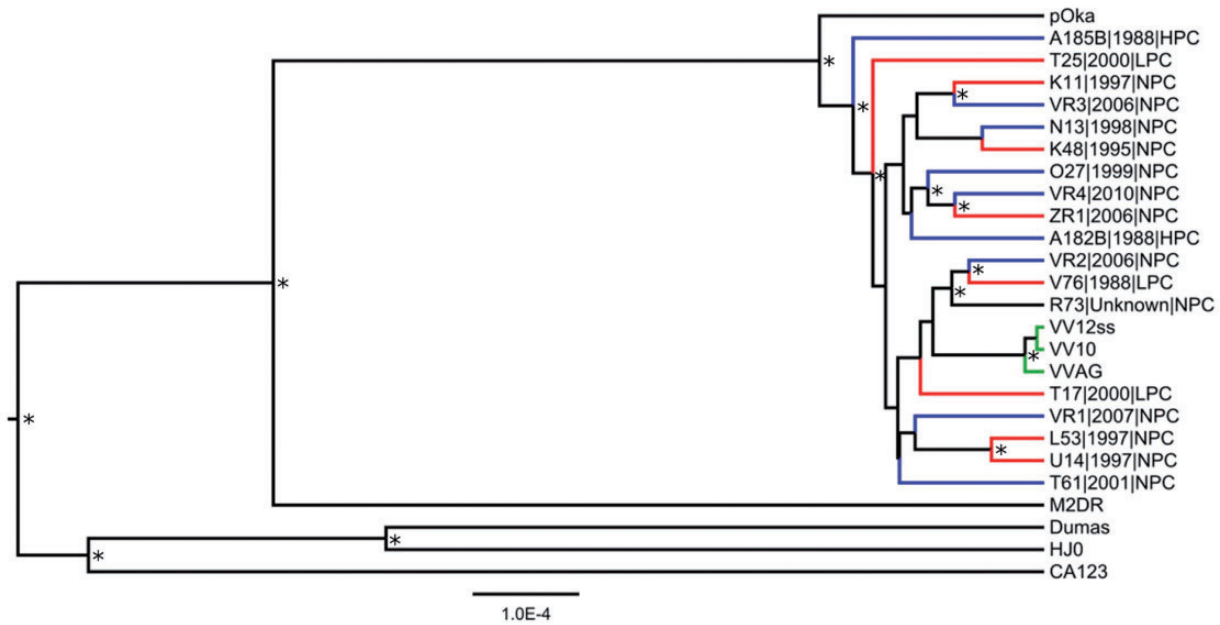

Fig. 2. Vaccine batches and vaccine rash samples form a monophyletic group with parental pOka strain. The Bayesian phylogenetic tree was inferred using BEAST and the topology was similar to trees generated using maximum likelihood or constructed using the ORF sequences only (data not shown). Vaccine batches are indicated by green lines, Vaccine rash sequences are derived from either varicella-like rashes (blue) or zoster-like rashes (red). Vaccine rash names indicated the sample code (i.e., K11), the year of isolation (i.e., 1997), and the samples status (NPC, uncultured; LPC, minimal culturing, i.e., less than five passages; and HPC, more than five passages in culture). Representative wild-type VZV sequences from each geographical clade were also included (pOka, Dumas, M2DR, HJO, and CA123).

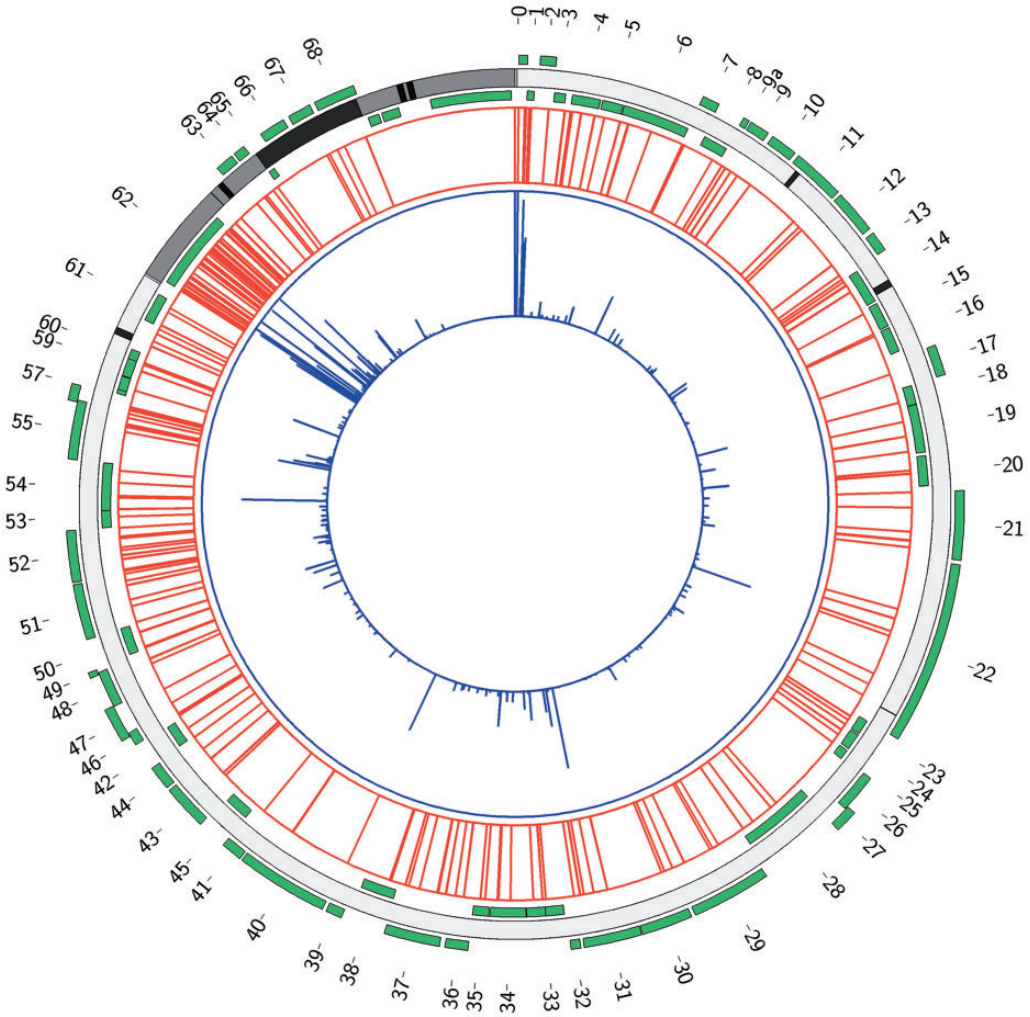

FIG. 3. Vaccine polymorphic sites are present throughout the genome with notable clusters in ORFS 0, 1, and 62. The circos genome map for vOka show the relative position of ORFs (green) in the U_L (light gray), U_S (dark gray), and I_R/T_R (mid gray) regions. Iterative repeat regions R1–R5 are shown as black bars. The 224 sites (indicated by red bars) at which both parental and vaccine alleles are present appear throughout the genome. The frequency (0–100%) of the vaccine allele is shown (in blue) at each site. Note that the T_R region has been left blank as this is an exact repeat of the I_R.

original vOka vaccines (supplementary table S2 and fig. S6, Supplementary Material online). These data are consistent with the conclusion that there exist one or more bottlenecks after inoculation of vaccine and the formation of rashes. There is a low level of polymorphism at novel loci which can be attributed to mutation because of the bottleneck. The rashes sequenced here were fixed for the vaccine allele at four amino acid coding positions in ORF 62 (105705, 106262, 107252, and 108111) (supplementary table S2, Supplementary Material online) and their presence may explain the mild presentation of most vaccine rashes compared with wild-type rashes (Tsolia et al. 1990; Cohen 2007). In addition to the 224 vaccine SNPs, we identified a further 39 polymorphic sites present in at least one vaccine and one vaccine rash and two present in two or more vaccine rashes but absent in the vaccines(supplementary table S3, Supplementary Material online). As we know that none of the rashes arose from the same VariVax vaccine batch (table 1), the most likely explanation is that for these 41 sites, the vaccine allele is present in the original vaccine at frequencies that are at the limit of detection by deep sequencing and therefore are not always detected as polymorphic.

De Novo Mutations Occurring after Inoculation

A total of 33 de novo mutations were identified in the consensus sequences of all rashes. Sixteen of these were nonsynonymous and occurred in eight rashes (VR1, VR2, A182B, L53, O27, T25, T17, and T61; table 2). The fixed mutations in rash A182B, a highly passaged isolate, for which plenty of DNA was available, were confirmed by Sanger sequencing. In the six rashes sequenced by the MiSeq method (supplementary table S4, Supplementary Material online), 67% of the new mutations that were not fixed were largely present at allele frequencies of 10% or less and the resequencing of K11, N13, and L53 confirmed of 62/68 (91.2%) of the new mutations in these rashes.

Selection for Rash-Forming Alleles

We have shown the segregating allele frequencies in the Merck vOka vaccine are sufficiently nonvariable between

Table 2. De novo mutations occurring post-inoculation.

Position (Dumas)	Wild Type	Vaccine	ORF	Amino Acid	Type	Rash ID
89	T	C	N/A	N/A	Noncoding	T61
967	G	A	N/A	N/A	Noncoding	T61
6420	G	A	ORF 6	A720T	Nonsynonymous	T61
10540	C	T	ORF 8	A43V	Nonsynonymous	T25
12588	A	G	ORF 10	P143	Nonsynonymous	T17
21595	C	A	ORF 15	T295K	Nonsynonymous	A182
23756	G	T	ORF 17	M13I	Nonsynonymous	A182
37959	G	A	ORF 22	D1293N	Nonsynonymous	T61
38111	A	C	ORF 22	T1343	Nonsynonymous	T61
41323	G	A	ORF 22	E2468K	Nonsynonymous	L53
44265	T	C	ORF 25	I37V	Nonsynonymous	VR1
51190	G	A	ORF 29	A112T	Nonsynonymous	O27
51206	G	A	ORF 29	R117Q	Nonsynonymous	T61
53993	G	A	ORF 29	R1046Q	Nonsynonymous	T61
61629	C	T	ORF 33	S170	Synonymous	T61
62811	T	C	ORF 34	T273	Synonymous	K11
65074	C	T	ORF 36	Q90*	Nonsynonymous	T61
67603	A	G	ORF 37	S510	Synonymous	A185
72395	G	A	ORF 40	D286N	Nonsynonymous	T61
79183	G	A	ORF 43	V431I	Nonsynonymous	VR2
80298	G	T	N/A	N/A	Noncoding	T17
81940	G	A	ORF 42	L218	Synonymous	T61
82225	A	G	ORF 42	P123	Synonymous	V76
98893	T	C	ORF 56	L109P	Nonsynonymous	T61
102976	C	T	N/A	N/A	Noncoding	R73
106497	T	C	ORF 62	G879	Synonymous	A182
109139	T	C	N/A	N/A	Noncoding	A182
109749	T	G	N/A	N/A	Noncoding	A185
109751	T	G	N/A	N/A	Noncoding	A185
109752	T	G	N/A	N/A	Noncoding	A185
109754	A	G	N/A	N/A	Noncoding	A185
109927	A	G	N/A	N/A	Noncoding	A182
117055	C	A	ORF 68	T416	Synonymous	R52

batches, that most of the vaccine alleles can be considered to have been present in the particular vaccine batch inoculated into patients (supplementary fig. S5, Supplementary Material online). We carried out analyses to evaluate whether there was any consistent change in allele frequencies between vaccine and rashes. This analysis had to take into account the dramatic loss of genetic diversity in each rash, because it involves large allele-frequency changes from the vaccine. These large differences cause the average frequencies across all rashes to deviate from the vaccine frequencies: A pattern that could be misinterpreted as evidence of consistent trends, unless appropriate compensation is made. We used two approaches to allow for this effect: A generalized linear mixed model, and a generalized linear model with quasi-binomial error. Both approaches detected a trend for the wild-type allele to increase in frequency in the rashes ($P < 0.001$), the mixed effect model suggesting a significant effect only at the coding sites ($P < 0.001$ for nonsynonymous and synonymous sites), indicative of selection.

A second analysis was used to identify SNPs with the clearest evidence for selection: those at which the vaccine allele repeatedly changed in frequency, which could indicate loci with a role in rash formation. Of the 224 vaccine polymorphic sites, there was statistical power (≥80%) to perform a binomial analysis for 7: 59591 (ORF31), 94167 (ORF 54), 102002 (intergenic), 105356 (ORF 62), 106001 (ORF 62), 107797 (ORF 62), and 108838 (ORF 62). At six of these sites, five nonsynonymous and one noncoding, the vaccine allele decreased significantly in the rashes. The vaccine allele increased significantly at the remaining synonymous site (94167 [ORF 54]) (supplementary table S5, Supplementary Material online). Using rash genotyping data from the literature and results from SNP typing of material from 36 additional rashes, we confirmed there was selection in favor of the ancestral allele for three of the five nonsynonymous sites, 105356 and 107797 (both ORF 62) which have previously been reported (Quinlivan et al. 2007), and 106001 (ORF 62; fig. 4). We analyzed a further nine nonsynonymous, three noncoding and one synonymous sites (560 [ORF 0], 19036 [ORF 13], 58595 [ORF 31], 87306 [ORF 50], 89734 [ORF 51], 90318 [ORF 51], 97479 [ORF 55], 97748 [ORF 55], 105169 [intergenic], 107599 [ORF 62], 109137 [intergenic], 109200 [intergenic], 115269 [intergenic]) for which SNP typing data from the literature and/or genotyping of a further 36 rashes provided enough power to confirm a significant change in allele frequencies (supplementary table S5, Supplementary Material online). Of these, there was a significant decrease in the frequency of the vaccine allele at six nonsynonymous sites: 560 (ORF 0), 19036 (ORF 13), 58595 (ORF 31), 97748 (ORF 55), 97479 (ORF 55), and 105599 (ORF 62); two noncoding: 109137 (ORF 62 promoter) and 105169 (61/62 intergenic); and a significant increase in the vaccine allele at one synonymous site: 89734 in ORF 51 (fig. 4; supplementary table S5 [Supplementary Material online]). In addition to positions 105169, 105356, and 107797 (all ORF 62), which have previously been reported (Quinlivan et al. 2007), three of the loci under selection, sites 106001 (K1045E) in ORF 62 domain IV, 19036 (G199R) in ORF 13, and 97479 (V495A) in ORF 55 have not previously been identified as polymorphic in the vaccine. None of the nonsynonymous loci were associated with changes in silico in known or predicted HLA class 1 epitopes (data not shown).

To evaluate the severity of the population bottleneck, we provisionally attributed alleles with a frequency of less than 1% in rashes as postbottleneck mutations. Having discarded these allele frequencies, we modeled the size of the putative population bottleneck occurring in the vaccine rash after vaccination by fitting a beta-binomial distribution around the expectation of the vaccine (estimated from the mean allele frequencies of the three vaccines). One of the fitted parameters of the beta-binomial analysis represents the probability F that two alleles randomly drawn from within a population are identical by descent. The estimated F was 0.79 (0.77–0.81, 95% CI). Matching the allele frequency spectrum, and this value of F to the output from a stochastic model of ancestry under exponential growth (we found the best fitting growth rate corresponding to the observed F, using the MLE function of R (growth rate $r = 0.0703$, over 202 generations), we could explain the data with a model (supplementary fig. S7, Supplementary Material online) in which most rashes were likely to be descended from one or two genotypes

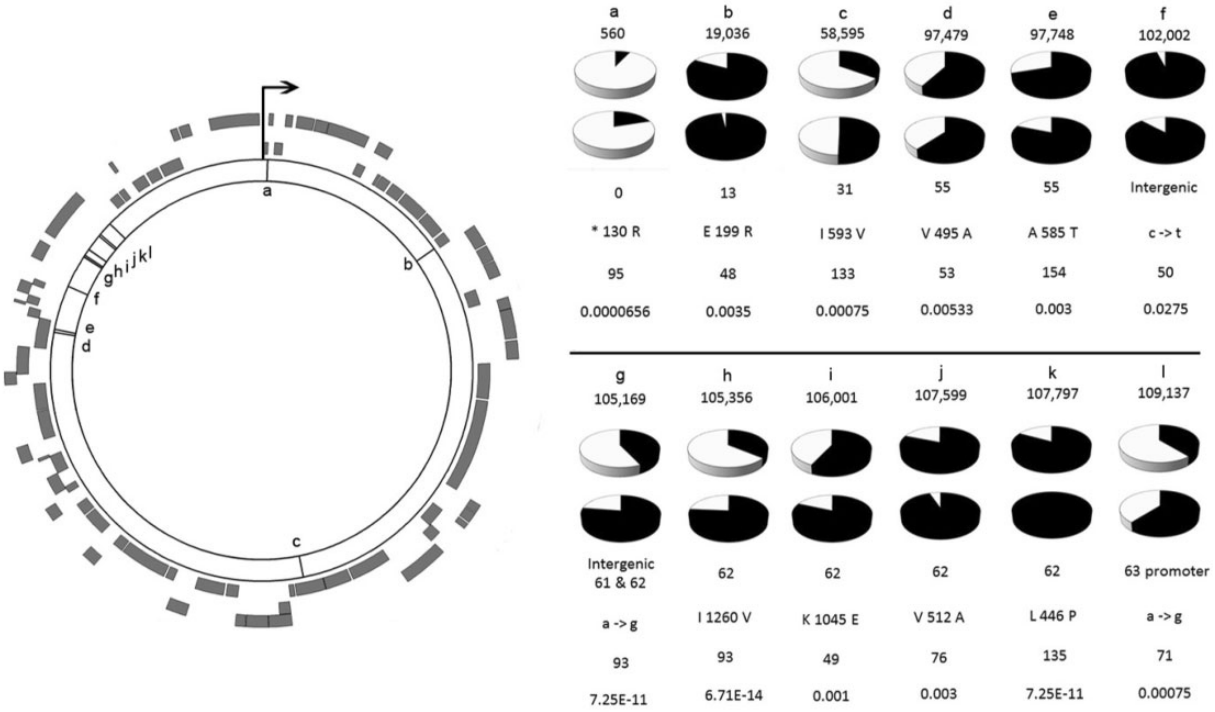

Fig. 4. Vaccine polymorphic sites under selection. Vaccine polymorphic sites under selection for the ancestral allele in vaccine rashes; 12 positions (9 nonsynonymous, 3 noncoding) are shown (*a*–*l*) in the circos genome plot with a close up of each site showing (top to bottom); the genome position of alleles (VZV Dumas coordinates), vaccine (upper pie chart), and vaccine rash (bottom pie chart) allele frequencies, the ORF affected, codon position and mutation, number of samples tested and (F) the FDR-corrected *P* value for each position. Note that a polymorphic site under selection at position 89734 is not shown in the figure as it encodes a synonymous mutation.

($P = 0.41$ and 0.51, respectively) and in which the majority of mutations (97%) occurred after bottleneck.

The availability of material from vOka varicella and zoster rashes together with the VariVax vaccine strain ancestral to both provided a unique opportunity to determine whether vOka rashes caused by viruses reactivating from latency differed genetically from those causing skin rashes directly, that is, within 42 days following inoculation (fig. 5). We found no significant differences in allele frequencies between viruses directly causing rash after vaccination (fig. 5, routes A and B) and those that established latency before reactivating to cause a zoster-rash (fig. 5, route C). The data presented here reflect only studies of the VariVax vaccine produced by Merck. Although all licensed vaccines are derived from the same attenuated vOka strain, the Biken and GSK vaccines have undergone different manufacturing processes. For example, the GSK vaccine was plaque purified and is known to have a higher percentage of vaccine SNPs overall (Kanda et al. 2011). Although it is reasonable to assume that the findings reported here apply to vaccine rashes derived from other manufacturer's vaccines, further sequencing is required to confirm this.

Discussion

We have exploited the current varicella immunization program in the United States and Europe to investigate the pathogenesis of the VZV vOka strain, a DNA virus that establishes latency in sensory neurons, and have identified a subset of vaccine loci at which selection of the ancestral allele influences the development of skin rashes. The VariVax varicella vaccine is known to be heterogeneous and genome sequencing in this study has identified at least 188 polymorphic sites additional to those previously recognized, of which 149 are present in all three sequenced vaccine preparations. Deep sequencing confirms that allele frequencies are highly correlated between temporally distinct batches (Kanda et al. 2011). Although lack of material prevented us sequencing the five iterative high G + C repeat regions in the genomes (R1–R5), these have not been found to differ from wild-type strains in prior analyses and so are considered unlikely to contribute to vaccine attenuation (Peters et al. 2006, 2012).

We show that following inoculation, the vaccine-associated alleles can persist in the viruses recovered from rashes. However, at most polymorphic vaccine loci, the rash virus is fixed for either the vaccine or the wild-type allele. A minority of vaccine loci remains polymorphic in rashes, and in most cases (72%) the minor allele is present at frequencies below 10% (fig. 3, supplementary table S2, Supplementary Material online). These data are consistent with a population bottleneck following vaccination. From the loss of polymorphism, it can be estimated that most genetic lineages ($F = 0.79$) are descended from the same ancestor in the bottleneck. Furthermore, in a model of exponential growth since the bottleneck, this value implies that a typical rash is descended from one to three ancestral genotypes. There is no evidence that the mutations that arise following inoculation are responsible for rash formation: they are rare and do not recur

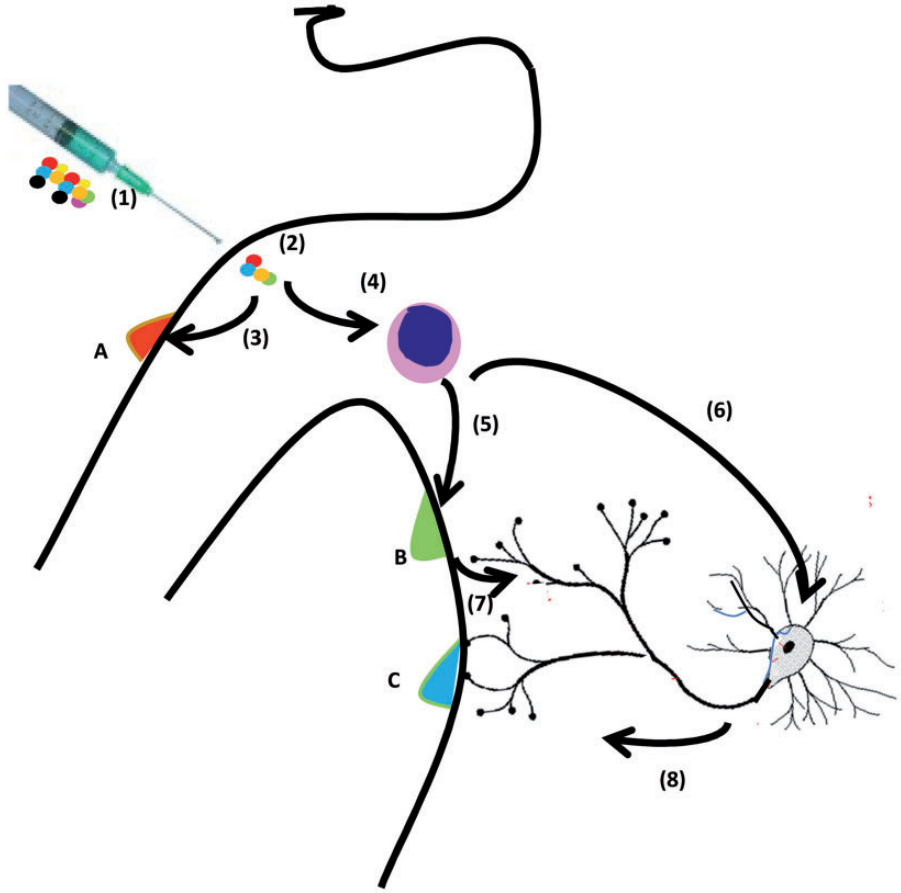

Fig. 5. Model of possible routes of vOka spread to the skin following inoculation. (1) Live-attenuated VZV vaccine, (2) replication at site of inoculation, (3) local spread to cause vaccine rash, (4) direct establishment of latency, (5) infection of T cell, (6) mediated transport to cause distal vaccine rash, (7) vireamic spread to establish latent ganglionic infection, (8) retrograde traversal of sensory neurons to establish latent ganglionic infection, and (9) reactivation from ganglia to cause zoster. (A), (B), and (C) are rash lesions caused by spread from the site of inoculation, hematogenous spread, and reactivation from latency. The genotypes of viruses recovered from these lesions cannot be distinguished by route of infection.

across the different rashes. Moreover, unlike some live-attenuated animal alphaherpesvirus vaccines, we see no evidence in our samples of recombination of vOka with wild-type virus in the rashes (Lee et al. 2012). As they occur at a low frequency in a population showing evidence of a severe bottleneck, we argue that low-level polymorphisms observed in the rashes are most likely to have arisen as mutations during replication of the vOka virus during the development of a vesicular lesion. From the allele frequency spectra, we estimated that 97% of mutations occur after bottleneck. Intriguingly, the rash genotypes show general evidence of selection against the vaccine allele: an increase in frequency for the ancestral allele (supplementary fig. S6, Supplementary Material online).

No one vaccine haplotype is responsible for all rash formation. Instead, the vOka haplotype is different in every lesion, a finding that has previously been inferred from limited SNP typing (Quinlivan et al. 2007). Rash-forming haplotypes are, however, not completely representative of the virus population in the vaccine. Complete viral genome sequences identify nine nonsynonymous loci in five proteins, at which the ancestral allele is most significantly favored in the rashes. Although two of these loci, positions 107797 (L446P) and 105356 (I1260V) in ORF 62 were already known (Quinlivan et al. 2007) three loci, 106001 (K1045E) in ORF 62 domain IV, 19036 (E199R) in ORF 13, and 97479 (V495A) in ORF 55 had not been previously identified by Sanger sequencing as polymorphic in the vaccine (fig. 4). Selection of the wild-type allele at 11 of 12 loci, suggests that vOka rash-causing viruses may be less attenuated for replication in skin, particularly as we found no in silico evidence that the observed amino acid changes are likely to increase escape from host immunity. The location of four selected positions in ORF 62, the major viral transactivating protein expressed early in VZV replication, supports the notion that mutations in this protein are critical for vaccine Oka attenuation in skin, although the possibility exists that some of these sites are in linkage disequilibrium (we were not able to determine this using the current methodology) while previous studies have shown these SNPs alone do not account for attenuation (Zerboni et al. 2005; Cohrs et al. 2006). Position 107797 (L446P in ORF62/71), a locus that has been consistently observed as the wild-type allele among vaccine rash isolates, appears to be essential for vOka rash formation; its position within region 2 of IE62 could affect the binding efficiency of the transcription complex that includes IE62, IE63 and host transcription factors (Lynch et al. 2002; Quinlivan et al. 2011). IE62 and IE63 are critical for viral

replication in skin and selection of the wild-type allele at position 101937 in the ORF63 promoter may therefore also contribute to optimal VZV replication in skin (fig. 4). Position *130R in ORF0p and I1260L in domain IV of ORF62 occur in other highly passaged VZV isolates and have been implicated as contributors to vaccine attenuation. The arginine at position *130R in ORF 0, a homolog of HSV UL56, is also present in VZV Ellen, a strain that is highly attenuated for replication in skin (Peters et al. 2012). The leucine at position I1260L in ORF62 has also been shown to evolve during passage of a wild-type strain, VZV strain 32, in repeated tissue culture monolayers (Tyler et al. 2007). The selection of the wild-type isoleucine in most vaccine rashes at this position may therefore be important for VZV replication in epidermal tissue. Although the functional significance of position 58595 (I593V) is not known, the SNP is located in the ectodomain of gB, which is essential for replication in skin (Zhang et al. 2010). gB is within a region of the vOka genome that has been identified by cosmid analysis as contributory to its attenuation for replication in epithelial tissue (Moffat et al. 1998). Positions 97748 (A585T) and 97479 (V495A) in ORF 55p, a homolog of the Herpes Simplex UL5 helicase protein, lie between predicted helicase domains III and IV, which are conserved in VZV (Zhu and Weller 1992). ORF 55p functions together with ORFs 6p and 52p to form the VZV helicase primase complex, which is essential for viral replication. Position 19036 (G199R) is located within the ORF 13 thymidylate kinase gene and also plays a part in viral replication. The vaccine allele at these loci may therefore contribute to attenuation of vOka strain replication, although this remains to be confirmed.

Whole-genome sequencing of vaccine-associated rashes arising soon after vaccination and comparison with those that followed reactivation from latency provided a unique opportunity to study the pathogenesis of VZV in its natural host. We found no difference in vOka haplotypes between rashes which occur soon after vaccination and those that occurred following a period of latency and reactivation from dorsal route ganglia (fig. 5). Thus, from our data, vaccine variants that persist following inoculation appear equally able to establish latency and reactivate. Moreover, there is little evidence for within host selection of neurotropic variants. We therefore conclude that neither preexisting nor new viral variation is likely to play a part in the establishment of vOka latency or reactivation (fig. 5).

Finally, it has been suggested previously that vOka vaccine, because of its heterogeneity, might evolve at a population level; similarly to RNA viral quasispecies (Gomi et al. 2002; Peters et al. 2012). For this to happen, the vOka mutation rate would have to be sufficient to maintain heterogeneity. Although our findings do indicate the accumulation of mutations as vOka replicates in the skin, mutation rates appear to be insufficient to recapitulate the diversity that was originally present in the vaccine. There appears to be even less opportunity for mutation before the virus reaches the skin (Moffat et al. 1998; Zerboni et al. 2005). We have found evidence of selection in these data, but it is not selection for new mutation-generated variants as in a classical quasispecies model. Instead, it is selection acting on the preexisting variation that was segregating in the vaccine. It appears that some vaccine alleles, which had drifted up to intermediate frequencies during the process of vaccine production, are then actively selected against on the route from inoculation to rash formation in the human body. These loci are therefore strong candidates implicated in the genetics of rash formation.

In summary, we used targeted enrichment technologies to recover sufficient VZV DNA from rashes caused by the vOka vaccine for whole viral genome sequencing. The method allowed variation in allele frequencies to be accurately quantified for assessment of genomic populations. We used this approach to investigate vOka rash formation and the results also shed light on the pathogenesis of VZV and the live attenuated vaccine strain vOka. We identify at least nine nonsynonymous residues in five ORFs across the VZV genome that are probably important for replication of VOka in the major target organ, skin. From our data, we find no evidence for population bottlenecks or viral selection related to latency or reactivation, and no evidence that vOka evolves as a quasispecies. Taken together, we predict that while vOka variation plays some role in the development of varicella-like rashes after inoculation, it plays little or no part in the pathogenesis of herpes zoster and related reactivation illnesses. The failure to identify neurotropic vOka strains provides important data for current efforts to develop vaccines that do not establish latency.

Materials and Methods

Repository of Data

Consensus sequences for all vaccines and vaccine rash samples sequenced in this study have been uploaded to the NR database on Genbank under accession numbers KF558371-91.

Sample Collection and Ethics

Diagnostic samples from patients with confirmed vaccine-associated rashes (both varicella- [$n = 8$] and zoster-like [$n = 12$]) collected in the UK and United States between 1988 and 2010, were retrieved from the Breuer lab cryobank. The clinical specimens, collected as part of standard clinical practice to genotype vaccine rashes occurring in recipients of vOka vaccine, were independently obtained from patients who developed skin rashes following immunization with the Merck vOka vaccine. All rashes were confirmed as vOka strain using previously published methods (Larussa et al. 1998; Parker et al. 2006). Rashes occurring within 42 days of vaccination were classified as varicella like. The exception was one case, T61, which was reported as a varicella-like rash 90 days after vaccination. Although by our definition this was thought to be more likely to be disseminated zoster, the case was included as varicella like. Cases presenting with the classical dermatomal rash were diagnosed clinically as herpes zoster. Samples were stripped of all personal identifiers other than details of vaccination and the type of rash, no patient information was available. The use of these specimens for research was approved by the East London and City Health Authority

Research Ethics Committee (P/96/046: Molecular typing of cases of VZV).

In addition, three batches of the VariVax live-attenuated VZV Vaccine preparation were collected in 2008 (VVAG [USA]), 2010 (VV10 [UK]), and 2012 (VV12 [UK]) and batch and lot numbers used to verify them as independent preparations (table 1).

Cell Culture

Cultured isolates (table 1) were passaged in melanoma cells (MeWo) that were propagated in culture media (MEM [Sigma] supplemented with 10% FBS and 1% nonessential amino acids). Viruses were harvested for DNA extraction 3–4 days after inoculation when syncytia was visible.

Library Construction, Targeted Enrichment, and Sequencing

Total DNA was extracted from each sample using the QiaAMP DNA mini kit (QIAGEN) according to manufacturer's instructions. DNA quantification was performed using a NanoDrop spectrophotometer and those with 260/280 ratios outside the range 1.7–2.1 and 260/230 ratios out the range 1.8–2.2 were further purified using the Zymoclean Genomic DNA Clean & Concentrator (Zymo Research Corp.). WGA using GenomiPhi V2 (GE Healthcare) was performed using 10 ng of starting material where <50 ng total DNA was available (table 1). Libraries were constructed as per the standard SS XT v1.3 (Agilent) and NEBNext (New England Biolabs) protocols, the latter modified to include a second PCR amplification step (6 cycles) to enable multiplexing using standard Illumina barcodes. Enrichment for VZV sequences was performed as described previously (Depledge et al. 2011). Samples were sequenced across several Illumina platforms (GAIIx, HiSeq, and MiSeq) according to availability. Specific sample preparation and sequencing metrics are shown in table 1. Concerns about bias introduced by SS and WGA methodologies were assuaged by resequencing the vaccine VV12 under multiple conditions (supplementary table S2, Supplementary Material online) as in all cases, the R^2 correlation between allele frequencies exceeded 0.93, and no identifiable bias was observed regardless of sample treatment.

Genome Assembly and Variant Calling

Each data set was parsed through QUASR (Watson et al. 2013) for duplicate removal and read-trimming (-q 30, -l 50) and subsequently aligned against pOka (AB097933.1) using BWA (Li and Durbin 2009). Resulting alignments were processed using SAMTools (Li et al. 2009) to generate pileup files for each sample. A consensus sequence for each data set was called with the QUASR module pileupConsensus and a 50% frequency threshold (i.e., no ambiguities were included). Variant profiling for each data set was performed using VarScan v2.2.11 (Koboldt et al. 2012) with the following parameters: basecall quality ≥20, read depth ≥50, independent reads supporting minor allele ≥2 per strand. In addition, variant calls showing directional strand bias ≥0.85 were excluded from further analyses. Consensus sequences were generated for each rash sample but iterative repeat regions (ORIS R1, R2, R3, R4, and R5) and the terminal repeat region were trimmed before analysis.

Consensus Sequence Analyses

DNA sequences were aligned using the program Mafft, v6 (Katoh 2002) with alignments checked manually; no insertions or deletions were inferred from the alignment. Prior to further analysis the VZV repeat regions R1, R2, R3, R4, R5, and OriS (Davison and Scott 1986; Hondo and Yogo 1988) were removed as these could not be accurately mapped given the limited DNA available. Bayesian phylogenetic trees were inferred using the program Beast v1.7.2 (Drummond et al. 2012). Two independent Monte Carlo-Markov chains were run for 50,000,000 iterations for the whole genome (minus repeat regions) with a thinning of 1,000. The most appropriate model of nucleotide substitution was selected using the program jModelTest (Posada 2008), which identified a general-time-reversible model of nucleotide substitution with a gamma-shaped rate distribution and a proportion of invariable sites (GTR + γ + I) as being the best site model. The phylogeny was reconstructed under a number of different tree priors including the Bayesian skyline, constant, and exponential growth models. Runs were checked for convergence and that an effective sample size of at least 200 had been achieved for all parameters. Runs were combined using LogCombiner v1.6.1, and TreeAnnotator v1.6.1 was used to obtain the highest clade credibility tree and posterior probabilities per node. Maximum likelihood phylogenies were also inferred using Mega v5.05 (Tamura et al. 2011) under the same model of evolution. All trees were visualized and drawn using Figtree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/, last accessed September 17, 2013). We tested for evidence of recombination using the programs GARD and SBP (Pond et al. 2006), and the BootScan method contained in the RDP3 program (Martin et al. 2010). The program BaTS (Parker et al. 2008) was used to test for an association between the Bayes-inferred phylogeny and the vaccine rash phenotype (varicella-like or zoster-like). BaTS estimates the association using the association index (Wang et al. 2001), parsimony score (Slatkin and Maddison 1989), and maximum monophyletic clade statistics.

Statistical Analyses

To compensate for intrinsic differences in the sequencing technologies used, their respective error profiles and fold differences between the mean read depths of individual samples, we identified a minimal criterion for defining biallelic sites that segregate our samples 1) total read depth ≥ 50, 2) the vaccine allele must appear ≥2 independent reads mapping to each strand, 3) the vaccine frequency must be ≥1%, and 4) the segregating site must appear in ≥2 vaccines and/or vaccine rash samples at the earlier-mentioned criteria. These criteria were applied only to data obtained using the MiSeq and HiSeq platforms and yielded a total of 368 sites that were segregating across the vaccines and vaccine rashes. The allele frequencies at these sites in the GAIIx-derived data were

subsequently included if the vaccine allele was a match (i.e., not a sequencing error). The frequency of the vaccine allele in vaccine rash samples at the majority of these sites was effectively binary, either 0% or 100%. 224 sites are polymorphic in two or more vaccine batches while sites that were polymorphic in just the one vaccine were excluded from further analysis. Subsequently, we calculated the mean vaccine allele frequency for each of these polymorphic sites in the three sequenced vaccines, VV10, VV12, and VVAG. We investigated any consistent changes in allele frequency from the vaccine to the vaccine rashes. The variance in vaccine rash allele frequencies appeared higher than expected and we modeled this over dispersion around the expectation (the rash frequency) using two approaches: a generalized linear mixed model and a generalized linear model with quasi-binomial error. With the rash allele frequencies as the response variable, we compared four models in the generalized linear model: no intercept, intercept only, coding/noncoding sites, and synonymous/nonsynonymous/intergenic region.

The majority of the 224 vaccine polymorphic sites in the vaccine rashes were effectively binary, that is, fixed for either the vaccine or the wild-type allele. Consequently, to identify any specific sites where the change in allele frequency cannot be explained by genetic drift alone, we compared the proportion of vaccine rashes with the vaccine allele to the expected proportion (vaccine mean). We estimated the effect size (based upon the mean vaccine allele frequency in the vaccines and the proportion of vaccine rashes fixed for the vaccine allele) and power (Cohen 1988), and identified those sites for which there was $\geq$80% power to compare the vaccine rashes with the vaccine. Finally, we applied two-tailed binomial exact tests to those sites with $\geq$80% power with the probability of "success" equal to the mean vaccine allele frequency in the vaccines. We corrected for multiple analyses using FDR (Hochberg and Benjamini 1990).

We also used the power analyses to identify a set of the vaccine polymorphic sites that could be analyzed with SNP typing of an additional 36 vaccine rashes, which were readily available and/or the inclusion of previously published data (Loparev et al. 2007; Quinlivan et al. 2007; Thiele et al. 2011). Using binomial exact tests, we compared the proportion of the vaccine allele at these sites in all of the vaccine rashes (sequenced, SNP typed, and previously published) with the mean frequency in the sequenced vaccines (with a FDR correction for multiple analyses). In addition, we reanalyzed those sites that we showed to have sufficient power from the sequencing (discussed earlier) with the addition of SNP typed and previously published data. Finally, we used generalized linear modeling with binomial error to identify any differences between varicella and zoster vaccine rashes at the 224 vaccine polymorphic sites. All binomial tests and the generalized linear modeling were performed in R (www.r-project.org/index.html, last accessed September 17, 2013). Effect size and power calculations were done using the "pwr" package in R, and the modeling of rash allele frequencies used the lme4 and VGAM packages.

To infer the size of the putative population bottleneck occurring in the vaccine rash after vaccination, we fitted a beta-binomial distribution around the expectation of the vaccine (estimated from the mean allele frequencies of the three vaccines). It has been shown that the distribution of allele frequencies (assuming that loci are biallelic) under genetic drift can be modeled as a beta-binomial distribution (Balding 2003). The probability that two alleles randomly drawn are identical by descent, F, is directly related to the correlation parameter of the beta distribution. We fitted the beta-binomial using the R package VGAM. Finally, we modeled the rash sample as the product of a bottleneck followed by exponential growth. We found the best fitting growth rate corresponding to the observed F, using the MLE function of R. The expected value of F was calculated in a model of postbottleneck exponential growth to a nominal end point of million-fold increase. F was calculated as $1 - \Pi(1\text{-}1/N_i)$, where N_i is the population size in generation i. We then used a Monte Carlo simulation of ancestry with these population sizes, by artificially drawing the ancestors of a sample back through the 201 preceding generations. The sample size matched our average coverage (1,942 reads). From 1,000 such possible ancestries, we generated the allele frequency spectra that would occur for mutations occurring post bottleneck and prebottleneck. The mutations were uniformly distributed across all ancestors (in each category). We matched the allele frequency spectrum of alleles that were not detectable in the vaccines, presumably new mutations occurring postbottleneck, and compared F_{IS} with the output from a stochastic model of ancestry with exponential growth (implemented in R).

Additional SNP typing

SNP typing at positions 59591, 90318, 94167, 97479, 97748, and 106001 was conducted on an additional 36 vaccine rash samples derived from a range of patients across the UK and the United States. PCRs comprising 30 amplification cycles were performed in 12.5 µl reactions using either the Extensor 2x Master Mix (Thermo Fisher Scientific, Hanover Park, IL) or the Takara Ex Taq Kit (Clontech Laboratories, Mountain View, CA). Each reaction of 2 µl was run on 1% agarose gel and visualized on a Biorad Molecular Imager Gel Doc XR + Imagine System (Bio-Rad Laboratories, CA). PCR products were cleaned up using ExoSap-It from USB (Affymetric, Cleveland, OH) according to manufacturer's specifications and sequenced on the 3130X/Genetic Analyzer from ABI (Life technologies, Foster City, CA). The sequence traces were analyzed with the Sequencher program from Gene Codes Corporation (Ann Arbor, MI). All primers used for SNP typing are shown in supplementary table S6 (Supplementary Material online).

Selection of Low-Level De Novo Variants

Mutations acquired postinoculation (i.e., de novo) were identified in data sets derived from the MiSeq under the following criteria; the variant allele occurred in $\geq$2 reads on opposite strands with a read depth $\geq$ 50 and a frequency exceeding 1%. To identify whether there was any selection operating upon the new mutations in the vaccine rashes, we used the Kolmogorov–Smirnov test to compare the total number of

nonsynonymous mutations with the sum of the synonymous and noncoding mutations in all of the rashes.

In Silico Epitope Prediction

As fitness selection is often linked to immune escape, the nonsynonymous sites under selection 59591, 97479, 97748, 105356, 106001, 107599, and 107797 were examined using Metaserver (MacNamara and Kadolsky 2009) and Epipred (Heckerman et al. 2006) to identify HLA class 1 epitopes mapping to the vaccine and wild-type peptides and to calculate the predicted binding scores. No significant hits were reported for any site.

Supplementary Material

Supplementary figures S1–S7 and tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Darren Marjenberg, Bram Herman, and Emily Leproust (Agilent Technologies) for technical advice relating to the SS Target Enrichment methodology and Becca Asquith (Imperial College) for discussion about epitope prediction. They thank Tony Brooks (UCL Genomics) and Charles Mein (QMUL Genomics) for Illumina sequencing. The authors acknowledge the infrastructure support provided the MRC Centre for Molecular Medical Virology grant G0900950, the NIHR UCL/UCLH Biomedical Research Centre, the EU FP7 PathSeek grant and the use of the UCL Legion High Performance Computing Facility, and associated support services, in the completion of this work. The work was supported by the MRC grant G0700814; the MRC Centre grant G0900950 to D.P.D.; the NIHR UCL/UCLH Biomedical Research centre to S.K. and J.B.; MRC grant G0700814 to M.J.; National Institutes of Health grants NS064022 and EY08098 to P.R.K.; and Research to Prevent Blindness Inc. and The Eye & Ear Institute of Pittsburgh. P.R.K. acknowledges support from Michael B. Yee.

References

Arvin AM. 2001. Varicella-zoster virus: molecular virology and virus-host interactions. *Curr Opin Microbiol.* 4:442–449.

Balding DJ. 2003. Likelihood-based inference for genetic correlation coefficients. *Theor Popul Biol.* 63:221–230.

Banovic T, Yanilla M, Simmons R, Robertson I, Schroder WA, Raffelt NC, Wilson YA, Hill GR, Hogan P, Nourse CB. 2011. Disseminated varicella infection caused by varicella vaccine strain in a child with low invariant natural killer T cells and diminished CD1d expression. *J Infect Dis.* 204:1893–1901.

Barrett-Muir W, Scott FT, Aaby P, Scott FT, Aaby P, John J, Matondo P, Chaudhry QL, Siqueira M, Poulsen A, Yaminishi K, Breuer J. 2003. Genetic variation of varicella-zoster virus: evidence for geographical separation of strains. *J Med Virol.* 70(1 Suppl):S42–S47.

Breuer J, Schmid DS. 2008. Vaccine Oka variants and sequence variability in vaccine-related skin lesions. *J Infect Dis.* 197(Suppl):S54–S57.

Chaves SS, Haber P, Walton K, Wise RP, Izurieta HS, Schmid DS, Seward JF. 2008. Safety of varicella vaccine after licensure in the United States: experience from reports to the vaccine adverse event reporting system, 1995-2005. *J Infect Dis.* 197(2 Suppl):S170–S177.

Chen JJ, Gershon A a, Li ZS, Lungu O, Gershon MD. 2003. Latent and lytic infection of isolated guinea pig enteric ganglia by varicella zoster virus. *J Med Virol.* 70(1 Suppl):S71–S78.

Civen R, Chaves SS, Jumaan A, Wu H, Mascola L, Gargiullo P, Seward JF. 2009. The incidence and clinical characteristics of herpes zoster among children and adolescents after implementation of varicella vaccination. *Pediatr Infect Dis J.* 28:954–959.

Cohen J. 1988. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates.

Cohen JI. 2007. Varicella-zoster vaccine virus: evolution in action. *Proc Natl Acad Sci U S A.* 104:7–8.

Cohrs R, Gilden D, Gomi Y. 2006. Comparison of virus transcription during lytic infection of the Oka parental and vaccine strains of varicella-zoster virus. *J Virol.* 80:2076–2082.

Davison AJ, Scott JE. 1986. The complete DNA sequence of varicella-zoster virus. *J Gen Virol.* 67:1759–1816.

Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda RK, Leproust E, Kellam P, Breuer J. 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One* 6: e27805.

Drummond A, Suchard M, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29: 1969–1973.

Gomi Y, Sunamachi H, Mori Y, Nagaike K, Takahashi M, Yamanishi K. 2002. Comparison of the complete DNA sequences of the Oka varicella vaccine and its parental virus. *Society* 76:11447–11459.

Goulleret N, Mauvisseau E, Essevaz-Roulet M, Quinlivan M, Breuer J. 2010. Safety profile of live varicella virus vaccine (Oka/Merck): five-year results of the European varicella zoster virus identification program (EU VZVIP). *Vaccine* 28:5878–5882.

Heckerman D, Kadie C, Listgarten J. 2006. Leveraging information across HLA alleles/supertypes improves epitope prediction. *J Comput Biol.* 14:736–746.

Hochberg Y, Benjamini Y. 1990. More powerful procedures for multiple significance testing. *Stat Med.* 9:811–818.

Hondo R, Yogo Y. 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.

Kanda RK, Quinlivan ML, Gershon AA, Nichols RA, Breuer J. 2011. Population diversity in batches of the varicella Oka vaccine. *Vaccine* 29:3293–3298.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30: 3059–3066.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22:568–576.

Kramer JM, LaRussa P, Tsai WC, Carney P, Leber SM, Gahagan S, Steinberg S, Blackwood RA. 2001. Disseminated vaccine strain varicella as the acquired immunodeficiency syndrome-defining illness in a previously undiagnosed child. *Pediatrics* 108:e39–e39.

Larussa P, Steinberg S, Arvin A, Dwyer D, Burgess M, Menegus M, Rekrut K, Yamanishi K, Gershon A. 1998. Polymerase chain reaction and restriction fragment length polymorphism analysis of varicella-zoster virus isolates from the United States and other parts of the world. *J Infect Dis.* 178:64–66.

Lee S, Markham P, Coppo M. 2012. Attenuated vaccines can recombine to form virulent field viruses. *Science* 337:2012.

Levin MJ, DeBiasi RL, Bostik V, Schmid DS. 2008. Herpes zoster with skin lesions and meningitis caused by 2 different genotypes of the Oka varicella-zoster virus vaccine. *J Infect Dis.* 198: 1444–1447.

Levy O, Orange JS, Hibberd P, et al. (12 co-authors). 2003. Disseminated varicella infection due to the vaccine strain of varicella-zoster virus, in a patient with a novel deficiency in natural killer T cells. *J Infect Dis.* 188:948–953.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Loman NJ, Misra R V, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol.* 30:434–439.

Loparev VN, Rubtcova E, Seward JF, Levin MJ, Schmid DS. 2007. DNA sequence variability in isolates recovered from patients with post-vaccination rash or herpes zoster caused by Oka varicella vaccine. *J Infect Dis.* 195:502–510.

Lynch JM, Kenyon TK, Grose C, Hay J, Ruyechan WT. 2002. Physical and functional interaction between the varicella zoster virus IE63 and IE62 proteins. *Virology* 302:71–82.

MacNamara A, Kadolsky U. 2009. T-cell epitope prediction: rescaling can mask biological variation between MHC molecules. *PLoS Comput Biol.* 5:e1000327.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26(19):2462–2463.

Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 12:R112.

Moffat J, Zerboni L, Kinchington P. 1998. Attenuation of the vaccine Oka strain of varicella-zoster virus and role of glycoprotein C in alphaherpesvirus virulence demonstrated in the SCID-hu mouse. *J Virol.* 72:965–974.

Parker J, Rambaut A, Pybus OG. 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol.* 8:239–246.

Parker SP, Quinlivan M, Taha Y, Breuer J. 2006. Genotyping of varicella-zoster virus and the discrimination of Oka vaccine strains by TaqMan real-time PCR. *J Clin Microbiol.* 44:3911–3914.

Peters GA, Tyler SD, Carpenter JE, Jackson W, Mori Y, Arvin AM, Grose C. 2012. The attenuated genotype of varicella-zoster virus includes an ORF0 transitional stop codon mutation. *J Virol.* 86:10695–10703.

Peters GA, Tyler SD, Grose C, Severini A, Gray MJ, Upton C, Tipples GA. 2006. A full-genome phylogenetic analysis of varicella-zoster virus reveals a novel origin of replication-based genotyping scheme and evidence of recombination between major circulating clades. *J Virol.* 80:9850–9860.

Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22(24):3096–3098.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341.

Quinlivan M, Breuer J, Schmid DS. 2011. Molecular studies of the Oka varicella vaccine. *Expert Rev Vaccines.* 10:1321–1336.

Quinlivan MA, Gershon AA, Nichols RA, La Russa P, Steinberg SP, Breuer J. 2006. Vaccine Oka varicella-zoster virus genotypes are monomorphic in single vesicles and polymorphic in respiratory tract secretions. *J Infect Dis.* 193:927–930.

Quinlivan ML, Gershon AA, Al Bassam MM, Steinberg SP, LaRussa P, Nichols RA, Breuer J. 2007. Natural selection for rash-forming genotypes of the varicella-zoster vaccine virus detected within immunized human hosts. *Proc Natl Acad Sci U S A.* 104:208–212.

Quinlivan ML, Gershon AA, Steinberg SP, Breuer J. 2004. Rashes occurring after immunization with a mixture of viruses in the Oka vaccine are derived from single clones of virus. *J Infect Dis.* 190:793–796.

Rentier B, Gershon AA. 2004. Consensus: varicella vaccination of healthy children. *Pediatr Infect Dis J.* 23:379–389.

Sengupta N, Taha Y, Scott FT, Leedham-Green ME, Quinlivan M, Breuer J. 2007. Varicella-zoster-virus genotypes in East London: a prospective study in patients with herpes zoster. *J Infect Dis.* 196:1014–1020.

Sharrar RG, LaRussa P, Galea SA, Steinberg SP, Sweet AR, Keatley RM, Wells ME, Stephenson WP, Gershon AA. 2000. The postmarketing safety profile of varicella vaccine. *Vaccine* 19:916–923.

Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow measured from the phylogenies of alleles. *Genetics* 123:603–613.

Takahashi M, Okuno Y, Otsuka T, Osame J, Takamizawa A. 1975. Development of a live attenuated varicella vaccine. *Biken J.* 18:25–33.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Thiele S, Borschewski A, Küchler J, Bieberbach M, Voigt S, Ehlers B. 2011. Molecular analysis of varicella vaccines and varicella-zoster virus from vaccine-related skin lesions. *Clin Vaccine Immunol.* 18: 1058–1066.

Tillieux SL, Halsey WS, Thomas ES, Voycik JJ, Sathe GM, Vassilev V. 2008. Complete DNA sequences of two oka strain varicella-zoster virus genomes. *J Virol.* 82:11023–11044.

Tsolia M, Gershon AA, Steinberg SP, Gelb L. 1990. Live attenuated varicella vaccine: evidence that the virus is attenuated and the importance of skin lesions in transmission of varicella-zoster virus. National Institute of Allergy and Infectious Diseases Varicella Vaccine Collaborative Study Group. *J Pediatr.* 116:184–189.

Tyler SD, Peters GA, Grose C, Severini A, Gray MJ, Upton C, Tipples GA. 2007. Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. *Virology* 359:447–458.

Wang TH, Donaldson YK, Brettle RP, Bell JE, Simmonds P. 2001. Identification of shared populations of Human immunodeficiency Virus Type 1 infecting microglia and tissue macrophages outside the central nervous system. *J Virol.* 75:11686–11699.

Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P. 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci.* 368:20120205.

Zerboni L, Hinchliffe S, Sommer MH, et al. (11 co-authors). 2005. Analysis of varicella zoster virus attenuation by evaluation of chimeric parent Oka/vaccine Oka recombinant viruses in skin xenografts in the SCIDhu mouse model. *Virology* 332:337–346.

Zhang Z, Selariu A, Warden C, Huang G, Huang Y, Zaccheus O, Cheng T, Xia N, Zhu H. 2010. Genome-wide mutagenesis reveals that ORF7 is a novel VZV skin-tropic factor. *PLoS Pathog.* 6:e1000971.

Zhu LA, Weller SK. 1992. The six conserved helicase motifs of the UL5 gene product, a component of the herpes simplex virus type 1 helicase-primase, are essential for its function. *J Virol.* 66: 469–479.