You Will Never Walk Alone: Codispersal of JC Polyomavirus with Human Populations

Diego Forni,*[,1] Rachele Cagliani,[1] Mario Clerici,[2,3] Uberto Pozzoli,[1] and Manuela Sironi [1]

[1]Scientific Institute, IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Lecco, Italy

[2]Department of Physiopathology and Transplantation, University of Milan, Milan, Italy

[3]IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

*Corresponding author: E-mail: diego.forni@lanostrafamiglia.it.

Associate editor: Meredith Yeager

Abstract

JC polyomavirus (JCPyV) is one of the most prevalent human viruses. Findings based on the geographic distribution of viral subtypes suggested that JCPyV codiverged with human populations. This view was however challenged by data reporting a much more recent origin and expansion of JCPyV. We collected information on ~1,100 worldwide strains and we show that their geographic distribution roughly corresponds to major human migratory routes. Bayesian phylogeographic analysis inferred a Subsaharan origin for JCPyV, although with low posterior probability. High confidence inference at internal nodes provided strong support for a long-standing association between the virus and human populations. In line with these data, pairwise F_{ST} values for JCPyV and human mtDNA sampled from the same areas showed a positive and significant correlation. Likewise, very strong relationships were found when node ages in the JCPyV phylogeny were correlated with human population genetic distances (nuclear-marker based F_{ST}). Reconciliation analysis detected a significant cophylogenetic signal for the human population and JCPyV trees. Notably, JCPyV also traced some relatively recent migration events such as the expansion of people from the Philippines/Taiwan area into Remote Oceania, the gene flow between North-Eastern Siberian and Ainus, and the Koryak contribution to Circum-Arctic Americans. Finally, different molecular dating approaches dated the origin of JCPyV in a time frame that precedes human out-of-Africa migration. Thus, JCPyV infected early human populations and accompanied our species during worldwide dispersal. JCPyV typing can provide reliable geographic information and the virus most likely adapted to the genetic background of human populations.

Key words: **JC polyomavirus, codispersal, human populations, phylogeography, molecular dating.**

Article

Introduction

JC polyomavirus (JCPyV) is one of the most common human viruses, with worldwide seroprevalence estimates of 44–90% (Moens et al. 2017). The virus is usually acquired early in life and persists asymptomatically in the urinary tract and lymphoid organs. About one third of healthy adults persistently shed JCPyV in the urine (Hirsch et al. 2013). In immunocompromised subjects such as patients with acquired immunodeficiency syndrome, patients receiving therapeutic monoclonal antibodies, and transplant recipients, JCPyV infection can develop into progressive multifocal leukoencephalopathy (PML), a lethal demyelinating disease of the central nervous system.

Like other members of the *Polyomaviridae* family, JCPyV is a nonenveloped virus with a small (~5 kb) dsDNA genome. Although JCPyV is specific to humans, a large number of polyomaviruses have been detected in mammals, birds, fish, and arthropods (Buck et al. 2016). Evolutionary analyses indicated that host–virus associations have been relatively stable over long time frames and that polyomaviruses have often codiverged with their hosts (Buck et al. 2016; Madinda et al. 2016).

Investigations of the recent evolutionary history of JCPyV provided contrasting results. The identification of distinct JCPyV subtypes in different geographic areas led to the suggestion that the virus infected our species before the out-of-Africa expansion and codispersed with human populations (Agostini et al. 1997; Sugimoto et al. 1997; Ryschkewitsch et al. 2000; Stoner et al. 2000; Sugimoto, Hasegawa, Kato, et al. 2002; Pavesi 2003; Hirsch et al. 2013). Thus, JCPyV was proposed as a molecular marker of human migrations (Agostini et al. 1997; Sugimoto et al. 1997; Pavesi 2003). Consistently, when the assumption of codivergence was incorporated in phylogenetic analyses, long-term estimates of polyomavirus evolutionary rates in the range of 10^{-7}–10^{-8} substitutions per site per year were obtained (Krumbholz et al. 2009; Buck et al. 2016; Madinda et al. 2016). However, data based on different approaches challenged this view and estimated orders of magnitude faster evolutionary rates and, consequently, a much more recent origin of JCPyV (Shackelton et al. 2006; Kitchen et al. 2008). These findings and the lack of significant congruence between the virus tree and a simplified human population phylogeny led to the conclusion that JCPyV originated in relatively recent times and its population size increased in the last 350 years

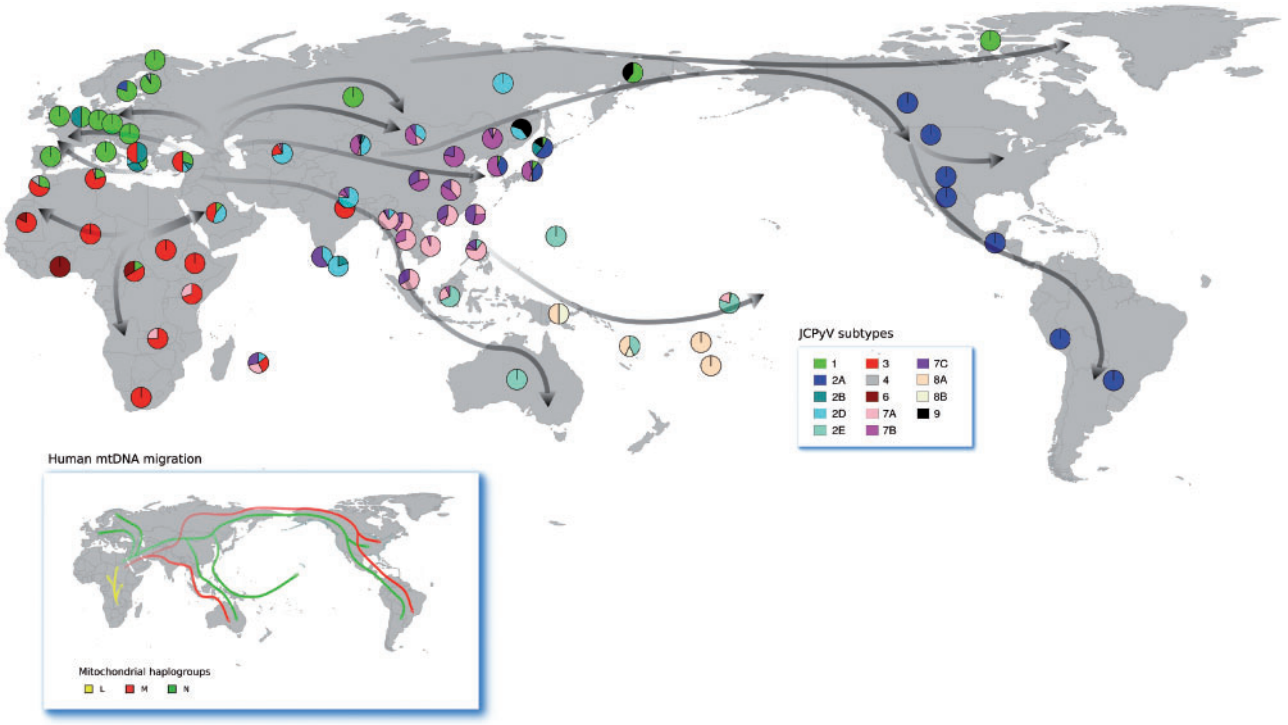

FIG. 1. Worldwide distribution of JCPyV. Worldwide distribution of known and inferred JCPyV subtypes using 431 full genomes and 701 sequences spanning the VT intergenic region. Subtype colors are indicated in the legend and gray arrows represent major known human migration routes according to Nielsen et al. (2017). The insert indicated the worldwide dispersal of mtDNA haplogroups (Stewart and Chinnery 2015).

(Shackelton et al. 2006). This scenario is nonetheless difficult to reconcile with some epidemiological findings (Sharp and Simmonds 2011).

We used multiple and complementary analyses and data sets to assess the evolutionary history of JCPyV and to test the hypothesis that it codiverged and codispersed with human populations. Distinct molecular evolution and phylogenetic approaches yielded consistent results, indicating that JCPyV originated before the out-of-Africa migration and spread worldwide along with its human hosts.

Results

Geographic Distribution of JCPyV Subtypes

To assess JCPyV geographic distribution, we analyzed a total of 1,132 isolates with known geographic origin by complementing 431 JCPyV complete genomes with 701 partial sequences spanning the VT intergenic region (supplementary table S1, Supplementary Material online). These JCPyV sequences could be unambiguously assigned to a known subtype using a classification evolutionary algorithm (see Materials and Methods). Nine strains clustered together but could not be assigned to a known subtype; they are thus provisionally referred to as subtype 9. Mapping of JCPyV subtypes onto a world map confirmed the preferential association of distinct subtypes with specific geographic regions, although it was also clear that multiple subtypes are transmitted in several areas (fig. 1).

Superimposition of established major human migratory routes (Nielsen et al. 2017) revealed a good correspondence with the distribution of JCPyV subtypes (fig. 1).

A similarity was also evident with the dispersal trajectories of mitochondrial DNA (mtDNA) haplogroups (Stewart and Chinnery 2015), whereby two major mtDNA lineages, M and N, emerged from Africa (where lineages L0–L3 are present). In the Old World, lineage N spread to Eurasia and, in Europe, originated haplogroup H, which is now prevalent in this continent and also represented in Northern Africa (resembling the distribution of subtypes 1 and 4; Badro et al. 2013). Lineage M remained mainly confined to Asia and Near Oceania (in analogy with the distribution of subtypes 2a–d, 7a–c, and 8a, b; fig. 1).

However, these observations may simply represent incidental resemblances and different possible scenarios may account for the observed distribution of JCPyV.

Long-Standing Association between JCPyV and Human Populations

We thus set out to obtain information on the phylogenetic and geographic relationships of JCPyV sequences. For this aim, we used a set of 431 complete genomes with known geographic origin (supplementary table S1, Supplementary Material online). The genome alignment was screened for the presence of recombination using GARD (Genetic Algorithm Recombination Detection; Kosakovsky Pond et al. 2006) and four methods implemented in the RDP4 software (Martin et al. 2017). In particular, RDP, GENECONV, MaxChi, and Chimera (Sawyer 1989; Smith 1992; Martin and Rybicki 2000; Posada and Crandall 2001; Martin et al. 2017) were used because they showed good power in previous simulation analyses

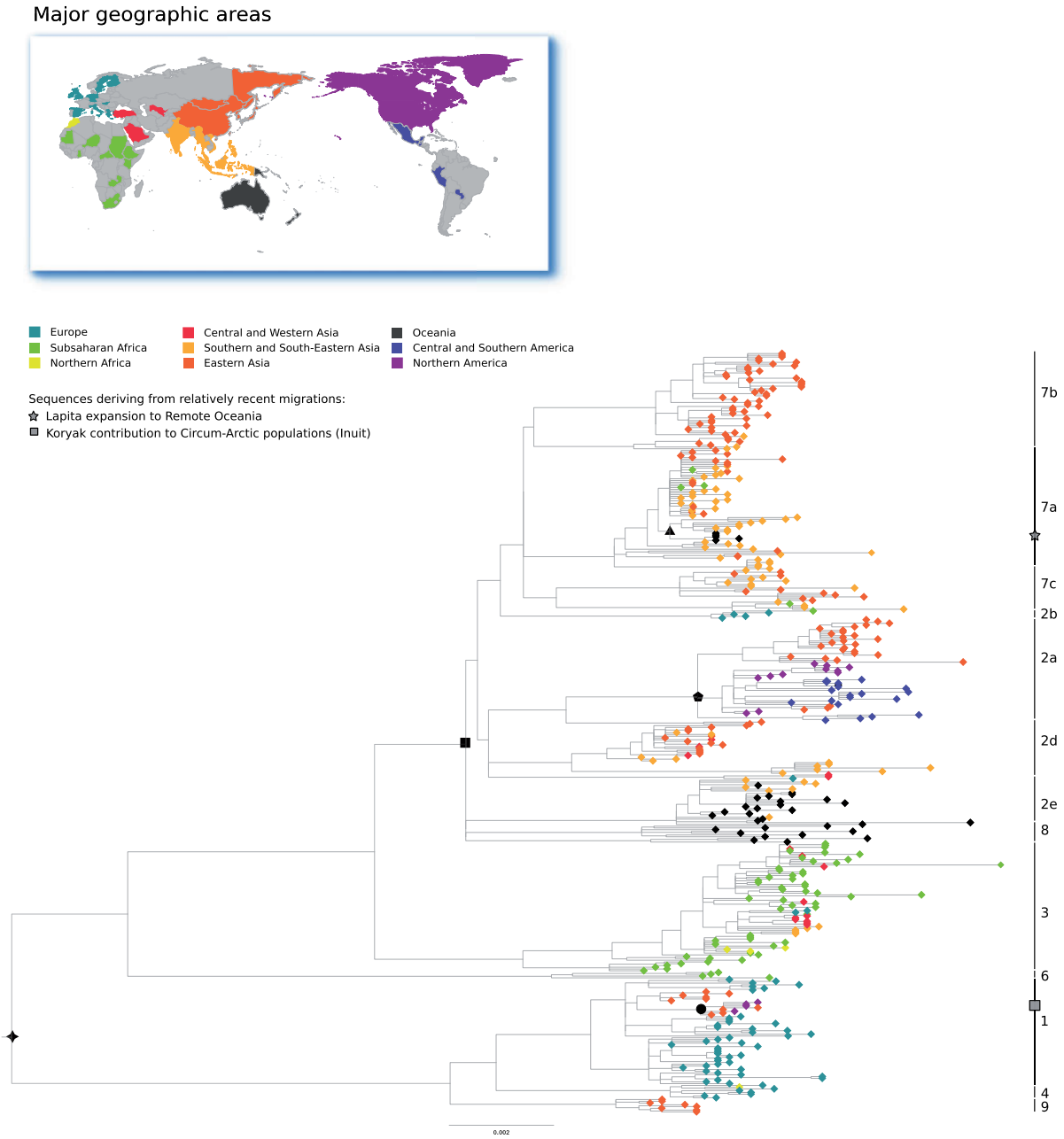

FIG. 2. JCPyV phylogeny. Maximum likelihood tree for JCPyV genomes. The tree was constructed with PhyML using PtrovPyV8 (not shown) as an outgroup. All bootstrap values were higher than 90. Tips are colored according to their geographic origin (see legend). Nodes used in the correlation analysis in figure 3 are marked with symbols. Clades are labeled based on JCPyV subtypes. The insert shows the geographic assignments of JCPyV genomes.

(Posada and Crandall 2001; Bay and Bielawski 2011). The cut-off *P* value was set to 0.01 and no evidence of recombination was detected using any method. Also, no evidence of substitution saturation was detected by Xia's index (Xia et al. 2003; Xia 2013).

JCPyV sequences were assigned to nine major areas (fig. 2) and a maximum likelihood phylogenetic tree was obtained. In line with previous reconstructions (Shackelton et al. 2006), the African lineages were not basal in the JCPyV phylogeny. Whereas this is not consistent with the codispersal hypothesis, an ancestral African lineage may have gone extinct or remain unsampled. An alternative possibility is that the basal lineage only found in Europeans and Asians was transmitted to modern humans by extinct hominin populations (e.g., Neanderthals or Denisovans; see Discussion).

Phylogenetic analysis provided several indications of a long-standing association between JCPyV and human populations. For instance, all viral strains sequenced from non-Inuit native Americans formed a single clade and clustered with sequences of Eastern Asian origin (fig. 2). This is clearly consistent with the notion that the Americas were initially populated from Beringia (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018). Inuit sequences clustered with those derived from Koryaks (from Siberia), in line with a

more recent gene flow from populations related to Koryaks into Circum-Arctic native American populations (~5.5 kya) (fig. 2; Moreno-Mayar, Vinner, et al. 2018; Flegontov et al. 2019). As for Oceanian sequences, they formed two clades that included or were closely related to strains from Southern and South-Eastern Asia (fig. 2). This pattern parallels the dynamics of the peopling of Oceania, whereby early migrants from South-East Asia reached Australia and Papua-New Guinea to subsequently spread to Remote Oceania (Skoglund et al. 2016). Later migrations from the Taiwan/Philippines area (Lapita culture) brought populations with different ancestry to Remote Oceania (Skoglund et al. 2016; Lipson et al. 2018). Finally, three of the four Northern African sequences branched from Subsaharan African strains. Conversely, sequences from Central and West Asia, which would be expected to represent an early branching event in human migration history, were interspersed in the phylogeny (fig. 2). However, it should be kept in mind that human movements in the recent past have also influenced the spread of pathogens and their present distribution. For instance, human movements along the Silk Road were previously shown to have contributed to the spread of *Yersinia pestis* (Morelli et al. 2010).

We next reasoned that, if JCPyV established long-standing associations with human populations, a correlation should be observed between the genetic diversity of human populations living in different geographic areas and the diversity of JCPyV sampled in those same locations. To test this prediction, we collected a set of 385 complete mtDNA sequences with similar geographic distribution as the JCPyV strains (Northern African strains were removed as the sample size was too small, see Materials and Methods; supplementary table S2, Supplementary Material online). Pairwise genetic differences were calculated as F_{ST} for both JCPyV and mtDNA. A positive and significant correlation was observed between the F_{ST} estimates obtained for the virus and for human mtDNA (fig. 3*A*), strongly supporting the view that JCPyV dispersed worldwide with its human hosts.

To further test the codispersal hypothesis, we used previously calculated F_{ST} values for human populations obtained from nuclear DNA variants (Cavalli-Sforza and Feldman 2003) and correlated them with the corresponding node ages (estimated as substitutions/site) of the JCPyV phylogeny (supplementary table S3, Supplementary Material online). An extremely strong linear correlation was observed, indicating that there is a very good correspondence between the relative genetic divergence of human populations and of JCPyV subtypes (fig. 3*B*).

Codispersal of JCPyV with Human Populations

We thus decided to apply phylogeographic and molecular dating approaches implemented in BEAST to gain further insight into the dispersal pattern of JCPyV. Given the results obtained above, we ran two distinct molecular dating and phylogeographic analyses, using two different external calibration points based on human population history. The first one set the most recent common ancestor (MRCA) of the Northern American/Southern American JCPyV strains as the split time (around 16 kya) of Northern Native American (NNA)/Southern Native American (SNA) populations (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018; fig. 4 and table 1). In the second analysis, we used the first migratory event to Australia and New Guinea (~51 kya) (Nielsen et al. 2017) as a calibration point for the split time of the JCPyV Oceanian strains (table 1).

Results were similar between the two analyses, with a TMRCA (time to the MRCA) of all JCPyV strains of ~135 kya and ~131 kya using the two calibration points (table 1). BEAST also estimated the most likely origin of JCPyV to be in Subsaharan Africa, albeit with relatively low confidence (fig. 4, supplementary fig. S1, Supplementary Material online). These results are consistent with the hypothesis that JCPyV originated before the major out-of-Africa migration event, which occurred 60–65 kya (Nielsen et al. 2017; Molinaro and Pagani 2018).

We then compared the location and timing of the branching of JCPyV strains with known events in human history. Overall, a very good correspondence was found (fig. 4). Thus, non-Inuit native American sequences were inferred with high probability to have an East-Asian origin (fig. 4, supplementary fig. S1, Supplementary Material online). Their split time was estimated to have occurred ~23–26 kya (fig. 4, table 1), in full agreement with the estimated date (i.e., ~24–26 kya) for the first peopling of the Americas from Beringia (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018). The inferred NNA/SNA split using the Oceania peopling calibration was dated at 14 kya (estimated split time of human populations around 16 kya) (table 1). Using both calibrations, the timing of the more recent migration from North-Eastern Siberia to Northern America was also roughly consistent with recent estimates (Moreno-Mayar, Vinner, et al. 2018; Flegontov et al. 2019; table 1 and fig. 4). Finally, in line with known human migration events, both Oceanian clades were inferred with high probability to have a Southern and South-Eastern Asian origin (fig. 4 and supplementary fig. S1, Supplementary Material online) and their split times were in good agreement with human settlement history, although we obtained estimates older than expected for the peopling of Remote Oceania (table 1).

It is now appreciated that the estimated rates of viral evolution scale negatively with the time frame over which they are measured (Duchene et al. 2014; Aiewsakun and Katzourakis 2016). This time-dependent rate phenomenon (TDRP), which is most likely due to a combination of purifying selection and saturation effects, can lead to severe underestimation of the time of origin of viral species and genera (Duchene et al. 2014; Aiewsakun and Katzourakis 2016). We thus wished to assess whether the results obtained above are affected by the TDRP. For this aim, we applied a recently developed Bayesian statistical inference approach that accommodates rate variation through time (epoch model). We performed two analyses by setting different epoch structures (see Materials and Methods), as proposed by Membrebe et al. (2019). Internal calibrations were set at the two nodes corresponding to the human migration events used in the analyses above. In addition, we set the divergence of JCPyV from Pan troglodytes verus polyomavirus

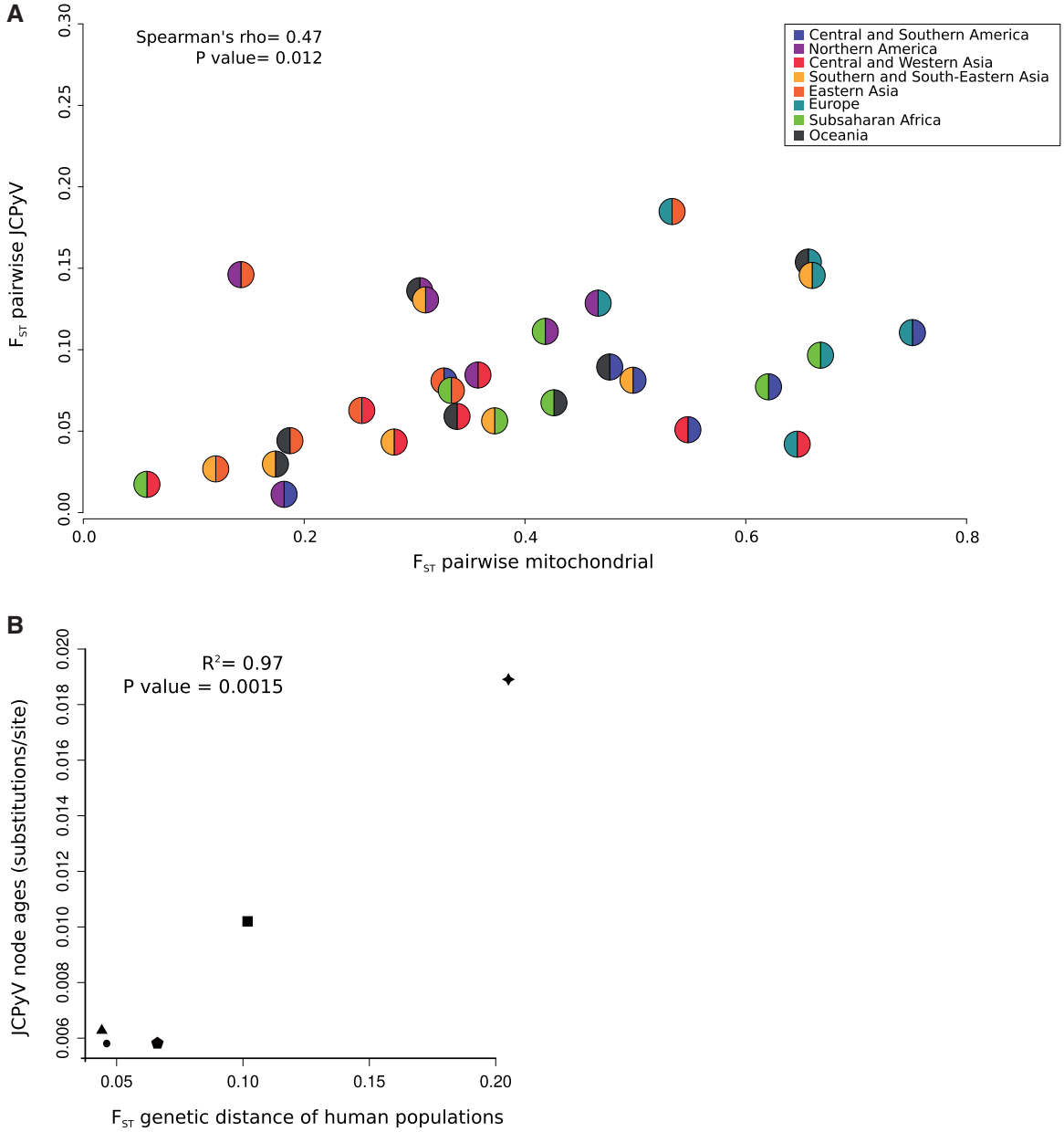

FIG. 3. Human–virus genetic diversity. (*a*) Plot of pairwise F_{ST} values calculated for JCPyV genomes grouped on the basis of their geographic location and pairwise F_{ST} values calculated on human mtDNA sequences sampled in the same locations (see Materials and Methods for details and supplementary table S3, Supplementary Material online). Spearman's correlation rho is also reported, along with the *P* value. (*b*) Correlation between JCPyV node ages (expressed as substitutions/site) and human population genetic distance expressed as F_{ST} (Cavalli-Sforza and Feldman 2003). Symbols denote nodes in figure 2.

8 (PtrovPyV8) as the accepted human/chimpanzee lineage split time (i.e., 6 Ma; Scally et al. 2012; Prado-Martinez et al. 2013). The two models provided similar TMRCA estimates (182–212 kya), which were roughly consistent with the results obtained with the calibration approaches described above (table 1, supplementary figs. S1 and S2, Supplementary Material online). A very good consistency with known human migration events was also obtained for the two internal nodes corresponding to the first peopling of the Americas and the gene flow from Koryak-related to Inuit populations (table 1). Again, the time estimate for the peopling of Remote Oceania was older than expected (table 1). However, it is worth noting that analysis of mtDNA diversity in Taiwan and Oceania also indicated divergence times older than 5 kya for Austronesian haplogroups (Duggan et al. 2014; Ko et al. 2014), most likely resulting from complex early population structure and migratory fluxes across Mainland and Insular Southeastern Asia.

Congruence between JCPyV and Human Population Phylogenies

Finally, to formally test for a congruence between the JCPyV and human population phylogenetic trees, we used CoRe-Pa (Merkle et al. 2010), an event-based method for

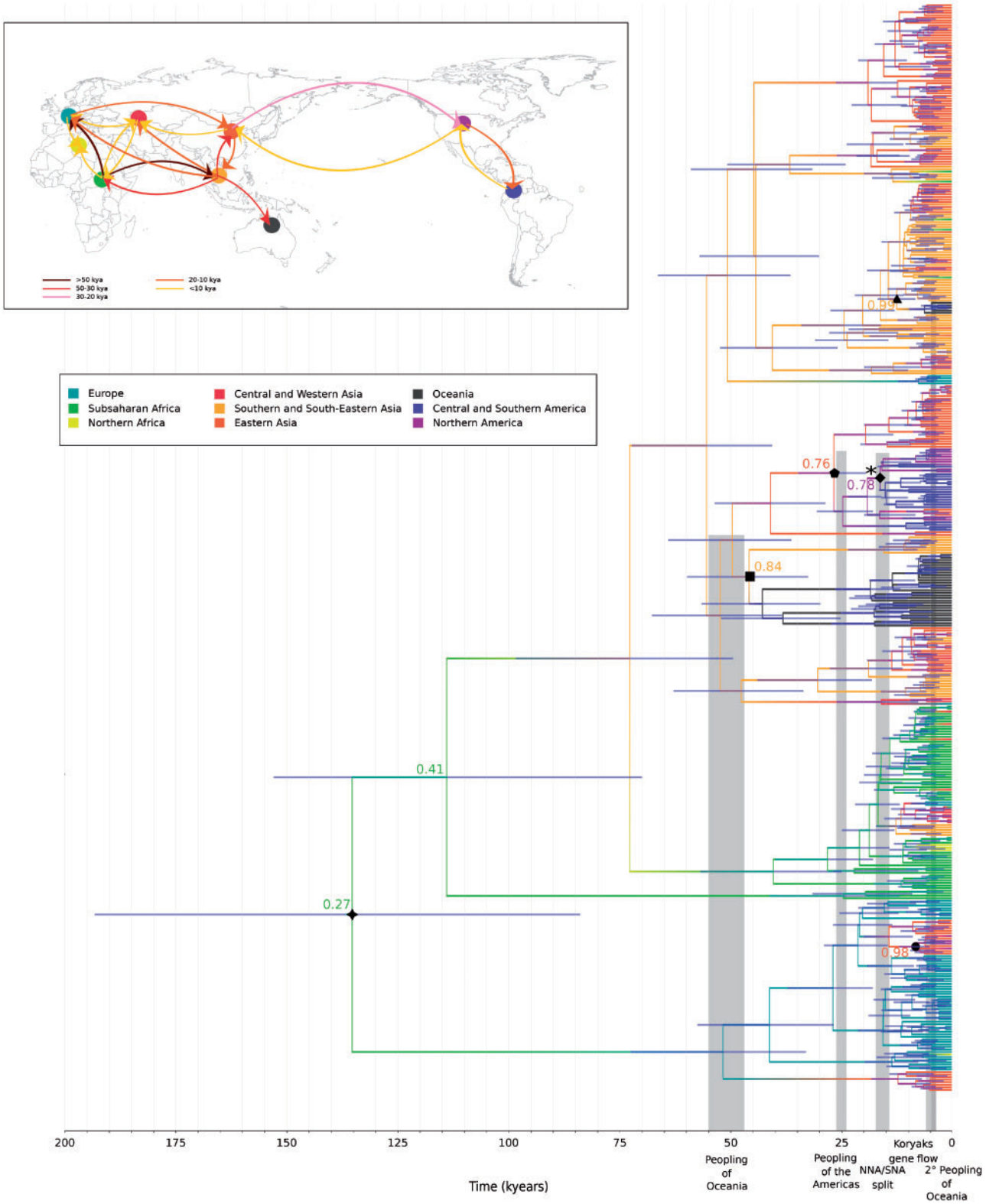

FIG. 4. Timescaled JCPyV phylogenetic tree and SpreaD3 analysis. Dated maximum credibility tree estimated using complete JCPyV genomes. Branch lengths represent the evolutionary time (in thousand of years); a timescale grid is shown at the tree base; node bars indicate 95% HPD intervals of node ages. The branches are colored according to inferred ancestral location (see legend); posterior support for relevant locations (colored numbers) is shown. Relevant nodes are indicated with symbols (as in figs. 2 and 3) and the asterisk indicates the node used as a calibration point. The time-frames of relevant human migrations are also reported. NNA, Northern Native Americans; SNA, Southern Native Americans. The insert shows inferred JCPyV spread patterns as visualized using the SpreaD3 tool. Circle colors denote geographic locations and arrow colors represent inferred spread times (see legend). Note that, because samples were assigned to nine major areas, geographic coordinates are arbitrary and do not correspond to any individual sample.

Table 1. Estimated Dates of JCPyV Phylogenetic Events.

Event	Molecular Dating, Calibration 1	Molecular Dating, Calibration 2	TDR Correction Epoch Structure 1[a]	TDR Correction Epoch Structure 2[a]	Corresponding Human Migration Event
MRCA	135.3 [C.I. 83.8–193.4]	131.0 [C.I. 47.1–207.5]	212 [C.I. 178–246]	182 [C.I. 155–208]	65 (Nielsen et al. 2017; Molinaro and Pagani 2018)
Oceanian strain split	45.7 [C.I. 32.4–59.7]	Calibrated with the human migration event (51, SD: 2)	Calibrated with the human migration event (51, SD: 2)	Calibrated with the human migration event (51, SD: 2)	47–55 (Nielsen et al. 2017)
American strain split	26.5 [C.I. 19.3–34.6]	23.2 [C.I. 15.4–35.5]	23.3 [C.I. 16.5–28]	22 [C.I. 16.6–27.6]	23.9–26 (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018)
MRCA of North American/South American strains	Calibrated with the human migration event (16.1, SD: 0.7)	14.3 [C.I. 5.6–22.7]	Calibrated with the human migration event (16.1, SD: 0.7)	Calibrated with the human migration event (16.1, SD: 0.7)	17.5–14.6 (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018)
Inuit strain split	8.2 [C.I. 4.9–11.8]	7.3 [C.I. 3.1–12.2]	13.4 [C.I. 7–18.2]	12.2 [C.I. 7.1–17.7]	4.9–6.2 (Flegontov et al. 2019)
Remote Oceania strain split	12.4 [C.I. 8.1–16.6]	17.8 [C.I. 7.6–17.4	22 [C.I. 15.6–26.2]	20.2 [C.I. 15.3–25.3]	3.0–5.0 (Lipson et al. 2018)

NOTE.—Time is expressed in thousands of years ago (kya).
[a]See Materials and Methods for details on the two epoch structures.

cophylogenetic analysis. In particular, we exploited previous data deriving from nuclear markers to construct a phylogenetic tree of 44 human populations from which at least one complete JCPyV sequence was available. The viral tree was obtained by pruning the phylogenetic tree to retain sequences ($n = 70$) corresponding to the 44 human populations (see Materials and Methods; supplementary table S1, Supplementary Material online). Because of lack of details on the ethnic origin of African subjects from whom JCPyV sequences were derived, the African clades were merged into two branches (Central and Southern Africa) based on geographic information.

CoRe-Pa applies several random cost schemes to the data and ranks cophylogenetic reconstructions based on their quality values (q_c). The program yielded 85 reconstructions; the three preferred ones had 33–36 codivergence events, a large number of sortings, 20–24 duplications, and no >13 population transitions (table 2). The other solutions had considerably worse quality values ($q_c > 0.015$). Analysis of random JCPyV-human population associations indicated that the probability of obtaining 33 or more cospeciation events is low ($P = 0.044$; table 2), supporting the hypothesis that codivergence contributed to the making of JCPyV genetic diversity. Indeed, a very good correspondence was observed between the two phylogenies, which was even more evident when phylogeographic information was superimposed (fig. 5). Most population transitions involved relatively terminal nodes and some of these reflected known recent migrations. For instance, the 2E strain from Kiribati most likely reflects the above-mentioned migration from the Philippines/Taiwan area into Remote Oceania (Skoglund et al. 2016), and the subtype 1 strain sampled from Ainus may have been derived from the contact between North-Eastern Siberian and Ainu peoples (Jeong et al. 2016). Clearly, population transitions at the tips of the phylogeny may also reflect very recent migration/admixture events.

Discussion

We provide several lines of evidence indicating that JCPyV has been infecting humans for a long time and accompanied our species during worldwide dispersal. This idea was first proposed in 1997, on the basis of the distribution of viral subtypes in different geographic areas (Agostini et al. 1997; Sugimoto et al. 1997). Several subsequent studies analyzed JCPyV genomes sampled in various locations, including strains hosted by relatively isolated populations (Agostini et al. 2001; Jobe et al. 2001; Sugimoto, Hasegawa, Zheng, et al. 2002; Zheng et al. 2003; Ikegaya et al. 2005; Takasaka et al. 2006). Epidemiologically, the codispersal hypothesis provided convincing explanations for several findings such as the presence of East-Asian like subtypes in Native Americans and Oceanians.

Later works however dismissed this possibility and indicated that JCPyV evolves much more rapidly than previously estimated and, thus, that it originated in recent times (Shackelton et al. 2006; Kitchen et al. 2008). Nonetheless, the molecular dating approach used for those inferences did not account for the TDRP, which is now recognized as

Table 2. Results of Cophylogenetic Analyses.

Reconstruction (q_c)	Total Cost	Cospeciations (Cost)	Sortings (Cost)	Duplications (Cost)	Population Transitions (Cost)	*P* Value for Cospeciation Events
1 (0.0042)	24.88	33 (0.192)	141 (0.044)	24 (0.260)	12 (0.502)	0.044
2 (0.0080)	24.27	34 (0.185)	148 (0.041)	24 (0.255)	11 (0.517)	0.016
3 (0.0090)	24.33	36 (0.165)	144 (0.042)	20 (0.291)	13 (0.501)	<0.001

a major factor responsible for the underestimation of the age of viral lineages (Aiewsakun and Katzourakis 2016). Also, these studies largely relied on tip calibration in a time when testing for the existence of a temporal signal was not common practice.

An important argument raised against the codispersal hypothesis was the observation that African JCPyV strains do not fall basal in the viral phylogeny. This is reflected in the analyses reported herein and clearly represents a notable difference from the evolutionary pattern of human populations. At least three possibilities may account for this discrepancy. First, one or more ancestral African lineages may have gone extinct. Viral lineage extinction events have previously been reported for other DNA viruses, including HBV (Krause-Kyora et al. 2018; Muhlemann, Margaryan, et al. 2018) and parvovirus B19 (Toppinen et al. 2015; Pyoria et al. 2017; Muhlemann, Jones, et al. 2018), both of them long-standing human pathogens. Second, an extant basal African clade may have remained unsampled. This is not unlikely, as sequences derived from Subsaharan Africa only account for ~14% of those available for analyses. Third, the European/Asian basal lineage may have been derived from extinct hominins who populated Eurasia and with whom modern humans interbred (e.g., Neanderthals and/or Denisovans; Nielsen et al. 2017). Indeed, such a host-switch event was previously proposed to have introduced the human papillomavirus 16A lineage into modern human populations (Pimenoff et al. 2017). This scenario seems however unlikely for JCPyV, as the Neanderthal/Denisovan lineage separated from that of modern humans around 400–600 kya (Nielsen et al. 2017), a timeframe which is not consistent with our dating of the JCPyV TMRCA (low-bound estimate: 246 kya).

Beyond the discrepant placing of the African lineages, the human population and viral phylogenies are not fully congruent but, as noted elsewhere, they are not expected to be (Sharp and Simmonds 2011). In fact, as the phylogenies of single genetic markers do not necessarily recapitulate the evolutionary history of a species, the same occurs with an infecting virus. Moreover, whereas genetic variants are always vertically transmitted, this is not necessarily true for viruses, although in the case of JCPyV vertical transmission is probably common (Boldorini et al. 2011). Finally, human migration patterns were complex and, in several regions, occurred through multiple waves. Importantly, human movements after the initial dispersal have been constant and have contributed to the spread of infecting species. Thus, the effect of recent migrations is expected to be superimposed on the signal determined by codispersal. Despite these observations, the reconciliation analysis we conducted showed a significantly high number of cospeciation events, strongly supporting the codispersal model for JCPyV evolution. Indeed, reconciliation and phylogeographic analyses also revealed patterns that can be ascribed to relatively recent migratory events such as the Koryak contribution to Circum-Arctic American populations, the peopling of Remote Oceania by the Lapita culture from the Philippines/Taiwan, and the gene flow between Ainus and Northern Siberians (Jeong et al. 2016; Skoglund et al. 2016; Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018).

In this sense, deep knowledge of human evolutionary history allowed the reconstruction of JCPyV migration history rather than the other way round. Whatever the perspective, JCPyV can provide reliable estimates on the geographic origin of a sample. Thus, given its high prevalence in all human populations, typing of JCPyV subtypes may provide relevant information for trace samples such as forensic specimens or ancient biological remains.

The lifelong persistence, low pathogenicity and high transmissibility of JCPyV are most likely key factors that allowed maintenance of the virus in early small human communities and the subsequent worldwide spread. Because of its long-standing association with our species, the possibility clearly exists that JCPyV has adapted to the genetic background of different human populations. Recent data indicated that, in Europe, JCPyV seropositivity is negatively associated with the HLA-DRB1*15 allele, which is, in turn, deferentially distributed worldwide (Sundqvist et al. 2014; Santos et al. 2016). Thus, human immunogenetic features may have contributed to the differentiation of the JCPyV subtypes. This possibility deserves investigation, especially in light of the increasing incidence of PML caused by the AIDS epidemic and, more recently, by the development of novel immunosuppressive drugs.

Materials and Methods

Sequences, Alignments, and Phylogenetic Reconstruction

We retrieved all available full-length JCPyV genomes from the NCBI database (http://www.ncbi.nlm.nih.gov/, last accessed February 15, 2019). We removed sequences representing artificial clones, highly passaged strains, viruses derived from subjects of mixed ancestry (e.g., Hispanic) or from migrants (e.g., Japanese living in the USA), as well as sequences from nonnative Americans (as they are likely to have reached America in relatively recent times due to the European expansion). Viruses derived from PML patients were also excluded, as they are known to frequently harbor disease-

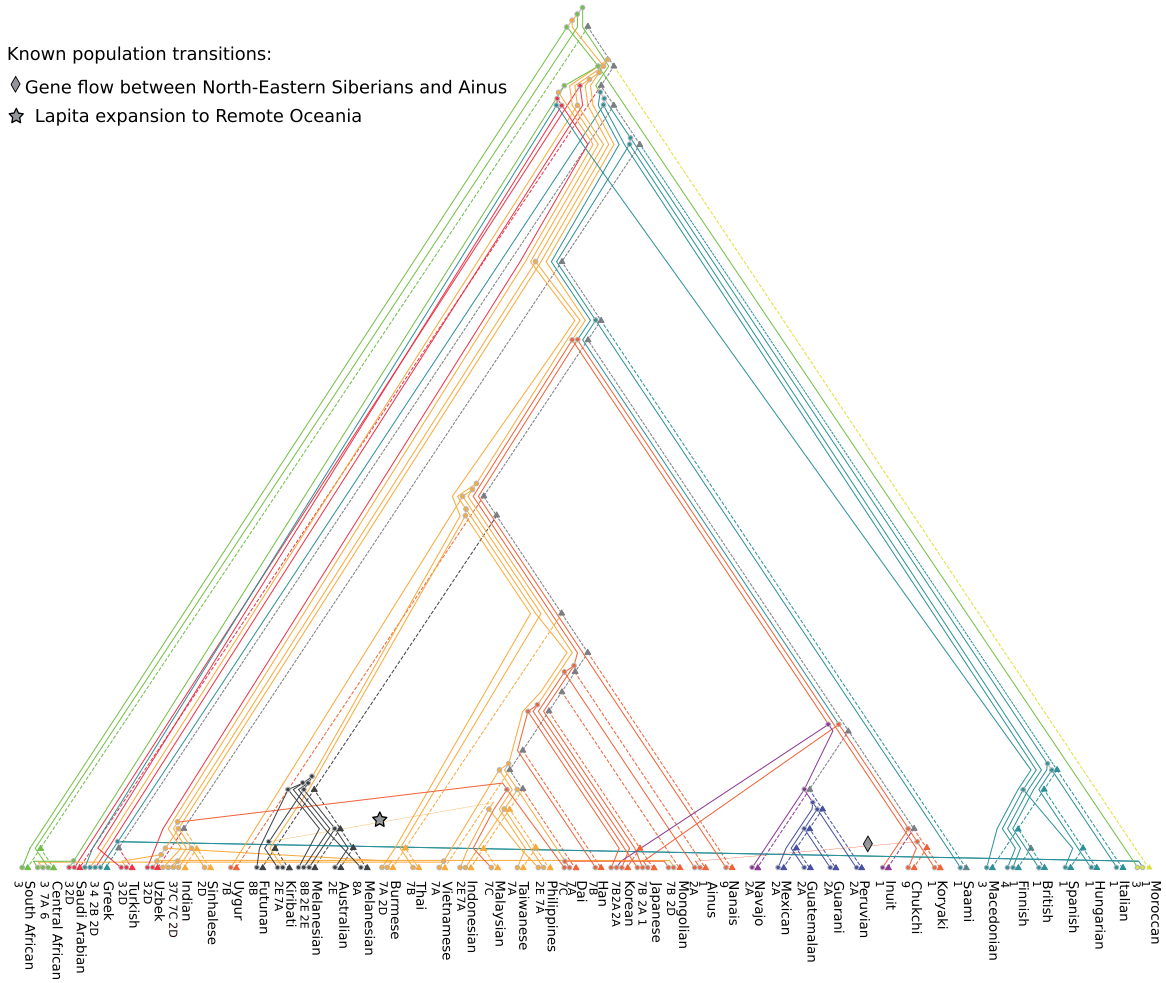

FIG. 5. Cophylogenetic analysis of JCPyV and human population phylogenies. The best CoRe-Pa reconstruction is shown. Nodes in the JCPyV and human populations trees are indicated with dots and triangles, respectively. The JCPyV phylogeny is represented with solid lines, whereas dashed lines denote the human population phylogeny. The two trees are colored according to major geographic areas. In particular, node and branch colors in the JCPyV phylogeny correspond to phylogeographic inference (fig. 4). In the human phylogeny, colors are used to represent populations from the same geographic areas; gray is used for internal nodes and branches when geographic inference cannot be established.

associated rearrangements and point mutations (Sunyaev et al. 2009; Hirsch et al. 2013).

Thus, for phylogenetic reconstruction and BEAST analyses, 431 full-length sequences with unequivocal geographic origin were used (supplementary table S1, Supplementary Material online).

JCPyV strains sequenced for the VT intergenic region, which consists of a genomic region between VP1 and the large T antigen, were also retrieved from NCBI with the same criteria applied for the full-length JCPyV genomes (supplementary table S1, Supplementary Material online). These sequences were used to study the geographic distribution of JCPyV subtypes. In all other analyses only complete genomes were included.

We used MAFFT (Katoh and Standley 2013) to generate multiple sequence alignments. The alignment of the VT intergenic region was 615 bp in length. The full-genome alignment was 5,775 bp long. For this latter alignment, a HKY substitution model with gamma distribution of rate variation (HKY+G) was selected using JmodelTest 2 (Guindon and Gascuel 2003; Darriba et al. 2012).

Phylogenetic reconstruction was performed with PhyML 3.1 (Guindon et al. 2009) using the HKY+G model and the PtrovPyV8 sequence (NCBI Accession Number: NC_028635) as an outgroup. 100 bootstrap replicates were done.

Strains were assigned into nine macro-areas based on the United Nation geographical subregions (http://unstats.un.org/unsd/methods/m49/m49regin.htm, last accessed February 15, 2019), namely Central and Southern America, Central and Western Asia, Eastern Asia, Europe, Northern Africa, Northern America, Oceania, Southern and South-Eastern Asia, Subsaharan Africa.

Subtype Inference

Subtypes were assigned using the Subtype Classification Using Evolutionary ALgorithms (SCUEAL) procedure from the HYPHY package (Pond et al. 2005; Kosakovsky Pond et al. 2009). This method uses a maximum likelihood multi-model inference approach to assign each query sequence to a specific genotype based on a set of reference sequences with known genotypes. We generated two different alignments of the VT intragenic region: The first one comprised all strains

from the whole genome data set with information of the subtype (reference alignment), and the other one comprised all the other strains (whole genome and VT intergenic region) with unknown subtype (query sequences). We then ran SCUEAL and we considered correctly assigned only query sequences with >0.50 support (posterior probability).

F_{ST} Calculation

Pairwise genetic differences in terms of F_{ST} values were calculated using the R package PopGenome (Pfeifer et al. 2014). In particular the 431 JCPyV genomes were grouped based on the same geographic areas as in the phylogenetic analysis (supplementary table S1, Supplementary Material online), considering only groups with at least five sequences (i.e., Northern African strains were excluded from the analysis). Pairwise F_{ST} values were then calculated for the remaining eight groups.

For each geographic location we selected the same number of mtDNA sequences as the JCPyV strains sampled in that location or, where the JCPyV strains outnumbered the mtDNA sequences deposited in GenBank, all the mitochondrial sequences for the region were included. A set of 385 mtDNA sequences were selected without any consideration about mitochondrial haplogroup (supplementary table S2, Supplementary Material online). Pairwise F_{ST} values were calculated with the same approach used for the viral strains. Correlation was estimated using Spearman's Rank Test.

Phylogeography and Molecular Dating

To reconstruct the geographic and temporal origin of JCPyV, inferences were carried out using the discrete model (Lemey et al. 2009) implemented in the Bayesian Evolutionary Analysis by Sampling Trees (BEAST, version 2.5) software (Bouckaert et al. 2014). All complete JCPyV genomes (supplementary table S1, Supplementary Material online) plus PtrovPyV8 (NCBI Accession Number: NC_028635) as an outgroup were used, with the nine macro-areas defined by the United Nation geographical subregions as discrete traits.

A HKY substitution model with gamma distribution of rate variation (HKY + G) was selected using JmodelTest 2 (Guindon and Gascuel 2003; Darriba et al. 2012). The path sampling tool implemented in BEAST was performed to choose between a constant, an exponential, or a coalescent Bayesian skyline tree prior (50 steps, 1,000,000 iterations each, 10% burn-in; supplementary table S4, Supplementary Material online). A Bayes factor test was applied to compare the different likelihoods. Factors between 3 and 20 suggest positive support, 20–150 strong support and >150 overwhelming support. The coalescent Bayesian skyline tree prior was favored (supplementary table S4, Supplementary Material online).

Analyses were performed using a Bayesian Markov Chain Monte Carlo (MCMC) method with a relaxed log normal clock.

Two distinct analyses were run by applying a prior distribution to a different calibration point. The first run was generated by setting a gamma prior distribution to the MRCA between the Northern American/Southern American JCPyV strains using the NNA/SNA split time (median: 16.1 kya, standard deviation: 0.7 kya) (Moreno-Mayar, Vinner, et al. 2018; Moreno-Mayar, Potter, et al. 2018). In the second one, we used the first peopling of Oceania as a calibration date for the divergence of the branch leading to Oceanian strains (median: 51 kya, standard deviation: 2 kya)(Nielsen et al. 2017), again using a prior gamma distribution.

For each of the analyses, we ran two chains for 200 million states with a step of 10,000. After discarding a 10% burn-in, assessing convergence with Tracer (Rambaut et al. 2018), and merging the two runs, the output was analyzed using TreeAnnotator (Bouckaert et al. 2014) and visualized with FigTree (http://tree.bio.ed.ac.uk/, last accessed February 20, 2019).

The phylogenetic reconstructions were analyzed and visualized by the spatial phylogenetic reconstruction of evolutionary dynamics using data-driven documents (SpreaD3) software (Bielejec et al. 2016). SpreaD3 projects the BEAST MCMC tree onto a geographic map. In particular, it combines the geographic coordinates of discrete traits (in term of latitude and longitude) with the ancestral location states and it reconstructs viral dispersal over time using the internal node dating estimated by BEAST. Migration events are represented in figure 4 using colored arrows that correspond to different time-frames.

To account for the TDRP, we applied a recent model proposed by Membrebe et al. (2019) and implemented in BEAST v1.10.4 (Suchard et al. 2018) to the same data set used for the analysis described above.

This model is based on drawing inference from the data using an epoch modeling approach in which the evolutionary rate parameter (r) can vary in each time interval according to the defined epoch structure (supplementary fig. S4, Supplementary Material online; Membrebe et al. 2019).

We performed two analyses by setting two different epoch structures (in both analyses time intervals were exponentially distributed and expressed in units of millions of years) as proposed by Membrebe et al. (2019): The first one has time intervals in the range of $0 < 10^{-5} < 10^{-4} < 10^{-3} < 10^{-2} < 10^{-1} < 10^{0} < 10^{1} < 10^{2} < \infty$ and in the second one the time range was reduced with intervals set at $0 < 5{*}10^{-5} < 5{*}10^{-4} < 5{*}10^{-3} < 5{*}10^{-2} < 5{*}10^{-1} < \infty$.

The TDRP model requires internal node calibrations distributed along the time-scale of the phylogeny. We used the same calibration nodes used in the phylogeographic analysis and we fixed these nodes to the corresponding human migration events (i.e., the first peopling of Oceania, 51 kya, and the split of Northern and Southern Native Americans, 16.1 kya), plus we fixed the divergence of JCPyV and PtrovPyV8 at the accepted human/chimpanzee lineage split time (i.e., 6 Ma, Scally et al. 2012; Prado-Martinez et al. 2013).

We analyzed our data set with a general time-reversible (GTR) substitution model and a Yule speciation (Gernhard 2008) prior on the tree, as suggested by Membrebe et al. (2019). Analyses were run with BEAST v1.10.4 (Suchard et al. 2018) with two chains of 100 million states each (10% burn-in), checked for convergence with Tracer (Rambaut et al. 2018), and the output was analyzed using TreeAnnotator (Bouckaert et al. 2014).

Cophylogeny

For human populations, a tree of 142 populations was obtained from a previous work (Mallick et al. 2016). The tree was manually edited to include additional populations for which viral sequences were available. In particular, Koryaks, Inuits and Navajos, Ainus, Southwest Pacific populations, and Sinhaleses were included based on their phylogenetic position as reported in literature (Friedlaender et al. 2008; Reich et al. 2012; Duda and Jan 2016; Jeong et al. 2016; Liu et al. 2017; supplementary table S1, Supplementary Material online). The resulting tree was then pruned to include only populations with at least one available complete JCPyV sequence. Because we had no information on the ethnicity of Subsaharan African subjects from which JCPyV sequences were derived, and because Subsaharan African populations are genetically diverse, we collapsed (i.e., merged) all populations in two branches (Central Africa and Southern Africa). The final tree had 44 tips (supplementary table S1, Supplementary Material online).

For the viral tree, sequences were included using the following criteria: 1) only sequences with a correspondence in the human population tree were retained; 2) for each population, one viral sequence was retained for each genotype; 3) in analogy to human populations, sequences of the same genotype from different Subsaharan African countries were collapsed based on their origin from Central or Southern Africa. The phylogenetic tree of these 70 viral sequences was obtained by pruning the BEAST tree generated earlier (supplementary table S1, Supplementary Material online).

Cophylogeny between JCPyV isolates and human populations was investigated using CoRe-PA version 0.5.1 (Merkle et al. 2010). CoRe-Pa assigns costs to four coevolutionary events: Cospeciation, duplication, sorting/loss, and host switching (herein intended as population transition). The program uses a parameter-adaptive approach to search for optimal cost values and to compute minimal-cost reconstructions of the evolutionary history between hosts and viruses. CoRe-PA analysis was performed with automatic estimation of the optimal cost setting and computed reconstructions of 5,000 random cost sets. Statistical significance was assessed with 1,000 random virus–host associations.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This study was supported by the Italian Ministry of Health ("Ricerca Corrente 2016-2018" to M.S. and "Ricerca Corrente 2018" to D.F.).

References

Agostini HT, Deckhut A, Jobes DV, Girones R, Schlunck G, Prost MG, Frias C, Perez-Trallero E, Ryschkewitsch CF, Stoner GL. 2001. Genotypes of JC virus in east, central and southwest Europe. *J Gen Virol*. 82(5):1221–1331.

Agostini HT, Yanagihara R, Davis V, Ryschkewitsch CF, Stoner GL. 1997. Asian genotypes of JC virus in native Americans and in a pacific island population: markers of viral evolution and human migration. *Proc Natl Acad Sci USA*. 94(26):14542–14546.

Aiewsakun P, Katzourakis A. 2016. Time-dependent rate phenomenon in viruses. *J Virol*. 90(16):7184–7195.

Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, Johnsrud B, Khazen G, Matisoo-Smith E, Soria-Hernanz DF, et al. 2013. Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern middle eastern populations with European and African populations. *PLoS One* 8(1):e54616.

Bay RA, Bielawski JP. 2011. Recombination detection under evolutionary scenarios relevant to functional divergence. *J Mol Evol*. 73(5–6):273–286.

Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. 2016. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol*. 33(8):2167–2169.

Boldorini R, Allegrini S, Miglio U, Paganotti A, Cocca N, Zaffaroni M, Riboni F, Monga G, Viscidi R. 2011. Serological evidence of vertical transmission of JC and BK polyomaviruses in humans. *J Gen Virol*. 92(5):1044–1050.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 10(4):e1003537.

Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, Katz JP, Pipas JM, McBride AA, Camus AC, et al. 2016. The ancient evolutionary history of polyomaviruses. *PLoS Pathog*. 12(4):e1005574.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat Genet*. 33(Suppl):266–275.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 9(8):772.

Duchene S, Holmes EC, Ho SY. 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc R Soc B*. 281:20140732.

Duda P, Jan Z. 2016. Human population history revealed by a supertree approach. *Sci Rep*. 6:29890.

Duggan AT, Evans B, Friedlaender FR, Friedlaender JS, Koki G, Merriwether DA, Kayser M, Stoneking M. 2014. Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the austronesian expansion. *Am J Hum Genet*. 94(5):721–733.

Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Culleton BJ, et al. 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570(7760):236–240.

Friedlaender JS, Friedlaender FR, Reed FA, Kidd KK, Kidd JR, Chambers GK, Lea RA, Loo JH, Koki G, Hodgson JA, et al. 2008. The genetic structure of pacific islanders. *PLoS Genet*. 4(1):e19.

Gernhard T. 2008. The conditioned reconstructed process. *J Theor Biol*. 253(4):769–778.

Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. 537:113–137.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52(5):696–704.

Hirsch HH, Kardas P, Kranz D, Leboeuf C. 2013. The human JC polyomavirus (JCPyV): virological background and clinical implications. *APMIS* 121(8):685–727.

Ikegaya H, Zheng HY, Saukko PJ, Varesmaa-Korhonen L, Hovi T, Vesikari T, Suganami H, Takasaka T, Sugimoto C, Ohasi Y, et al. 2005. Genetic diversity of JC virus in the Saami and the Finns: implications for their population history. *Am J Phys Anthropol*. 128(1):185–193.

Jeong C, Nakagome S, Di Rienzo A. 2016. Deep history of East Asian populations revealed through genetic analysis of the Ainu. *Genetics* 202(1):261–272.

Jobe DV, Friedlaender JS, Mgone CS, Agostini HT, Koki G, Yanagihara R, Ng TCN, Chima SC, Ryschkewitsch CF, Stoner GL. 2001. New JC virus

(JCV) genotypes from papua new guinea and Micronesia (type 8 and type 2E) and evolutionary analysis of 32 complete JCV genomes. *Arch Virol*. 146:2097–2113.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
Kitchen A, Miyamoto MM, Mulligan CJ. 2008. Utility of DNA viruses for studying human host history: case study of JC virus. *Mol Phylogenet Evol*. 46(2):673–682.
Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, Stoneking M, Ko YC. 2014. Early austronesians: into and out of Taiwan. *Am J Hum Genet*. 94(3):426–436.
Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 23(10):1891–1901.
Kosakovsky Pond SL, Posada D, Stawiski E, Chappey C, Poon AF, Hughes G, Fearnhill E, Gravenor MB, Leigh Brown AJ, Frost SD. 2009. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol*. 5(11):e1000581.
Krause-Kyora B, Susat J, Key FM, Kuhnert D, Bosse E, Immel A, Rinne C, Kornell SC, Yepes D, Franzenburg S, et al. 2018. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* 7:36666.
Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. 2009. Phylogenetics, evolution, and medical importance of polyomaviruses. *Infect Genet Evol*. 9(5):784–799.
Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol*. 5(9):e1000520.
Lipson M, Skoglund P, Spriggs M, Valentin F, Bedford S, Shing R, Buckley H, Phillip I, Ward GK, Mallick S, et al. 2018. Population turnover in remote Oceania shortly after initial settlement. *Curr Biol*. 28(7):1157–1165.e7.
Liu X, Lu D, Saw WY, Shaw PJ, Wangkumhang P, Ngamphiw C, Fucharoen S, Lert-Itthiporn W, Chin-Inmanu K, Chau TN, et al. 2017. Characterising private and shared signatures of positive selection in 37 Asian populations. *Eur J Hum Genet*. 25(4):499–508.
Madinda NF, Ehlers B, Wertheim JO, Akoua-Koffi C, Bergl RA, Boesch C, Akonkwa DB, Eckardt W, Fruth B, Gillespie TR, et al. 2016. Assessing host–virus codivergence for close relatives of merkel cell polyomavirus infecting African great apes. *J Virol*. 90(19):8531–8541.
Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206.
Martin DP, Murrell B, Khoosal A, Muhire B. 2017. Detecting and analyzing genetic recombination using RDP4. *Methods Mol Biol*. 1525:433–460.
Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16(6):562–563.
Membrebe JV, Suchard MA, Rambaut A, Baele G, Lemey P. 2019. Bayesian inference of evolutionary histories under time-dependent substitution rates. *Mol Biol Evol*. 36(8):1793–1803.
Merkle D, Middendorf M, Wieseke N. 2010. A parameter-adaptive dynamic programming approach for inferring cophylogenies. *BMC Bioinformatics* 11(1 Suppl):S60-2105-11-S1-S60.
Moens U, Krumbholz A, Ehlers B, Zell R, Johne R, Calvignac-Spencer S, Lauber C. 2017. Biology, evolution, and medical importance of polyomaviruses: an update. *Infect Genet Evol*. 54:18–38.
Molinaro L, Pagani L. 2018. Human evolutionary history of Eastern Africa. *Curr Opin Genet Dev*. 53:134–139.
Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, et al. 2010. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet*. 42(12):1140–1143.
Moreno-Mayar JV, Potter BA, Vinner L, Steinrucken M, Rasmussen S, Terhorst J, Kamm JA, Albrechtsen A, Malaspinas AS, Sikora M, et al. 2018. Terminal Pleistocene Alaskan genome reveals first founding population of native Americans. *Nature* 553(7687):203–207.
Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, Allentoft ME, Vimala T, Racimo F, Pinotti T, et al. 2018. Early human dispersals within the Americas. *Science* 362(6419):eaav2621.
Muhlemann B, Margaryan A, Damgaard PB, Allentoft ME, Vinner L, Hansen AJ, Weber A, Bazaliiskii VI, Molak M, Arneborg J, et al. 2018. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc Natl Acad Sci USA*. 115:7557–7562.
Muhlemann B, Jones TC, Damgaard PB, Allentoft ME, Shevnina I, Logvin A, Usmanova E, Panyushkina IP, Boldgiv B, Bazartseren T, et al. 2018. Ancient hepatitis B viruses from the bronze age to the medieval period. *Nature* 557:418–423.
Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* 541(7637):302–310.
Pavesi A. 2003. African origin of polyomavirus JC and implications for prehistoric human migrations. *J Mol Evol*. 56(5):564–572.
Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 31(7):1929–1936.
Pimenoff VN, de Oliveira CM, Bravo IG. 2017. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. *Mol Biol Evol*. 34(1):4–19.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.
Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA*. 98(24):13757–13762.
Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499(7459):471–475.
Pyoria L, Toppinen M, Mantyla E, Hedman L, Aaltonen LM, Vihinen-Ranta M, Ilmarinen T, Soderlund-Venermo M, Hedman K, Perdomo MF. 2017. Extinct type of human parvovirus B19 persists in tonsillar B cells. *Nat Commun*. 8:14930.
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol*. 67(5):901–904.
Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. 2012. Reconstructing native American population history. *Nature* 488(7411):370–374.
Ryschkewitsch CF, Friedlaender JS, Mgone CS, Jobes DV, Agostini HT, Chima SC, Alpers MP, Koki G, Yanagihara R, Stoner GL. 2000. Human polyomavirus JC variants in Papua New Guinea and Guam reflect ancient population settlement and viral evolution. *Microbes Infect*. 2(9):987–996.
Santos E, McCabe A, Gonzalez-Galarza FF, Jones AR, Middleton D. 2016. Allele frequencies net database: improvements for storage of individual genotypes and analysis of existing data. *Hum Immunol*. 77(3):238–248.
Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol*. 6(5):526–538.
Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483(7388):169–175.
Shackelton LA, Rambaut A, Pybus OG, Holmes EC. 2006. JC virus evolution and its association with human populations. *J Virol*. 80(20):9928–9933.
Sharp PM, Simmonds P. 2011. Evaluating the evidence for virus/host coevolution. *Curr Opin Virol*. 1(5):436–441.
Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C, Petchey F, Fernandes D, et al. 2016. Genomic insights into the peopling of the southwest pacific. *Nature* 538(7626):510–513.
Smith JM. 1992. Analyzing the mosaic structure of genes. *J Mol Evol*. 34(2):126–129.

Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet*. 16(9):530–542.

Stoner GL, Jobes DV, Fernandez Cobo M, Agostini HT, Chima SC, Ryschkewitsch CF. 2000. JC virus as a marker of human migration to the Americas. *Microbes Infect*. 2(15):1905–1911.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 4(1):vey016.

Sugimoto C, Hasegawa M, Kato A, Zheng HY, Ebihara H, Taguchi F, Kitamura T, Yogo Y. 2002. Evolution of human polyomavirus JC: implications for the population history of humans. *J Mol Evol*. 54(3):285–297.

Sugimoto C, Hasegawa M, Zheng HY, Demenev V, Sekino Y, Kojima K, Honjo T, Kida H, Hovi T, Vesikari T, et al. 2002. JC virus strains indigenous to northeastern Siberians and Canadian Inuits are unique but evolutionally related to those distributed throughout Europe and Mediterranean areas. *J Mol Evol*. 55(3):322–335.

Sugimoto C, Kitamura T, Guo J, Al-Ahdal MN, Shchelkunov SN, Otova B, Ondrejka P, Chollet JY, El-Safi S, Ettayebi M, et al. 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc Natl Acad Sci USA*. 94(17):9191–9196.

Sundqvist E, Buck D, Warnke C, Albrecht E, Gieger C, Khademi M, Lima Bomfim I, Fogdell-Hahn A, Link J, Alfredsson L, et al. 2014. JC polyomavirus infection is strongly controlled by human leucocyte antigen class II variants. *PLoS Pathog*. 10(4):e1004084.

Sunyaev SR, Lugovskoy A, Simon K, Gorelik L. 2009. Adaptive mutations in the JC virus protein capsid are associated with progressive multifocal leukoencephalopathy (PML). *PLoS Genet*. 5(2):e1000368.

Takasaka T, Kitamura T, Sugimoto C, Guo J, Zheng HY, Yogo Y. 2006. Phylogenetic analysis of major African genotype (Af2) of JC virus: implications for origin and dispersals of modern Africans. *Am J Phys Anthropol*. 129(3):465–472.

Toppinen M, Perdomo MF, Palo JU, Simmonds P, Lycett SJ, Soderlund-Venermo M, Sajantila A, Hedman K. 2015. Bones hold the key to DNA virus history and epidemiology. *Sci Rep*. 5:17226.

Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol*. 30(7):1720–1728.

Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. *Mol Phylogenet Evol*. 26(1):1–7.

Zheng HY, Sugimoto C, Hasegawa M, Kobayashi N, Kanayama A, Rodas A, Mejia M, Nakamichi J, Guo J, Kitamura T, et al. 2003. Phylogenetic relationships among JC virus strains in Japanese/Koreans and Native Americans speaking Amerind or Na-dene. *J Mol Evol*. 56(1):18–27.