

# Possible European Origin of Circulating Varicella Zoster Virus Strains

Chiara Pontremoli,<sup>1,\*</sup> Diego Forni,<sup>1</sup> Mario Clerici,<sup>2,3</sup> Rachele Cagliani,<sup>1</sup> and Manuela Sironi<sup>1</sup>

<sup>1</sup>Istituto di ricovero e cura a carattere scientifico (IRCCS) E. Medea, Bioinformatics, Bosisio Parini, and <sup>2</sup>Department of Physiopathology and Transplantation, University of Milan, and <sup>3</sup>IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

(See the Editorial Commentary by Breuer, on pages 1213–5.)

Varicella zoster virus (VZV) is the causative agent of chickenpox and shingles. The geographic distribution of VZV clades was taken as evidence that VZV migrated out of Africa with human populations. We show that extant VZV strains most likely originated in Europe and not in Africa. Europe was also identified as the ancestral location for most internal nodes of the VZV phylogeny, including the ancestor of clade 5 strains. We also show that strains from clades 1, 2, 3, and 5 derived a major proportion of their ancestry from each of 4 ancestral populations. Conversely, viruses from other clades displayed variable levels of admixture. Some low-level admixture was also observed for clade 5 genomes, but only for non-African viruses. This pattern indicates that the clade 5 VZV strains do not represent recent introductions from Africa due to migratory fluxes. These data have also relevance for the definition and classification of VZV clades.

**Keywords.** varicella zoster virus; phylogeography; origin; admixture

Varicella zoster virus (VZV) is a human herpesvirus belonging to the genus *Varicellovirus* of the Alphaherpesvirinae subfamily (family Herpesviridae). Also known as human alphaherpesvirus 3, VZV has a double-strand DNA genome about 125 kb in length.

On first infection, VZV causes varicella (chickenpox) and establishes lifelong latency in sensory and autonomic ganglia. VZV reactivation, usually during adulthood, causes herpes zoster (shingles). Both chickenpox and herpes zoster can lead to severe complications, including encephalitis, vasculopathy, and pneumonia [1–3]. The introduction of a VZV vaccine based on a live attenuated strain (vOka) has reduced the health burden imposed by VZV in several countries [4]. However, the virus causes a considerable disease-associated morbidity and mortality in regions where vaccination coverage is limited [4].

VZV is the only human herpesvirus to be transmitted primarily by aerosolization. Thus, horizontal transmission is common and typically associated with chickenpox outbreaks [4]. In temperate regions, varicella is highly contagious and virtually all susceptible (nonvaccinated) individuals acquire VZV early in life [4]. In tropical areas, VZV infectivity is lower, and chickenpox usually occurs at a later age [5]. Herpes zoster is also contagious, and the ability of VZV to reactivate after decades and

to originate new transmission chains was suggested to represent an evolutionary strategy of survival in small human populations (eg, hunter-gatherer early human groups) [6].

In fact, several lines of evidence indicate that VZV originated in Africa and spread worldwide with its human hosts. First, virus-host cospeciation has probably driven the evolution of varicelloviruses [7]. Second, the closest relative of VZV (with approximately 70% sequence identity) is *Cercopithecine alphaherpesvirus 9* (also known as simian varicella virus), which infects Old World monkeys [8–10]. Third, molecular dating approaches using the split time of the Cercopithecidae and Hominoidea lineages as a calibration point dated the origin of VZV about 110 000 years ago [7]—that is, before the onset of major human out-of-Africa migrations, which started about 60 000 years ago [11]. However, an approach based on tip dating estimated that the most recent common ancestor (MRCA) of VZV existed 309 years ago [12]. Although this result might have been influenced by the lack of temporal signal in the VZV data set, a more recent estimate, based on the calculation of within-patient evolutionary rates of vaccine strains, dated the MRCA of VZV about 5000 years ago [13]. In line with this observation, VZV substitution rates calculated from interhost data were estimated to be similar to those of other herpesviruses and thus inconsistent with the out-of-Africa migration hypothesis [12, 14, 15].

On the basis of genome sequence similarity, VZV strains have been divided into 9 clades: 7 established (designated 1–6 and 9) and 2 provisional (VII and VIII) [16, 17]. The major VZV clades tend to be associated with specific geographic regions. In particular, clades 1, 3, and 6 are mainly transmitted in Europe,

Received 12 February 2019; editorial decision 21 March 2019; accepted 2 May 2019; published online May 3, 2019.

Correspondence: Chiara Pontremoli, Bioinformatics, Scientific Institute, IRCCS E.MEDEA, Via Don L. Monza 20, 23842 Bosisio Parini, Italy (chiara.pontremoli@bp.inf.it).

The Journal of Infectious Diseases® 2020;221:1286–94

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com. DOI: 10.1093/infdis/jiz227

North America, and Australia, whereas clade 2 is primarily Asian. Clades 4 and 5 tend to be more widespread, and clade 5 is the only one to be reported in Africa [18]. Such spatial distribution was also taken as evidence that VZV originated in Africa and dispersed with human populations [6, 19, 20]. However, no formal testing of the African origin of VZV strains has ever been provided. In the current study, we used different approaches to reconstruct the phylogeography of extant VZV clades and to analyze VZV population structure. Multiple lines of evidence suggest a European origin for circulating VZV strains.

## METHODS

### Sequences, Alignments, Linkage, and Recombination Analyses

We retrieved the sequences for 110 human VZV strains from the Virus Pathogen Resource (ViPR; <https://www.viprbrc.org/>) and National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>) databases. We only used complete genome sequences, and we removed laboratory, vaccine, and highly passaged strains. A list of accession number is reported in [Supplementary Table 1](#). We used MAFFT software (version 7.392) [21] to generate multiple sequence alignments.

A subset of partial sequences of open reading frames (ORFs) 31 (2729 base pairs [bp]) through ORF38 (462 bp)–ORF54 (436 bp) for 16 Indian VZV strains were also obtained from the ViPR and NCBI databases. These ORFs were selected to maximize the number of available Asian sequences and the length of the sequenced regions.

From the complete genome alignment, 799 parsimony-informative sites (sites that contain  $\geq 2$  types of nucleotides, each with a minimum frequency of 2) were obtained. These sites were used as the input for linkage disequilibrium (LD) and population structure analyses. LD was evaluated using LIAN software (version 3.7) [22] to test the null hypothesis of linkage equilibrium for all loci. The Monte Carlo test option was chosen, with 10 000 iterations.

We screened alignments for the presence of recombination using Recco software (version 0.93) [23], a method based on cost minimization, as well as with 4 methods implemented in the RDP4 program. The Recco software's output includes a *P* value for the whole data set that, controlling for false-positives, provides an indication as to whether a significant amount of recombination is detectable in the whole alignment. In this case, the *P* value was equal to .0099. Sequences were considered to be recombinants if the number of savings was  $>20$ , and the sequence *P* value was  $<.01$ , as suggested elsewhere [24]. Break points were retrieved from the Recco software output for individual recombination events.

For the RDP4 program, 4 methods (RDP, GENECONV, MaxChi, and Chimera) [25–28] were used because they showed good power in previous simulation analyses [27]. The cutoff *P* value was set to .05.

### Phylogeography

To reconstruct the geographic origin of extant VZV strains, we assigned sequences to 8 macro areas based on the United

Nation geographic subregions [29]: Europe, Eastern Asia, South-Eastern Asia, Southern Asia, Northern Africa, Western Africa, Northern America, and Central America. To obtain inference of evolutionary rates and geographic origin, we used the discrete model [30] implemented in the Bayesian Evolutionary Analysis by Sampling Trees (BEAST; version 2.5) software [31]. We selected a General Time Reversible (GTR) substitution model with gamma distribution plus invariant sites (GTR+G+I) using JmodelTest 2 [32, 33]. We ran the path sampling tool implemented in BEAST to choose between a constant, an exponential, or a coalescent Bayesian skyline tree prior (50 steps, with 1 000 000 iterations each).

A Bayes factor test was applied to compare the different likelihoods. Factors of 3–20 suggest positive support; 20–150, strong support; and  $>150$ , overwhelming support. The coalescent Bayesian skyline tree prior was favored (see [Supplementary Table 2](#)). Analyses were performed using a Bayesian Markov chain Monte Carlo (MCMC) method with a relaxed log normal clock. We performed 2 different runs, with 100 million iterations each, and we sampled every 10 000 steps after a 10% burn-in. Runs were combined after checking for convergence. We generated a maximum clade credibility tree using TreeAnnotator software (version 2.5.0) [31], which was visualized using FigTree software (version 1.4.2.) (<http://tree.bio.ed.ac.uk/>). We also inferred geographic origin using the Bayesian binary MCMC (BBM) and S-DIVA (Statistical Dispersal-Vicariance Analysis (S-DIVA) methods implemented in the RASP (Reconstruct Ancestral State in Phylogenies) program [34–36].

For both BBM and S-DIVA, 10 000 BEAST-generated trees and a consensus tree were used as an input. Two BBM chains were run for 100 000 generations with estimated state frequencies, [37] a gamma-distributed among-site rate variation, sampling every 100 generations, and a maximum of 2 areas for each node. S-DIVA was run with default parameters and a maximum of 2 areas for each node. BBM allows null character status information for a portion of input sequences. This property was exploited to run 100 analyses to check for consistency against the skewed geographic origin of available sequences. In particular, 100 distributions were generated for BBM analysis so as to include the same number of sequences from Africa and Europe ( $n = 25$ ). For each of these distribution, location probabilities were recorded for the MRCA and selected internal nodes. Probabilities were averaged across the 100 runs.

### Network and Population Structure

The neighbor-net split network was constructed with SplitsTree software (version 4.13.1) [38] using uncorrected *p* distances (the proportion of nucleotide sites at which two sequences being compared are different) and all polymorphic sites. We analyzed the structure of extant VZV populations using the STRUCTURE program (version 2.3) [39]. The allele frequency spectrum

parameter ( $\lambda$ ) was estimated by using the “estimate  $\lambda$ ” model for  $K = 1$ , as suggested elsewhere [40]. A value of  $\lambda = 0.493$  was obtained and used in all analyses.

The linkage model with correlated allele frequencies was applied [40]. Map distances were set equal to parsimony-informative site physical distances. The optimal number of populations was determined by running the model for  $K = 1$  to  $K = 10$ . For each  $K$ , 10 runs were performed with MCMC run lengths of 50 000 and 20 000 burn-in steps. The Evanno method was used to select the optimal  $K$  [41]. Results of independent runs were analyzed by permutating clusters using the CLUMPAK program [42] to generate the  $Q$ -value matrix.

## RESULTS

### LD and Recombination

We obtained a list of complete or almost complete VZV genomes from public databases. After we excluded vaccine-derived strains and highly passaged/attenuated viruses, a set of 110 sequences with known geographic origin remained (see [Supplementary Table 1](#)).

Recombination has been documented for VZV [7, 17, 43]. Therefore, we investigated the level of LD using LIAN software (version 3.7), which tests the null hypothesis of linkage equilibrium across loci [22]. Statistically significant LD was detected (Monte Carlo simulations, 10 000 repetitions;  $P < 10^{-4}$ ). However, the resulting standardized index of association was equal to 0.06, consistent with moderate LD (the index of association is expected to be 0 under linkage equilibrium) and, therefore, with some level of recombination.

We thus screened the full genome alignment for the presence of recombination events. Four methods implemented in RDP4 detected no recombination. Conversely, Recco software identified 2 regions (approximately 38 000–43 000 and approximately 102 000–111 000; positions refer to the 03-500 strain, GenBank accession no. DQ479957) where recombination break points cluster ([Figure 1A](#)). The locations of these break points are consistent with a previous analysis [13].

### VZV Phylogeography

Phylogeographic analysis was carried out on the larger nonrecombining region (about 58 000 bp long) ([Figure 1A](#)). We used 3 phylogeographic methods to determine the most likely origin of extant VZV strains. In all these analyses, sequences were assigned to geographic areas (see Methods) ([Figure 1B](#)), which represent the character states for ancestral state reconstruction.

Inference of geographic origin was initially obtained using the discrete model [30] implemented with BEAST software [31]. The path sampling tool was used to select among tree priors and the Bayesian skyline model was favored (see [Supplementary Table 2](#)). BEAST software assigned the highest probability to Europe as the origin of extant VZV strains, with

a posterior probability of 0.5814 ([Figure 1B](#)). It also inferred a likely European origin for most internal nodes of the phylogeny, including the MRCA of clade 5 strains, which include all the African sequences ([Figure 1B](#)). Very similar results were obtained with the BBM and S-DIVA methods [34–36], which assigned a European origin to the ancestral node of the VZV phylogeny with high probabilities (97.27% for BBM and 91.0% for S-DIVA) ([Figure 2A](#)).

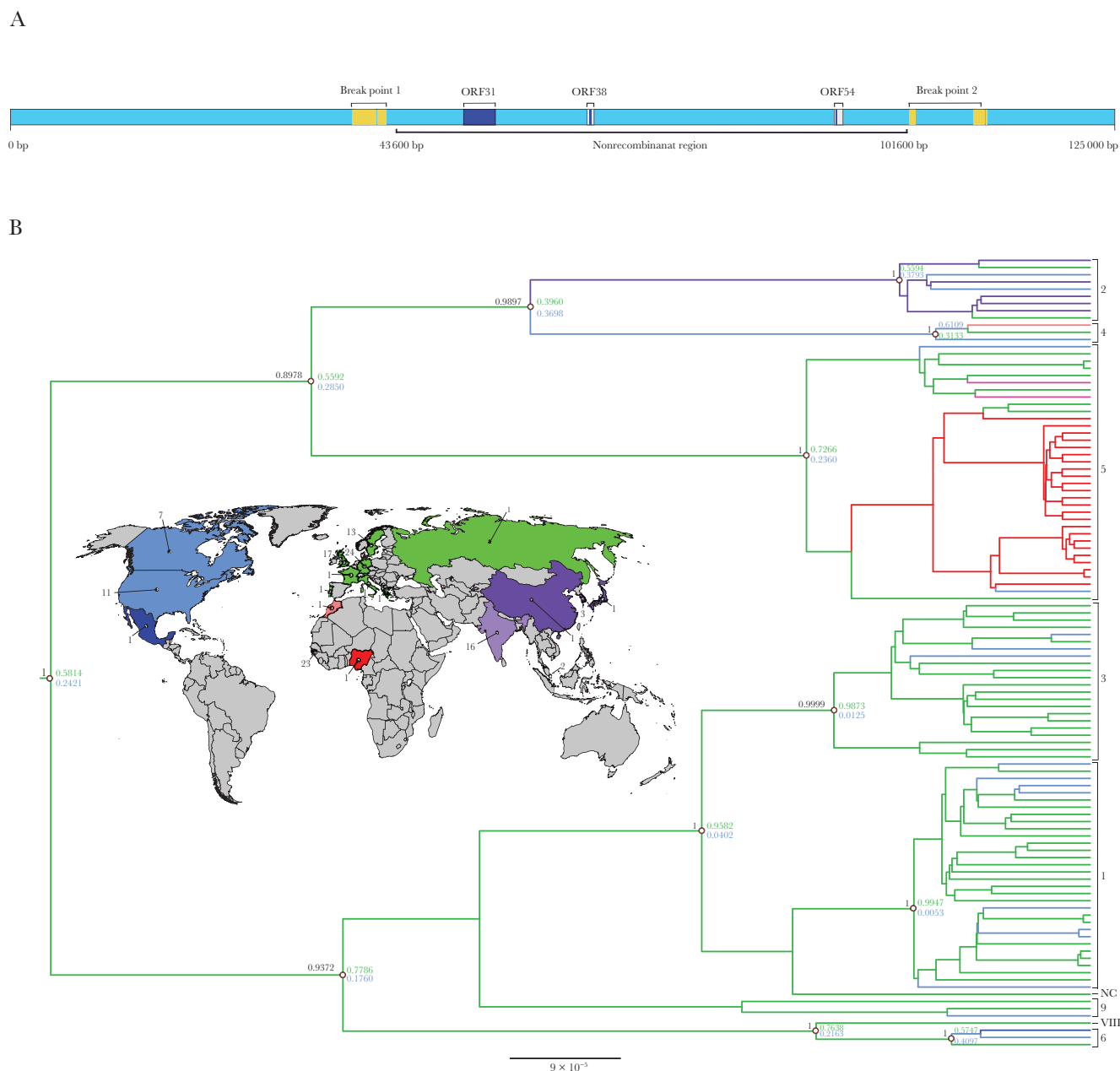
Because most VZV sequences were collected in Europe ([Figure 1B](#), see also [Supplementary Table 1](#)), we wished to assess whether this skewed distribution affected phylogeographic inference. We thus ran 100 BBM analyses by omitting the character state for 34 randomly selected European strains. This number was selected to obtain a data set that included the same number of sequences from Europe and Africa. We calculated the mean probabilities at selected nodes by averaging over the 100 BBM runs ([Figure 2B](#)). Europe was identified as the most likely origin of all extant VZV clades (the second-highest probability was assigned to Northern America) as well as the ancestral location of the MRCAs of clades 1, 3, and 5 ([Figure 2B](#)).

The data set of complete VZV genomes includes only few samples from Asia. This underrepresentation can clearly affect the accuracy of phylogeographic inference for this continent. To partially obviate this lack of sequence data, we performed an additional analysis on a shorter region deriving from the concatenation of partial ORF31-ORF38-ORF54 sequences. These ORFs are located within the nonrecombining region ([Figure 1A](#)), and they were partially sequenced in 16 Indian VZV samples. By adding these samples to those deriving from complete genomes, we obtained a data set with 23 Asian sequences. For this analysis, sequences were assigned to continents (instead of areas). A drawback of this approach is that phylogenetic relationships can be reconstructed with limited confidence when small regions are analyzed. Thus, we limited inference to the MRCA of the whole phylogeny, with no attempt to reconstruct the ancestral location of internal nodes. BEAST, BBM, and S-DIVA analyses still assigned the highest probability to Europe as the ancestral location of the VZV clade MRCA ([Figure 2C](#)).

### Population Structure and Admixture

We next sought to investigate the ancestry and admixture of VZV populations. To this aim, we used STRUCTURE, a program that relies on a Bayesian statistical model for clustering genotypes into populations without information on their geographic origin. STRUCTURE identifies distinct genetic populations (or clusters) that compose the overall population and assigns each individual to  $\geq 1$  of those populations. The  $K$  parameter defines the number of subpopulations and the more likely number of clusters is usually chosen by comparing models with different value of  $K$ .

Complete genome sequences of the 110 strains was used for STRUCTURE analysis with the linkage model [39, 40]. This

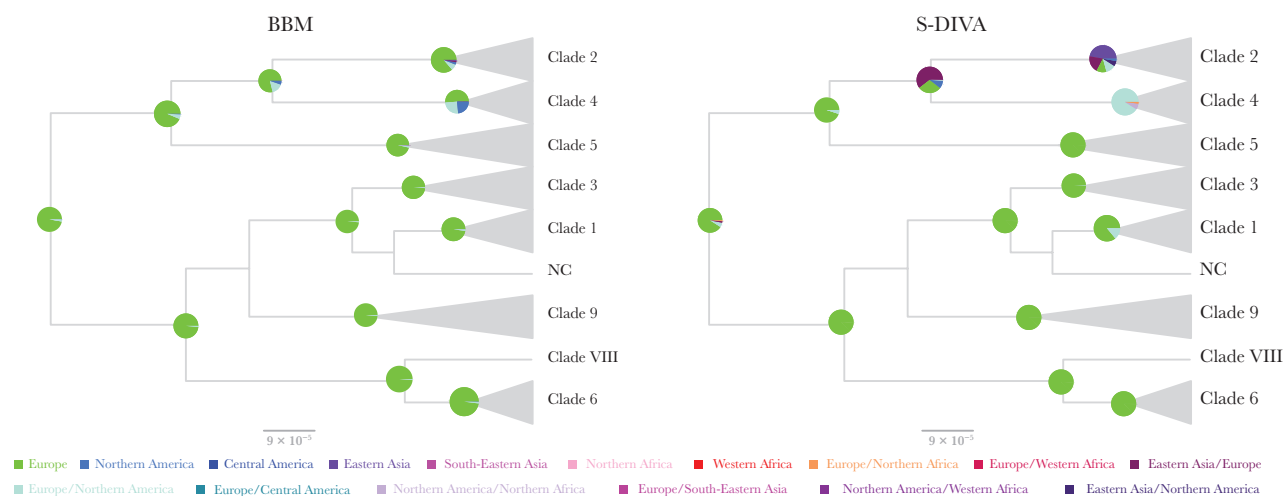


**Figure 1.** Varicella zoster virus (VZV) recombination and phylogeography. *A*, Schematic representation of the VZV genome. Positions refer to the 03-500 strain (GenBank accession no. [DQ479957](#)). The larger nonrecombinant region we used for phylogeographic analysis is shown, as well as the open reading frame (ORF) 31, ORF38, and ORF54 portions (blue) used for the analysis with the additional Indian strains. The location of the recombination break points obtained using Recco software is also reported. *B*, Geographic distribution of the VZV strains and Bayesian Evolutionary Analysis by Sampling Trees (BEAST) maximum credibility tree for the larger nonrecombinant region. Branches are colored according to inferred ancestral location; the posterior support for relevant nodes is shown in black, and the posterior support for location is colored according to the location to which it refers. VZV clades are indicated to the right of the branches; not classified (NC) refers to the Cti/UK/CSF/2909/2011 strain (ID KP771889), a possible interclade recombinant [17]. The number of available VZV sequences from each location is reported on the map. All refer to complete genomes with the exclusion of the Indian strains.

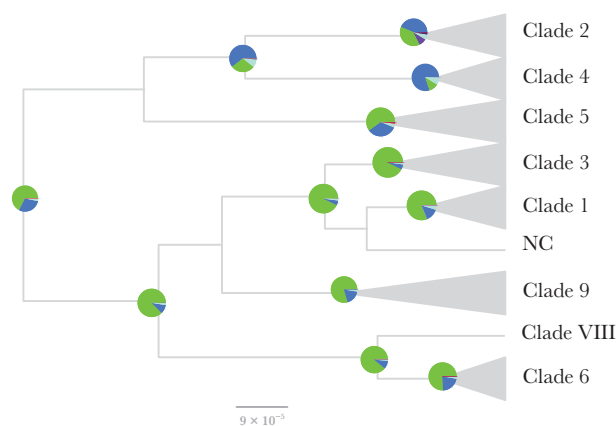
model uses LD among loci and assumes that discrete genome “chunks” are inherited from ancestral populations. The optimal number of populations ( $K$ ) was estimated to be equal to 4 using the  $\Delta K$  method (Figure 3A) [41] and individual ancestry coefficients were highly consistent across replicate runs. The population clusters inferred with STRUCTURE showed

very good correspondence to VZV clades and, therefore, appreciable congruence to geographic origin (Figure 3B and C). VZV genomes from clades 1, 2, 3, and 5 had a major proportion of their ancestry from each of the 4 ancestral populations. Based on the known geographic distribution of VZV clades [18], these ancestral populations are thus referred to as

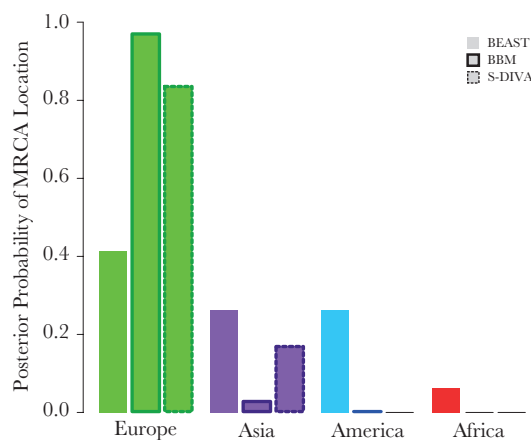
A



B



C



**Figure 2.** Phylogeographic analyses. A, Phylogeny of the nonrecombinant region; branches were collapsed at the clade level. Pie charts represent the ancestral location probability of the respective area obtained by the Bayesian binary Bayesian Markov chain Monte Carlo (BBM) analysis (*left*) and Statistical Dispersal-Vicariance Analysis (S-DIVA) (*right*). B, Same phylogeny as in A, with pie charts representing probabilities averaged over 100 BBM runs with a similar number of sequences from Europe and Africa. In BBM and S-DIVA analyses, combined regions appear because we allowed 2 areas per node. C, Posterior probability distributions for the location of the varicella zoster virus most recent common ancestor (MRCA) obtained using Bayesian Evolutionary Analysis by Sampling Trees (BEAST), BBM, and S-DIVA. Results refer to the region deriving from the concatenation of partial open reading frame (ORF) 31–ORF38–ORF54 sequences. Abbreviation: NC, not classified (Cti/UK/CSF/2909/2011 strain).

Europe-1, Asia-2, Europe-3, and Africa/Asia-5. Conversely, viruses from clades 4, 6, 9, and VIII showed variable levels of admixture from the 4 populations. Some low-level admixture was also observed for clade 5 genomes, but only for non-African viruses (Figure 3C).

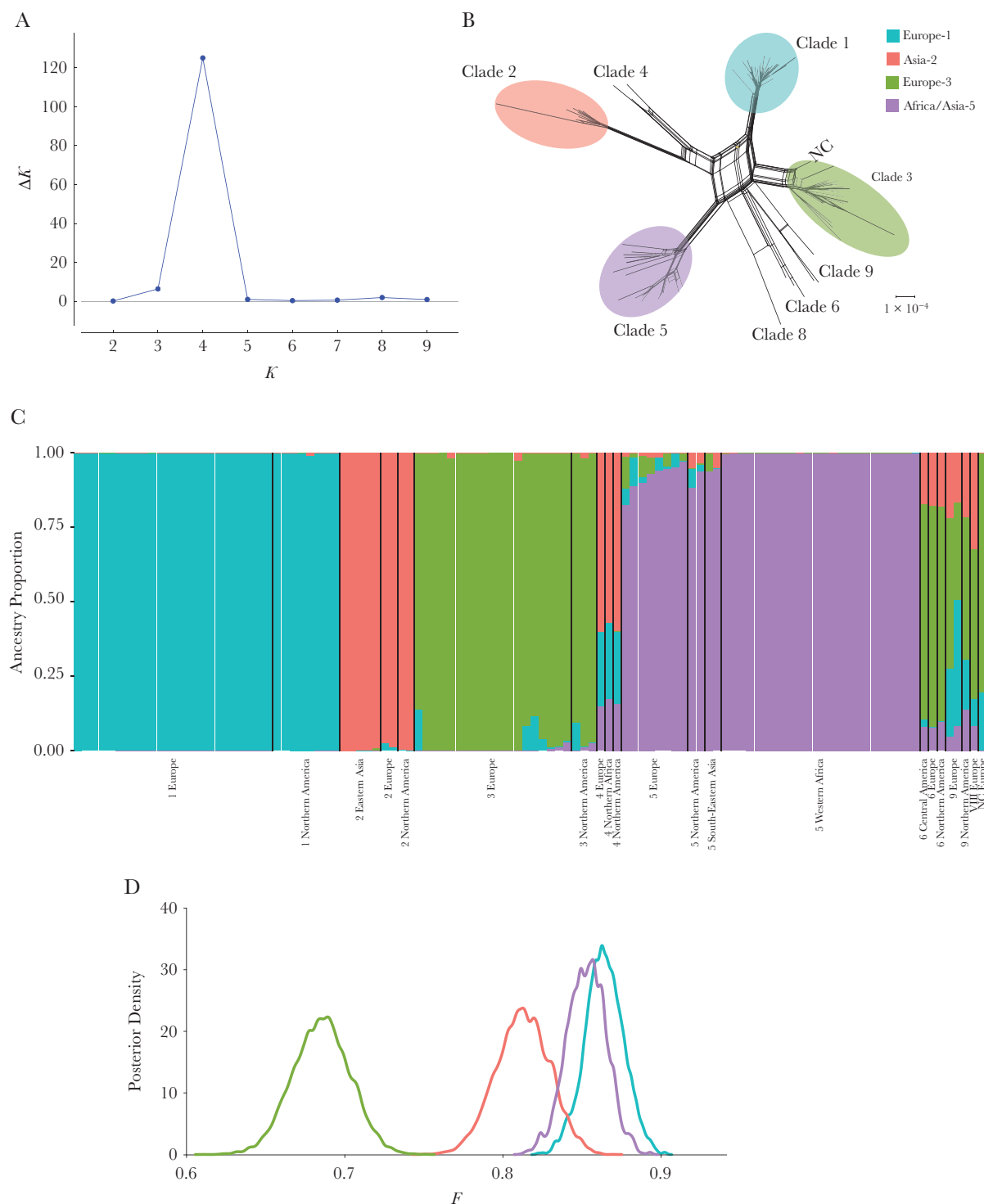
The linkage model also estimates the level of drift, measured as the  $F$  parameter (analogous to fixation index [ $F_{ST}$ ]), of each population from a hypothetical ancestral population.  $F_{ST}$  is a measure of genetic differentiation between populations based on allele frequencies. The  $F$  parameter, in an analogous way, estimates the drift of each population from the ancestral frequencies of a common population. The  $F$  value for the Europe-3 population was substantially lower than those for the 3 other populations (Figure 3D), indicating that clade 3 viruses

diverged less than the other clades from the ancestral VZV population. This result is thus consistent with a European origin of VZV.

## DISCUSSION

Herpesviruses infect a wide range of animals and most likely originated millions of years ago [6, 44]. Several studies have shown that the phylogenetic relationships among herpesviruses very often mirror those among their hosts, indicating that viral lineages frequently arose through cospeciation [6, 45, 46]. It is thus reasonable to assume that VZV is no exception and that, as its human host, the virus originated in Africa and subsequently dispersed worldwide. Based on this assumption, previous works suggested a parallelism between the geographic distribution of





**Figure 3.** Varicella zoster virus (VZV) population structure. *A*, Analysis of optimal  $K$  for STRUCTURE analysis.  $\Delta K$  is calculated as described in [41]. The peak of this distribution is the optimal  $K$  used in the subsequent STRUCTURE analysis. *B*, Neighbor-net split network of 110 complete VZV genome sequences. Viruses that derived >80% of their genome from a single ancestral population are shaded in the corresponding population color. Not classified (NC) refers to the CII/UK/CSF/2909/2011 strain (ID KP771889). *C*, Bar plot representing the result from the STRUCTURE linkage model. Each vertical line represents a VZV strain. The y-axis shows the proportion of ancestry from each of the 4 ancestral populations. Strains are ordered on the basis of clade assignment and of geographic origin. *D*, Distributions of posterior  $F$  values for the 4 ancestral populations.

VZV clades and that of human mitochondrial DNA: in both cases the out-of-Africa migration originated lineages that reached Europe and others that populated Asia [6, 19, 20].

Herein we applied a phylogeographic approach to reconstruct the origin of extant VZV clades. We applied 3 different methods, and we found no evidence that extant VZV clades originated in Africa; conversely, both phylogeography and population structure analysis indicated a likely European origin for circulating VZV strains. Clearly, a major caveat of analyses herein relates to the overrepresentation of European samples and to the skewed distribution of the overall data set. In fact, few Asian sequences were available, and most African strains derived from Guinea Bissau, with a single strain from Nigeria. Notably, the Nigerian strain, despite having been sampled 10 years later, was very similar to the Guinea Bissau sequences, suggesting a limited diversity of VZV in sub-Saharan Africa. This view is also supported by single-nucleotide polymorphism typing studies indicating that clade 5 viruses dominate transmissions in other African countries [47]. Nonetheless, these observations are still based on a limited number of strains and may not represent the real situation in the continent.

We addressed the problem of skewed strain origin both by resampling with omission of character states for European sequences and by the analysis of a smaller genome region that was sequenced in Asian VZV strains. Although results confirmed a likely European origin, these strategies cannot compensate for the lack of sequence information from several geographic areas. Thus, data herein should not be regarded as a definite indication that VZV originated in Europe but as a reassessment of the hypothesis that the geographic segregation of VZV clades is consequent to early human out-of-Africa migration patterns [6, 19, 20]. In particular, our analyses do not imply that VZV did not exist in Africa at the time of human dispersal but rather suggest that the representation and spatial distribution of circulating strains are due to much more recent events. This conclusion is in agreement with the observation that data from randomly collected viruses provide an accurate picture of VZV strain prevalence across the world and, therefore, that VZV clade diversity is most likely higher in Europe (excluding recent introductions) than in Africa [47].

In the absence of viral sequences recovered from ancient samples, as is the case for VZV, analysis of viral evolution is inevitably limited to extant sequences, meaning to lineages that have survived. Sequencing of other DNA viruses from ancient samples clearly indicated that present-day viral genetic diversity does not necessarily correspond to the diversity present in the past. In fact, viral lineage extinction was documented for hepatitis V virus [48, 49], parvovirus B19 [50], and variola virus [51]. Thus, owing to either selective or stochastic processes, VZV strains that infected early humans may have gone extinct and been replaced by those we observe today. This may be a common fate for viral lineages, especially for viruses capable

of horizontal transmission and rapid, epidemic spread [52]. Similar conclusions were reached by Weinert and coworkers [13], who incorporated viral latency in an evolutionary model to date the MRCA of extant VZV strains. They obtained a time estimate that placed the MRCA at most 8000 years ago, clearly inconsistent with the out-of-Africa origin. They noted, however, that this estimates refers to extant VZV strains, not excluding the possibility that VZV existed in much earlier times in sub-Saharan Africa.

Data herein also have bearing on the debate concerning the evolutionary origin of major VZV clades [17, 19, 43, 44]. STRUCTURE analysis indicated that the majority of clade 1, 2, 3, and 5 strains derived most of their genetic ancestry from 1 of the 4 ancestral populations. Of course, as noted elsewhere in the case of recombination patterns [17], this conclusion may change if additional VZV lineages are sequenced and found to have contributed to the ancestry of these clades. Conversely, the observation that clades 4, 6, 9, and VIII have mosaic genomes deriving from different ancestral populations is not affected by this caveat (at most, the number of contributing populations may increase). Thus, STRUCTURE results suggest clades 4 and 6/9 originated from independent ancestral recombination events, because viral genomes in the 2 clades have different contributions from distinct ancestral populations. Clade 9 and clade 6 viruses, as well as the single-clade VIII genome have very similar ancestry components, which may derive from sequential recombination events. Because the 2008 VZV nomenclature meeting proposed to define clades as evolutionarily distinct viral subspecies [53], reconsideration of current nomenclature might be warranted.

Finally, we note that low-level admixture was also observed for clade 5 genomes, but only for viruses sampled outside Africa. Phylogeographic analysis placed the origin of clade 5 in Europe, suggesting that VZV was introduced from this continent into Africa, where most (if not all) transmitted viruses belong to clade 5 [47]. Thus, geographic isolation might be responsible for the limited opportunity of admixture of clade 5 in Africa. The observed admixture pattern also suggests that the clade 5 VZV strains included in this analysis do not represent very recent introductions from Africa due to migratory fluxes.

## SUPPLEMENTARY DATA

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

## Notes

**Financial support.** This work was supported by the Italian Ministry of Health (grant RC 2016–2018 to M. S.).

**Potential conflicts of interest.** All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

- Kennedy PGE, Gershon AA. Clinical features of varicella-zoster virus infection. *Viruses* **2018**; 10. doi:10.3390/v10110609.
- Gershon AA, Breuer J, Cohen JL, et al. Varicella zoster virus infection. *Nat Rev Dis Primers* **2015**; 1:15016.
- Galetta KM, Gilden D. Zeroing in on zoster: a tale of many disorders produced by one virus. *J Neurol Sci* **2015**; 358:38–45.
- World Health Organization. Immunization, vaccines and biologicals: varicella. <https://www.who.int/immunization/diseases/varicella/>. Accessed 4 April 2019.
- Nichols RA, Averbeck KT, Poulsen AG, et al. Household size is critical to varicella-zoster virus transmission in the tropics despite lower viral infectivity. *Epidemics* **2011**; 3:12–8.
- Grose C. Pangaea and the out-of-Africa model of varicella-zoster virus evolution and phylogeography. *J Virol* **2012**; 86:9558–65.
- Zell R, Taudien S, Pfaff F, Wutzler P, Platzer M, Sauerbrei A. Sequencing of 21 varicella-zoster virus genomes reveals two novel genotypes and evidence of recombination. *J Virol* **2012**; 86:1608–22.
- Allen WP, Felsenfeld AD, Wolf RH, Smetana HF. Recent studies on the isolation and characterization of Delta herpesvirus. *Lab Anim Sci* **1974**; 24:222–8.
- Gray WL, Pumphrey CY, Ruyechan WT, Fletcher TM. The simian varicella virus and varicella zoster virus genomes are similar in size and structure. *Virology* **1992**; 186:562–72.
- Gray WL, Starnes B, White MW, Mahalingam R. The DNA sequence of the simian varicella virus genome. *Virology* **2001**; 284:123–30.
- Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. *Nature* **2017**; 541:302–10.
- Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol* **2010**; 27:2038–51.
- Weinert LA, Depledge DP, Kundu S, et al. Rates of vaccine evolution show strong effects of latency: implications for varicella zoster virus epidemiology. *Mol Biol Evol* **2015**; 32:1020–8.
- Depledge DP, Gray ER, Kundu S, et al. Evolution of cocirculating varicella-zoster virus genotypes during a chickenpox outbreak in Guinea-Bissau. *J Virol* **2014**; 88:13936–46.
- Muir WB, Nichols R, Breuer J. Phylogenetic analysis of varicella-zoster virus: evidence of intercontinental spread of genotypes and recombination. *J Virol* **2002**; 76:1971–9.
- Jensen NJ, Rivallier P, Tseng HF, et al. Revisiting the genotyping scheme for varicella-zoster viruses based on whole-genome comparisons. *J Gen Virol* **2017**; 98:1434–8.
- Norberg P, Depledge DP, Kundu S, et al. Recombination of globally circulating varicella-zoster virus. *J Virol* **2015**; 89:7133–46.
- Schmidt-Chanasit J, Sauerbrei A. Evolution and world-wide distribution of varicella-zoster virus clades. *Infect Genet Evol* **2011**; 11:1–10.
- Peters GA, Tyler SD, Grose C, et al. 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* **2006**; 80:9850–60.
- Wagenaar TR, Chow VT, Buranathai C, Thawatsupha P, Grose C. The out of Africa model of varicella-zoster virus evolution: single nucleotide polymorphisms and private alleles distinguish Asian clades from European/North American clades. *Vaccine* **2003**; 21:1072–81.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**; 30:772–80.
- Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Linkage analysis. Bioinformatics* **2000**; 16:847–8.
- Maydt J, Lengauer T. Recco: recombination analysis using cost optimization. *Bioinformatics* **2006**; 22:1064–71.
- Thézé J, Lowes S, Parker J, Pybus OG. Evolutionary and phylogenetic analysis of the hepaciviruses and pegiviruses. *Genome Biol Evol* **2015**; 7:2996–3008.
- Martin DP, Murrell B, Khoosal A, Muhire B. Detecting and analyzing genetic recombination using RDP4. *Methods Mol Biol* **2017**; 1525:433–60.
- Sawyer S. Statistical tests for detecting gene conversion. *Mol Biol Evol* **1989**; 6:526–38.
- Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **2001**; 98:13757–62.
- Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol* **1992**; 34:126–9.
- Statistics Division, Department of Economic and Social Affairs, United Nations. Methodology: standard country or area codes for statistical use (M49). <https://unstats.un.org/unsd/methodology/m49/>. Accessed 4 April 2019.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* **2009**; 5:e1000520.



31. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **2014**; 10:e1003537.
32. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **2012**; 9:772.
33. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **2003**; 52:696–704.
34. Yu Y, Harris AJ, Blair C, He X. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol* **2015**; 87:46–9.
35. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**; 19:1572–4.
36. Yu Y, Harris AJ, He X. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Mol Phylogenet Evol* **2010**; 56:848–50.
37. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **1981**; 17:368–76.
38. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **2006**; 23:254–67.
39. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* **2000**; 155:945–59.
40. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **2003**; 164:1567–87.
41. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **2005**; 14:2611–20.
42. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **2015**; 15:1179–91.
43. Norberg P, Liljeqvist JA, Bergström T, Sammons S, Schmid DS, Loparev VN. Complete-genome phylogenetic approach to varicella-zoster virus evolution: genetic divergence and evidence for recombination. *J Virol* **2006**; 80:9569–76.
44. McGeoch DJ, Gatherer D. Integrating reptilian herpesviruses into the family herpesviridae. *J Virol* **2005**; 79:725–31.
45. Davison AJ. Evolution of sexually transmitted and sexually transmissible human herpesviruses. *Ann N Y Acad Sci* **2011**; 1230:E37–49.
46. McGeoch DJ, Rixon FJ, Davison AJ. Topics in herpesvirus genomics and evolution. *Virus Res* **2006**; 117:90–104.
47. Breuer J. Molecular genetic insights into varicella zoster virus (VZV), the vOka vaccine strain, and the pathogenesis of latency and reactivation. *J Infect Dis* **2018**; 218:75–80.
48. Krause-Kyora B, Susat J, Key FM, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* **2018**; 7. doi:10.7554/eLife.36666.
49. Mühlemann B, Jones TC, Damgaard PB, et al. Ancient hepatitis B viruses from the bronze age to the medieval period. *Nature* **2018**; 557:418–23.
50. Mühlemann B, Margaryan A, Damgaard PB, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc Natl Acad Sci U S A* **2018**; 115:7557–62.
51. Duggan AT, Perdomo MF, Piombino-Mascali D, et al. 17th Century variola virus reveals the recent history of smallpox. *Curr Biol* **2016**; 26:3407–12.
52. Wertheim JO. Viral evolution: mummy virus challenges presumed history of smallpox. *Curr Biol* **2017**; 27:R119–20.
53. Breuer J, Grose C, Norberg P, Tipples G, Schmid DS. A proposal for a common nomenclature for viral clades that form the species varicella-zoster virus: summary of VZV Nomenclature Meeting 2008, Barts and the London School of Medicine and Dentistry, 24–25 July 2008. *J Gen Virol* **2010**; 91:821–8.