Recombination of Globally Circulating Varicella-Zoster Virus

Peter Norberg, a Daniel P. Depledge, b Samit Kundu, b Claire Atkinson, c Julianne Brown, d Tanzania Haque, e Yusuf Hussaini, e Ethne MacMahon, f Pamela Molyneaux, g Vasiliki Papaevangelou, h Nitu Sengupta, i Evelyn S. C. Koay, j Julian W. Tang, k Gillian S. Underhill, l Anna Grahn, m Marie Studahl, n Judith Breuer, o Tomas Bergström

Department of Infectious Diseases, Section for Clinical Virology, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden; a; MRC Centre for Medical Molecular Virology, Division of Infection and Immunity, London, United Kingdom; b; Department of Virology, Royal Free Hospital, London, United Kingdom; Microbiology, Virology and Infection Prevention and Control, Camelia Botnar Laboratories, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; Department of Microbiology, Ahmadu Bello University, Zaria, Kaduna State, Nigeria; c; Department of Infectious Diseases, St. Thomas’ Hospital, Guy’s & St. Thomas’ NHS Foundation Trust, London, United Kingdom; Department of Medical Microbiology, Aberdeen Royal Infirmery, Aberdeen, United Kingdom; ; Third Department of Pediatrics, University of Athens School of Medicine, University General Hospital ATTIKON, Athens, Greece; e; Ealing Services for Children with Additional Needs, London, United Kingdom; Molecular Diagnosis Centre, Department of Laboratory Medicine, National University Hospital, Singapore; Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom; ; Department of Clinical Microbiology, Queen Alexandra Hospital, Portsmouth, United Kingdom

ABSTRACT
Varicella-zoster virus (VZV) is a human herpesvirus, which during primary infection typically causes varicella (chicken pox) and establishes lifelong latency in sensory and autonomic ganglia. Later in life, the virus may reactivate to cause herpes zoster (HZ; also known as shingles). To prevent these diseases, a live-attenuated heterogeneous vaccine preparation, vOka, is used routinely in many countries worldwide. Recent studies of another alphaherpesvirus, infectious laryngotracheitis virus, demonstrate that live-attenuated vaccine strains can recombine in vivo, creating virulent progeny. These findings raised concerns about using attenuated herpesvirus vaccines under conditions that favor recombination. To investigate whether VZV may undergo recombination, which is a prerequisite for VZV vaccination to create such conditions, we here analyzed 115 complete VZV genomes. Our results demonstrate that recombination occurs frequently for VZV. It thus seems that VZV is fully capable of recombination if given the opportunity, which may have important implications for continued VZV vaccination. Although no interclade vaccine-wild-type recombinant strains were found, intraclade recombinants were frequently detected in clade 2, which harbors the vaccine strains, suggesting that the vaccine strains have already been involved in recombination events, either in vivo or in vitro during passages in cell culture. Finally, previous partial and complete genomic studies have described strains that do not cluster phylogenetically to any of the five established clades. The additional VZV strains sequenced here, in combination with those previously published, have enabled us to formally define a novel sixth VZV clade.

IMPORTANCE
Although genetic recombination has been demonstrated to frequently occur for other human alphaherpesviruses, herpes simplex viruses 1 and 2, only a few ancient and isolated recent recombination events have hitherto been demonstrated for VZV. In the present study, we demonstrate that VZV also frequently undergoes genetic recombination, including strains belonging to the clade containing the vOka strain.

Varicella-zoster virus (VZV) is a human herpesvirus belonging to the genus Varicellovirus of the Alphaherpesvirinae subfamily of the order Herpesvirales. VZV has a stable prevalence in all countries worldwide, and the primary infection is almost always symptomatic. Upon initial infection, the virus typically causes chicken pox, a highly contagious disease with tropism to skin and ganglia. In temperate climates, chicken pox is most commonly observed in children and is one of the classic diseases of childhood. In contrast, in tropical regions chicken pox often occurs in adolescents or adults, typically causing more-severe symptoms (1). Primary infection is followed by the establishment of lifelong latency of the virus in dorsal root and trigeminal ganglia. The virus may then reactivate at a later stage from one or more ganglia to cause herpes zoster, affecting the corresponding dermatomes (2). The virus may also cause various neurological complications, such as facial palsy, encephalitis, meningitis, and cerebral vasculitis (3–8).

The disease burden of primary and reactivated VZV motivated the development and use of a vaccine to prevent its spread and disease. A live attenuated vaccine strain, vOka (Varixav or Varilrix) was originally developed in Japan to prevent the spread of varicella among children (9). This vaccine, a heterogeneous mixture of related haplotypes, is routinely used in many countries, including Japan, South Korea, the United States, Canada, Australia, Germany, Costa Rica, Uruguay, and Qatar. More recently, the vaccine trademarked as Zostavax was licensed for use in the prevention of herpes zoster and postherpetic neuralgia in elderly peo-
perevirus strains was reported for pseudorabies virus, which is an alphaherpesvirus that infects pigs (44). The authors coinfected swine with two different attenuated vaccine strains, and virus isolates derived by serial plaque purification directly from tissue homogenates were characterized as recombinant and parental strains. Another recent study of infectious laryngotracheitis virus (ILTV), an alphaherpesvirus that infects poultry, analyzed distinct viruses (referred to as genotype classes 8 and 9) that were isolated from geographically distinct areas of Australia in 2008 (45). These strains are highly virulent and have been associated with outbreaks causing mortality rates of up to 17.6% (46). Three ILTV vaccines (live attenuated strains) are available in Australia: two closely related vaccines of Australian origin, SA2 and A20 (Pfizer, Australia), and one vaccine of European origin, Serva (Intervet). The vaccines of Australian origin are closely related and classified as class 1 (47). The vaccine of European origin, however, is classified as class 7 (48), and its genome is divergent from the vaccine strains of Australian origin, having only 99.2% nucleotide sequence identity (47). Whole-genome analysis of the virulent viruses of classes 8 and 9 demonstrated that these are recombinant viruses derived from the vaccine strains of Australian and European origin (45). It is thus possible for some live attenuated herpesvirus vaccine strains to recombine in vivo to form virulent progeny that spreads through the population. The authors also studied the pathogenicity of these recombinant viruses in specific-pathogen-free chickens and concluded that each recombinant strain had distinct in vivo genotypes with significantly increased virulence or replication compared with their parental strains. These findings gave rise to concerns about the use of attenuated herpesvirus vaccines under conditions that favor recombination (45).

Although recombination has been evidenced in VZV, little is known about how frequently recombination events occur within the host population. To evaluate the likelihood of the emergence of novel interclade recombinants as a result of increased migration and/or vaccination with attenuated viruses, we need to increase our understanding of VZV recombination in general and to determine how frequently this occurs in nature. The aim of the present study was thus to perform an evolutionary analysis of VZV and thoroughly map the extent of homologous recombination between circulating VZV strains. The purpose was to evaluate whether the use of attenuated VZV vaccines might present conditions that favor recombination. We analyzed 115 complete VZV genomes, derived from GenBank or newly sequenced here, from wild-type and vaccine strains using multiple advanced algorithms in parallel for comparison. Phylogenetic analysis revealed the presence of a novel clade of VZV, which was designated clade 6. We also present data that support frequent homologous recombination of VZV and provide strong statistical support for the existence of multiple inter- and intraclade recombination events. These results may have implications for our understanding of the future evolution of VZV and the use of live VZV vaccines.

MATERIALS AND METHODS

Ethics statement. Clinical specimens (diagnostic samples collected as part of standard clinical procedures) were independently obtained from patients with confirmed VZV infection and anonymized prior to this study. Consent for all children and adult participants was written and informed. The parent or guardian of any child participant provided informed consent on their behalf. The use of these specimens for research was approved.

Norberg et al.
DNA was extracted from each sample using the QiaAMP DNA minikit (Qiagen) according to the manufacturer’s instructions. Whole-genome amplification using the Genomic DNA Clean & Concentrator kit (Zymo Research) according to the manufacturer’s instructions. DNA quantification was performed using a NanoDrop spectrophotometer, and samples with 260/280 ratios outside the range of 1.7 to 2.1 and/or 260/230 ratios outside the range of 1.8 to 2.2 were further purified using the Genomic DNA Clean & Concentrator kit (Zymo Research) according to the manufacturer’s instructions. Whole-genome amplification using GenomiPhi V2 (GE Healthcare) was performed using 10 ng of DNA and yielded between 3,000 and 5,000 ng of DNA. Sequencing libraries were constructed using 3,000 ng of DNA as input, as per the standard SureSelect XT v1.5 protocols (Agilent). Enrichment for VZV DNA was performed as described previously (49, 50). The final sequence libraries were multiplexed and sequenced using either 2 x 150 bp or 2 x 250 bp paired-end kits across multiple runs on an Illumina MiSeq.

Genome assembly and variant calling. Sequence data sets were demultiplexed using BaseSpace, and individual data sets were subsequently parsed through Quasr (51) for duplicate removal and read-trimming (≤q 30, -1 50) and subsequently aligned against the VZV reference strain Dumas (NC_001348) using BWA (52). The resulting alignments were processed using SAMTools (17) 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 (53) with the following parameters: base call quality, ≥20; read depth, ≥50; independent reads supporting minor allele, ≥2 per strand. In addition, variant calling showing a directional strand bias of ≥0.85 were excluded from further analyses. Consensus sequences were generated for each rash sample, but iterative repeat regions R1, R2, R3, R4, and R5 (19, 20) as well as the terminal repeat region were trimmed prior to tree-building analyses.

DNA sequence analysis. DNA sequences were aligned using Kalign (54) implemented in eBioX (55). Phylogenetic networks were constructed using SplitsTree4 (56) with Uncorrected_P characters transformation, and all gaps and repeat regions were excluded prior to analysis. The informative sites analysis was performed using SimPlot (57), and the recombination analysis was performed by using the algorithms RDP, Geneconv, Chimaera, MaxChi, BootScan, SisScan, 3Seq, and LARD implemented in RDP4 (58) with default settings.

Phylogenetic analysis is the reconstruction of the evolutionary history of organisms, in this case of different VZV strains. Based on differences and similarities in their genomes, i.e., single nucleotide polymorphisms, the strains are typically arranged as external nodes in bifurcating trees, so-called phylogenetic or evolutionary trees. The strains are supposed to have descended from common ancestors that are represented in the phylogenetic tree as internal nodes to which the external nodes are attached. However, a limitation with these traditional phylogenetic trees is that they are bifurcating and assume that recombination has not contributed to the tree-building process. Therefore, traditional phylogenetic trees should reflect the true evolutionary history of the respective genomic segment, but none reflect the correct evolutionary history for the complete genomes. The traditional bifurcating phylogenetic tree based on the complete genome (D, top) cannot illustrate recombination events but instead places the recombinant strain separately from the other strains since it contain parts from both parental strains, which the algorithms interpret as separate evolution. Hence, this tree does not reflect the true evolutionary history of any genomic segment. The use of traditional phylogenetic trees should therefore be avoided when recombinant genomes are included in the analysis. The phylogenetic network (D, bottom), however, correctly illustrates the evolutionary relationships between the different strains. The parallel edges in the network illustrate the conflicting phylogenetic signals resulting from the multiple evolutionary histories found in the recombinant strain. In phylogenetic networks, parallel branches or internal “boxes” thus indicate that the data sets have conflicting phylogenetic signals. These conflicting signals may be a result of either convergent evolution or recombination.
Recombination analysis. The detection of recombinant strains was carried out by analyzing conflicting phylogenetic signals (SNPs). Phylogenetically incompatible sites in a sequence alignment can be a result of either recombination or of convergent evolution. For both reasons, the result of recombination. In addition, the informative site analysis tool implemented in the SimPlot program was used to detect and illustrate conflicting phylogenetic signals in selected strains, and putative interclade recombinants were also analyzed using the bootstrap algorithm implemented in SimPlot.

**Nucleotide sequence accession numbers.** Consensus sequences for all samples sequenced in this study are available in GenBank under the accession numbers shown in Table 1.

**RESULTS**

**Phylogenetic analysis.** We sequenced the complete genomes of 37 clinical strains (Table 1) isolated from patients in Europe and Singapore. We also accessed publicly available complete VZV genomes from GenBank. In total, 115 complete VZV genomes were caused by recombination or by convergent evolution. First, we used the program SplitsTree (56) to create phylogenetic networks that illustrate the evolutionary relationships between the strains. When phylogenetic incongruences were detected, we used the phi-test, and algorithms included in the RDP4 program (RDP, Geneconv, Chimera, MaxChi, BootScan, SisScan, 3Seq, and LARD) to perform a statistical analysis to determine whether the conflicting phylogenetic sites were the result of recombination. In addition, the informative site analysis tool implemented in the SimPlot program was used to detect and illustrate conflicting phylogenetic signals in selected strains, and putative interclade recombinants were also analyzed using the bootstrap algorithm implemented in SimPlot.

**Phylogenetic analysis.** We sequenced the complete genomes of 37 clinical strains (Table 1) isolated from patients in Europe and Singapore. We also accessed publicly available complete VZV genomes from GenBank. In total, 115 complete VZV genomes were caused by recombination or by convergent evolution. First, we used the program SplitsTree (56) to create phylogenetic networks that illustrate the evolutionary relationships between the strains. When phylogenetic incongruences were detected, we used the phi-test, and algorithms included in the RDP4 program (RDP, Geneconv, Chimera, MaxChi, BootScan, SisScan, 3Seq, and LARD) to perform a statistical analysis to determine whether the conflicting phylogenetic sites were the result of recombination. In addition, the informative site analysis tool implemented in the SimPlot program was used to detect and illustrate conflicting phylogenetic signals in selected strains, and putative interclade recombinants were also analyzed using the bootstrap algorithm implemented in SimPlot.

**Table 1** ID, accession numbers, source (body compartment), sample collection date, and pathology of the 37 strains sequenced.

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>Accession no.</th>
<th>Source (body compartment)</th>
<th>Sample collection date</th>
<th>Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ves/Cli/ POR/Ves/5001</td>
<td>KF771921</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/ BAL/UK/2402/2009</td>
<td>KF771904</td>
<td>BAL specimen</td>
<td>24/02/09</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/ Ves/UK/1001/2012</td>
<td>KF771920</td>
<td>Vesicle fluid</td>
<td>10/01/12</td>
<td>Varicella</td>
</tr>
<tr>
<td>Zos/Cli/ UK/1801/2012</td>
<td>KF771925</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>Var/Cli/ UK/1401/2012</td>
<td>KF771906</td>
<td>EDTA blood</td>
<td>14/01/12</td>
<td>Varicella (fatal)</td>
</tr>
<tr>
<td>Var/Cli/UK/BAL/1001/2012</td>
<td>KF771905</td>
<td>BAL specimen</td>
<td>10/01/12</td>
<td>Varicella (fatal)</td>
</tr>
<tr>
<td>Zos/Cli/ GRE/Ves/03/2012</td>
<td>KF771923</td>
<td>Vesicle fluid</td>
<td>01/03/12</td>
<td>Zoster</td>
</tr>
<tr>
<td>Var/Cli/ UK/Ves/1301/2013</td>
<td>KF771911</td>
<td>Vesicle fluid</td>
<td>13/01/13</td>
<td>Zoster with encephalitis</td>
</tr>
<tr>
<td>Var/Cli/ UK/Ves/0706/2013</td>
<td>KF771910</td>
<td>Vesicle fluid</td>
<td>07/06/13</td>
<td>Zoster with encephalitis</td>
</tr>
<tr>
<td>Var/Cli/ UK/Ves/2203/2013</td>
<td>KF771912</td>
<td>Vesicle fluid</td>
<td>22/03/13</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/ UK/Ves/2403/2013</td>
<td>KF771913</td>
<td>Vesicle fluid</td>
<td>24/03/13</td>
<td>Varicella (fatal)</td>
</tr>
<tr>
<td>Var/Cli/ UK/CSF/2912/2012</td>
<td>KF771908</td>
<td>CSF</td>
<td>29/12/12</td>
<td>Varicella with encephalitis</td>
</tr>
<tr>
<td>Var/Cli/ UK/Ves/0201/2013</td>
<td>KF771909</td>
<td>Vesicle fluid</td>
<td>02/01/13</td>
<td>Varicella with encephalitis</td>
</tr>
<tr>
<td>Var/Cli/ UK/CSF/0102/2013</td>
<td>KF771907</td>
<td>CSF</td>
<td>01/02/13</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>DE10-1515</td>
<td>KF771891</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-2480</td>
<td>KF771892</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-2660</td>
<td>KF771893</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-2704</td>
<td>KF771894</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-3378</td>
<td>KF771895</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-4367</td>
<td>KF771896</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-4582</td>
<td>KF771897</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-5454</td>
<td>KF771898</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-567</td>
<td>KF771899</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>DE10-581</td>
<td>KF771900</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>KVe-2313</td>
<td>KF771901</td>
<td>CSF</td>
<td>Not known</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>KVe-3127</td>
<td>KF771902</td>
<td>CSF</td>
<td>Not known</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>KVe-1390</td>
<td>KF771903</td>
<td>CSF</td>
<td>Not known</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>Zos/Cli/Ves/SING/9</td>
<td>KF771924</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Zoster</td>
</tr>
<tr>
<td>Cli/UK/CSF/2909/2011</td>
<td>KF771889</td>
<td>CSF</td>
<td>29/09/11</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>Cli/UK/CSF/3009/2011</td>
<td>KF771890</td>
<td>CSF</td>
<td>30/09/11</td>
<td>Encephalitis</td>
</tr>
<tr>
<td>Zos/Cli/CSF/SING/1008/2008</td>
<td>KF771922</td>
<td>CSF</td>
<td>10/08/08</td>
<td>Zoster and cerebellitis</td>
</tr>
<tr>
<td>Var/Cli/Ves/SING/1308/2008</td>
<td>KF771919</td>
<td>Vesicle fluid</td>
<td>13/08/08</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/Ves/GER/31/2005</td>
<td>KF771915</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/Ves/GER/43/2006</td>
<td>KF771916</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/Ves/ITA/51/2006</td>
<td>KF771918</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/Ves/GER/63/2006</td>
<td>KF771917</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
<tr>
<td>Var/Cli/Ves/FRA/98/2013</td>
<td>KF771914</td>
<td>Vesicle fluid</td>
<td>Not known</td>
<td>Varicella</td>
</tr>
</tbody>
</table>

**a** BAL, bronchoalveolar lavage.

**b** Dates are in the format day/month/year.
aligned and analyzed. All regions containing gaps or ambiguous sites were excluded prior to any analysis.

A phylogenetic network based on all strains was initially constructed. The network demonstrated a divergence into the previously described five clades designated 1 to 5 (15–17) but also showed six strains that did not cluster to any of these clades (Fig. 2.). One of the strains sequenced here, Var/Cli/Ves/France/98/2013, and strain Var160 available from GenBank clustered together and therefore fulfilled the requirements for clade designation agreed on during the VZV nomenclature meeting in 2008 (17). We designated this clade 6. Furthermore, strains Cli/UK/CSF/2909/2011 and Cli/UK/CSF/3009/2011 clustered separately as two novel clade candidates. Strains 1483/2005 and 457 also clustered separately as clade candidates as suggested previously (29).

**Interclade recombination.** Interclade recombinants are defined as recombinant strains containing two or more genomic segments that phylogenetically cluster closely to two or more clades, respectively, i.e., a recombinant strain that has descended from parental strains from different phylogenetic clades. If there are recombination crossovers that are present in all strains in a specific clade, we assume that these crossovers are descended from an ancient recombination event that affected a common ancestor to all strains in that clade. These recombination events are therefore classified as ancient (since the recombination event occurred prior to the divergence of that clade).

In the phylogenetic network based on all strains, there are four outliers (highlighted in red in Fig. 2) that we consider to be either interclade recombinant candidates or single representatives of
novel VZV clades. To further resolve the evolutionary history of VZV and to reveal ancient interclade recombination events, we created a consensus strain for each clade. Owing to the geographical separation of strains, there might be a sampling bias favoring strains from certain clades. Here, all clades or clade candidates are treated equally to resolve the extent of interclade recombination of VZV. A new phylogenetic network was created based on the consensus sequences from all strains within each of the six clades and the four outliers. In total, 10 sequences were included in the analysis. We also performed a bootscan analysis to detect and illustrate recombination crossovers. The phi-test for statistical analysis of recombination was applied on the data set, and the informative site analysis was applied on the recent interclade recombinant candidate strain Cli/UK/CSF/2909/2011.

The topology of the network (Fig. 3A) resembles that of the network including all strains and suggests conflicting phylogenetic signals. The topology is star shaped, which is typical when there has been an evolutionary history with an exponential population growth and/or massive recombination. Furthermore, the parallel internal branches illustrate phylogenetic incongruences, which is typical when there are several recombination events and/or convergent evolution. The phi-test demonstrated high statistical significance for recombination in favor over convergent evolution ($P < 0.00$), and the bootscan analysis demonstrated highly shifting phylogenetic support indicating numerous recombination crossovers (Fig. 3A). The RDP4 program detected three distinct recombination events in six recombinant clades (clades 3, 4, 5, and 6 and strains Cli/UK/CSF/2909/2011 and 1483/2005) (Fig. 3B). Although the bootscan analysis of interclade recombinant candidate strain Cli/UK/CSF/2909/2011 did not demonstrate convincing evidence for recombination, the informative site analysis of the same strain demonstrated segmented differences in phylogenetic support (Fig. 3C).

**Intraclade recombination.** An intraclade recombinant is de-
fined as a recombinant strain in a specific clade that has descended from parental strains from that same clade. All intraclade recombinants are, by definition, regarded as recent recombinant strains (since the recombination event occurred after the divergence of that clade).

To estimate the frequency of intraclade recombination events of VZV, we independently analyzed strains from each clade harboring four or more strains, e.g., clades 1, 2, 3, and 5 (most methods require at least four strains in order to analyze recombination events). First, five phylogenetic networks were constructed based on the strains from each clade, respectively, to detect and illustrate the presence of conflicting phylogenetic signals. Then, a phi-test for statistical analysis of recombination was applied to each data set to determine whether there was statistical support for recombination within each clade. In a further attempt to detect and classify intraclade recombination events with statistical significance, we applied the methods RDP, Geneconv, Chimaera, MaxChi, BootScan, SisScan, 3Seq, and LARD implemented in the RDP4 program on each sequence. Finally, we analyzed the distribution of conflicting sites among selected strains using the informative site analysis tool implemented in the SimPlot program. The informative sites analysis also made it possible to estimate the frequency of recombinant candidates, (i.e., strains with conflicting phylogenetic signals but too few SNPs that differ between the putative parental strains in order to obtain a statistically significant measure of recombination).

**Recombination analysis of strains within clade 1.** The network based on clade 1 strains demonstrated a few internal parallel edges suggesting phylogenetic incongruences (Fig. 4A). The existence of recombinant strains was statistically supported ($P = 0.041$) by the phi-test, and putative recombination crossovers were detected by using the informative sites analysis tool (Fig. 4B and C). In contrast, the RDP4 program was unable to detect any recombination events among the clade 1 strains with statistical significance.

**Recombination analysis of strains within clade 2.** The sequences of strains pOka and vOka from GenBank contain several ambiguous sites (denoted “N,” “Y,” etc.). As a precaution, all these sites were excluded prior to the analysis, resulting in a data set (denoted data set 1) with less information. In order to analyze additional information in the other strains, we also analyzed the original data set excluding strains pOka and vOka (denoted data set 2), i.e., two data sets were therefore analyzed in parallel for strains within clade 2.

The phylogenetic networks based on both data sets contained parallel branches demonstrating conflicting phylogenetic signals (Fig. 5A and C). The phi-test also provided high statistical significance ($P = 0.018$) for recombination among strains in data set 1 but not for data set 2 ($P = 0.069$). The RDP4 program detected several recombinant strains and recombination crossovers in both data sets (Fig. 5B and D), and recombination was statistically supported by the methods RDP, BootScan, MaxChi, Chimaera, and...
In addition, at least five recombinants and eight recombination crossovers were also detected and clearly visualized by the informative sites analysis tool (Fig. 6A to E).

**Recombination analysis of strains within clade 3.** Phylogenetic incongruences were clearly illustrated in the clade 3 network, but the phi-test showed that these were not statistically significant ($P = 0.23$) for recombination when all strains were analyzed (Fig. 7A). However, when putative nonrecombinants were manually identified and excluded from the analysis, the phi-test did show statistically significant ($P = 0.001$) signals for recombination. At least two recombinants and two recombination crossovers were independently detected by using the informative sites analysis tool (Fig. 7B and C) and the RDP4 program (Fig. 7D), respectively.

**Recombination analysis of strains within clade 5.** The analysis of clade 5 strains was similar to that of clade 3 strains. The network demonstrated internal parallel branches illustrating phylogenetic incongruences (Fig. 8A). The phi-test again resulted in statistically significant recombination when putative nonrecombinants were manually identified and excluded from the analysis ($P = 0.006$) but not when all strains were analyzed ($P = 0.39$). The informative sites analysis tool identified at least two different recombination events (Fig. 8B and C), and the methods MaxChi and Chimaera implemented in the RDP4 program identified one recombination event with statistical significance (Fig. 8D).

**DISCUSSION**

Here, we have demonstrated that homologous recombination occurs frequently between wild-type VZV strains. We have identified...
fied putative novel interclade recombinants and well-supported frequent intraclade recombination within clades 1, 2, 3, and 5. The VZV genome is significantly more conserved than those of HSV-1 and HSV-2. Excluding the variable repeat regions, there are approximately 30 to 200 SNP differences between any two VZV strains (60). DNA data from recombinant and parental strains may thus contain too few SNPs differing between the parental strains, resulting in no or only few parental-specific nucleotides transferred to the recombinant strains. It will hence be difficult to detect recombination events between similar parental strains or to classify them with high statistical significance. It is therefore likely that many of the strains analyzed here that contain conflicting phylogenetic signals but have putative parental strains that are too closely related to yield statistical significance for recombination are indeed true recombinants. Thus, the frequency of recombination is potentially even higher for VZV than reported here.

Little is known about when, how, and where recombination occurs, other than that recombination requires the simultaneous replication of two or more distinct viruses in the same cell. Recombination thus requires multiple infections with two or more distinct strains. It is, however, uncertain if such multiple infections occur through primary infections with heterogeneous virus populations, through simultaneous multiple primary infections from different sources, or by re-infections later in life, scenarios that all are probable. Furthermore, while a recombinant virus may be produced within a person, it is uncertain if the recombination event happens directly after infection or during latency or reactivation or what kind of mechanism may trigger recombination. Interestingly, two of the few putative interclade recombinants detected here (Cli/UK/CSF/3009/2011 and Cli/UK/CSF/2909/2011) were collected from cerebrospinal fluid (CSF) from patients with encephalitis. It may therefore be speculated that central nervous system (CNS) infections facilitate the replication of different viruses in the same cell or vice versa. Although this cannot explain all the recombination events of VZV, further studies on VZV strains collected from CSF may be warranted to better understand the underlying mechanisms of VZV recombination and of VZV CNS infections.

Clade 2, which harbors the vaccine strains, showed evidence of multiple intraclade recombination events. Although ambiguities remain about which strains are parental and which are recombinants, our data suggest that either the original vaccine strains or the vaccine strains isolated from lesions are intraclade recombinants. It is therefore likely that the vaccine strains have already been involved in recombination events, either in vivo or in vitro. The vaccine preparation has previously been shown to be heterogeneous (49, 61) and thus may have recombined during passaging in vitro or within the host following inoculation. In any case, it is only the detection of interclade vaccine/wild-type recombinants that can reliably reveal in vivo recombination between vaccine and wild-type strains after vaccination, since no strains from other clades are present in the vaccine batches (49).

However, vaccination with Zostavax will undoubtedly result in individuals infected with multiple strains, since most adults are already carrying latent VZV strains. Since Zostavax is based on clade 2 strains, most vaccinated persons living outside Asia (where clade 2 is predominant) will most likely also be carrying multiple strains from different clades. Despite thorough analysis, we did not, however, detect any interclade vaccine/wild-type recombi-
nants. These results may suggest that the probability of interclade VZV vaccine/wild-type recombinants emerging is low, but this may not necessarily be the case. First, we have analyzed only a limited number of strains (n = 115). Second, as mentioned above, it is unknown when and where the actual recombination event occurs. It may thus take some time after vaccination before a recombinant emerges and before we can detect it. The best chances of detection of vaccine/wild-type recombinants may therefore be to analyze reactivating zoster in elderly patients vaccinated with Zostavax, which makes an interesting avenue for further studies.

One intriguing question is when the recombination events detected here occurred. Unfortunately, exact dating of recombination events is troublesome and requires knowledge about the parental strains involved in the process. Assuming we actually had that knowledge, the different segments in the recombinant strains retrieved from respective parental strain could be identified and compared with the corresponding regions in the recombinant strains. This would then allow us to estimate the number of substitutions that have taken place since the recombination event. Estimations of the molecular clock rate (substitutions/site/year) could then be used in combination with the number of SNPs to calculate the number of years since the recombination event. See FIG 7 for an example of how this can be done using the RDP program and statistical support yielded by each method for the respective recombination event.

FIG 7 Intraclade recombination analysis of clade 3. (A) Phylogenetic network based on the complete genomes of strains from clade 3 to illustrate phylogenetic incongruences. Arrows highlight recombination crossovers supported by two or more mutations on each side of the crossover. (B) Recombinant strains identified by the RDP program and statistical support yielded by each method for the respective recombination event.
However, in this case we have no knowledge of the parental strains of the recombinants. At best, we might have DNA data from strains closely related to the parental strains, i.e., that are similar to the parental strains but differ from them by several SNPs. If we assumed that these are the true parental strains, these SNPs would mistakenly be interpreted to have originated after the recombination event. This would result in the predicted date of recombination being older than the real date. Due to these circumstances and given the uncertainties of the molecular clock rate, we have therefore not made any attempts to date single recombination events in this study, other than classifying them as recent or ancient based on the relation to clade divergence.

Several attempts have been made to distinguish recombinant clades from parental clades, i.e., which clades descend from recombinants and which are not recombinants (14–16). However, these kinds of analyses require assumptions on the total global genetic divergence and the exclusion of the possibility of other existing clades. Since we have only 115 genomes and the number of defined clades has increased as more strains are sequenced, such assumptions are likely to be erroneous. Furthermore, the recombination and parental history become increasingly complicated as more clades are included, and we cannot exclude that all clades have been involved in recombination events at some occasion. We therefore avoided distinguishing between parental and recombinant clades in the present study, but we conclude that multiple ancient interclade recombination events have occurred in the evolutionary history of VZV. We also suggest that the complexity of the recombination crossover patterns will increase as more strains are sequenced and more clades are discovered. Future studies that produce whole-genome sequences for strains from various geographic regions may shed further light on these issues.

It was suggested during the VZV nomenclature meeting in 2008 (17) that the five major VZV clades should be designated 1 to 5 and that novel strains should be designated members of an existing clade, based on phylogenetic clustering. The intention was also to introduce criteria so that each existing clade and future novel clades should have at least two whole genomes sequenced to be considered true clades. Although additional clades have been

![FIG 8 Intraclade recombination analysis of clade 5. (A) Phylogenetic network based on the complete genomes of strains from clade 5. (B and C) Simplot analysis of arbitrarily selected putative recombinant strains from clade 5 to illustrate phylogenetic incongruences. Arrows highlight recombination crossovers supported by two or more mutations on each side of the crossover. (D) Recombinant strains from clade 5 suggested by the RDP program and statistical support yielded by each method for respective recombination event.](http://jvi.asm.org/.../6,2017 by guest http://jvi.asm.org/ Downloaded from)
suggested (29, 62, 63), only clades 1 to 5 have hitherto fulfilled this requirement. One strain sequenced here (Var/Cli/Var/1006/2013), however, clusters phylogenetically with another strain (Var/Cli/Var/1006/2013; GenBank accession number KC112914.1), separately from strains from clades 1 to 5. The requirements for a sixth clade are therefore now fulfilled, and we designate this clade 6. Interestingly, these two strains that form clade 6 were collected in widely separated geographic regions. Strain Var160 was isolated from vesicle fluid from a 7-year-old boy in Mexico in 2007. Strain Var/Cli/Var/1006/2013, on the other hand, was isolated from vesicle fluid from a 2-year-old girl with chicken pox in France in 2013. This finding indicates that clade 6 may have a geographic distribution similar to that of clades 1 and 3, which are predominant in Europe and North America, although additional isolates are required to define its predominant geographic locations.

The high frequency of recombination detected here further supports previous suggestions that interclade recombination of VZV could dramatically increase in the future (22). Increased migration and the introduction of live vaccine strains from clade 2 in geographical regions, where strains from other clades are predominant, will undoubtedly erase previous geographical barriers between clades. The rates of interclade recombination should also increase as the geographic barriers diminish, and it is possible that most VZV strains will eventually have mosaic recombinant genomes similar to those in HSV-1 (22, 23). If so, a challenging question will be how to define a VZV clade in the future. It will be highly impractical to continue with the clade designation on a complete-genome level, since every unique interclade recombinant will cluster separately in a phylogenetic tree or network, as described for HSV-1 (22). Although it may take several decades before the majority of VZV strains are mosaic recombinants, the results from this study suggest that several wild-type recombinant strains have already emerged. Collaborative discussions about future VZV nomenclature, classification, and genotyping are thus important in the near future.

ACKNOWLEDGMENTS

We acknowledge the infrastructure support provided by MRC Centre for Molecular Medical Virology grant G0900950 and by the NIHR UCL/UCLH Biomedical Research Centre 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 MRC grant G0700814, an MRK New Investigator Award for D.P.D., and the NIHR UCL/UCLH Biomedical Research Centre for S.K. and J.B. This work was also supported by the Swedish Medical Research Council (grant number 521-2011-3297 to T.B.) and the LUA-ALF Foundations of Sahlgren’s University Hospital (grant number 145-841 to T.B.).

REFERENCES

5. Frenos S, Galli I, Chiappini E, de Martino M. 2007. An increasing incidence of chickenpox central nervous system complications in chil-


ACKNOWLEDGMENTS

We acknowledge the infrastructure support provided by MRC Centre for Molecular Medical Virology grant G0900950 and by the NIHR UCL/UCLH Biomedical Research Centre 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 MRC grant G0700814, an MRK New Investigator Award for D.P.D., and the NIHR UCL/UCLH Biomedical Research Centre for S.K. and J.B. This work was also supported by the Swedish Medical Research Council (grant number 521-2011-3297 to T.B.) and the LUA-ALF Foundations of Sahlgren’s University Hospital (grant number 145-841 to T.B.).

REFERENCES


Downloaded from http://jvi.asm.org on November 6, 2017 by guest


