 doi:10.20944/preprints202012.0310.v1

Article

Phylogeographical and evolutionary history of *variola major* virus; a question of timescales?

Annalisa Bergna[1], Carla Della Ventura[1], Rossella Marzo[1], Massimo Ciccozzi[2], Massimo Galli[1], Gianguglielmo Zehender[1,3] and Alessia Lai[1]*

[1] Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, Milan, Italy.;
[2] Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico of Rome, Rome, Italy;
[3] Coordinated Research Center "EpiSoMI", University of Milan, Milano, Italy

* Correspondence: e-mail: alessia.lai@unimi; Phone (+39) 0250319775.

Abstract: In order to reconstruct the origin and pathways of *variola* virus (VARV) dispersion, we analyzed 47 VARV isolates available in public databases and their SNPs.
The mean substitution rate of the whole genomes was $9.41x10^{-6}$ (95%HPD:8.5-11.3x10^{-6}) substitutions/site/year. The time of the tree root was estimated to be a mean 68 years (95%HPD:60.5–75.9).
The phylogeographical analysis showed that the Far East and India were the most probable locations of the tree root and of the inner nodes, respectively, whereas for the outer nodes it corresponded to the sampling locations.
The Bayesian Skyline plot showed that the effective number of infections started to grow exponentially in 1915-1920, peaked in the 1940s, and then decreased to zero.
Our results suggests that the VARV *major* strains circulating between 1940s-1970s probably shared a common ancestor originated in the Far East; subsequently moved to India, which became the center of its dispersion to eastern and southern Africa, and then to central Africa and the Middle East, probably following the movements of people between south-eastern Asia and the other places with a common colonial history.
These findings may help to explain the controversial reconstructions of the history of VARV obtained using long- and short- term calibrations.

Keywords: *Variola major*; phylogeographical analysis; long- term calibrations; short- term calibrations.

1. Introduction

Smallpox was once a severe infectious disease caused by the *variola virus* (VARV) whose outbreaks across the world were a major cause of mortality. It was responsible for hundreds of millions of deaths as it not only negatively affected populations' growth in the Old World, but also contributed to the crises affecting established civilizations in the New World when their populations were decimated after they came into contact with Europeans [1].

However, in 1980, it was officially certified that the intensified Smallpox Eradication Programme (SEP) approved by WHO in 1966 for eliminating endemic disease in Africa, Asia, and South America, and preventing its return to Europe [2], had ensured the global elimination of a disease for the first time in human history. The last natural case of smallpox was recorded in Somalia in 1977 [3, 4].

Two principal variants of VARV infections were *variola major*, which was associated with an overall case fatality rates (CFR) of around 20%, and *variola minor* (also called alastrim), which was common in western Africa and the New World in the late XIX century but had lower CFR of <1%. There was no credible description of smallpox in the Americas or sub-Saharan Africa before the westward exploration of the XV exported the disease to their aboriginal populations [5] .

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

Although it has been claimed that cases of smallpox dating back thousands of years were observed in Egypt, China and India, the timescale of the emergence of *variola* virus and its evolution are unclear because there are significant gaps in historical medical records and controversies concerning the ancient or more recent origin of various forms of smallpox [5].

However, despite differences in the methods used, the viral genes considered and the sequences included, all of the available studies propose similar dates for VARV, and similar substitution rates. The findings of molecular clock-based studies suggest that key events in the evolution of smallpox have occurred over the last two centuries [6] and that the VARV strains circulating in the XX century had a common ancestor dating back to about 500 or 1,000 years ago (YA) [6, 7], and Hughes *et al*. [8] have dated the divergence of *variola minor* and *variola major* to 700-1,000 YA.

In a previous study based on hemagglutinin sequences [1], we described the evolutionary history of the entire *Orthopoxvirus* (OPV) genus: it was estimated that the root of the VARV clade dated back to 720 YA (corresponding to about 1,300 AD) and that *variola major* and *minor* diverged about 500 YA (corresponding to about XV-XVI century). The main limitation of this study was that it was based on the analysis of partial conserved regions of the viral genome; moreover, almost all the available VARV sequences fall within a narrow time frame and are possibly the expression of a recent epidemic re-ignition.

The aim of this study was to analyze the whole genomes and single nucleotide polymorphisms (SNPs) of human VARV available in public databases (which include viral strains circulating between the 1940s and the 1970s) in order to describe VARV genealogy on the basis of XX century epidemics by means of a phylogeographical and temporal reconstruction, and to assess the correlations between the phylogeography and historical data of the latest epidemics.

2. Materials and Methods

2.1. Sequence datasets

Forty-seven whole genome (WG) isolates with known sampling time and location, were retrieved from public databases (Genbank at: http://www.ncbi.nlm.nih.gov/genbank/).

The VARV sequences came from Ethiopia (n=2), Somalia (n=3), Tanzania (n=1), Sudan (n=2), South Africa (ZA), Botswana (n=2), Sierra Leone (n=1), Nigeria (n=1), Ghana (n=1), Benin (n=1), Bangladesh (n=4), Nepal (n=1), India (n=4), Germany (n=1), the United Kingdom (n=4), Afghanistan (n=1), Iran (n=1), Kuwait (n=1), Pakistan (n=1), Syria(n=1), Yugoslavia (n=1), Japan (n=3), China (n=1), Korea (n=1), Indonesia (n=2), Mexico (n=1) and Brazil (n=1).

2.2 Single Nucleotide Polymorphisms

WG sequences with a length of 183,938 bp were aligned and cropped using the Mauve alignment system [9] (http://darlinglab.org/mauve/mauve.html) available in Geneious software v.9.1.8 (http://www.geneious.com), and then used to construct multiple genome alignments of large-scale evolutionary events. The obtained FASTA alignment was then used to generate 1,625 bp long SNPs using the MEGA program v.10.1 (http://www.megasoftware.net/).

2.3 Datasets, model selection and evolutionary rates estimate

In order to estimate the substitution rates, two datasets were constructed: the first consisting of the 47 WG sequences, and the second with an equal number of SNPs sequences. Given the lack of a representative number of *variola minor* sequences (P2), a subset of 40 strains exclusively containing P1 sequences (thus excluding alastrim) was used to create the most recent phylogeographical reconstruction relating to the last epidemic of VARV.

 doi:10.20944/preprints202012.0310.v1

The best-fitting nucleotide substitution for the data was estimated using a hierarchical likelihood ratio approach using Modeltest 3.7 [10], and the GRT (general time reversible) model was selected.

A Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST 1.8.4 [11] was used to estimate evolutionary rates and the time of the Most Recent Common Ancestor (tMRCA). The coalescent priors of a constant population size, exponential growth, logistic growth, and a piecewise-constant Bayesian skyline plot (BSP) [12], were tested under strict clock conditions as previously described [1]. The chains were run for 30 million generations and sampled every 3,000 steps until reaching convergence which was assessed on the basis of the effective sampling size (ESS>200) after a 10% burn-in using Tracer software v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Uncertainty in the estimates was indicated by 95% highest probability density (HPD) intervals, and the best fitting models were selected using a Bayes Factor (BF) with marginal likelihoods implemented in BEAST.

In accordance with Kass and Raftery [13], the strength of the evidence against the null hypothesis (H0) was evaluated as follows: 2lnBF <2 no evidence; 2–6 weak evidence; 6–10 strong evidence, and >10 very strong evidence. A negative 2lnBF indicates evidence in favor of H0. A less restrictive Bayesian skyline plot was used as the coalescent prior. Only values of >6 were considered significant. BF calculations were made using Tracer v.1.6.

The maximum clade credibility (MCC) tree was then selected from the posterior tree distribution after excluding a 10% burn-in using the TreeAnnotator program v.1.8.4 included in the BEAST package, and visualized using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). The tMRCA estimates were expressed as the mean number of years and 95%HPD years before the most recent sampling date, corresponding to 1977.

2.4 Phylogeographical analysis

The phylogeographical analysis of the subset of 40 P1 sequence was made using Beast 1.8.4 by assigning the isolates to six discrete groups on the basis of their sampling location: Africa (eastern Africa/EA, n=8; central Africa/CA, n=2; southern Africa/SA, n=4); India (IN, n=13), the Middle East (ME, n=6) and the Far East (FE, n=7). In the case of the strains isolated in Europe, where smallpox was eradicated in the 1940s, we used the probable place of origin of the infection rather than the sampling location. The previously selected evolutionary model, strict clock and BSP model were used. The analyses were made using the continuous-time Markov chain (CTMC) process over discrete sampling locations [14] and a Bayesian stochastic search variable selection (BSSVS) approach in order to allow diffusion rates to be zero with a positive prior probability.

In order to estimate the direction of dispersion, the sequence data were fitted for two discrete-trait models: a symmetrical model (which assumes non-zero rates of change between each pair of discrete states are equal) and the asymmetrical model, which assumes that the non-zero rates of change between each pair of discrete states are different. Comparison of the posterior and prior odds that the individual rates are non-zero provides a formal BF for testing the significance of the linkage between locations. Rates yielding a BF >3 were considered significant [14].

MCMC chains of the two dataset were run for 30 million generations and sampled every 3,000 generations until reaching convergence which was assessed on the basis of the effective sampling size (ESS ≥ 200) after a 10% burn-in by using Tracer software v.1.6. The final MCC tree was visualized using FigTree v. 1.4.2, which is freely available on the web (http://tree.bio.ed.ac.uk/software/figtree/), and the most probable location at each node was highlighted by labelling the branches with different colors.

In order to visualize diffusion rates over time, it is possible to convert the location-annotated MCC tree to a keyhole mark-up language (KML) file that is suitable for viewing with geo-referencing software using the SPREAD program (available at http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html). The migration routes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

indicated by the tree were visualized using Google Earth in order to provide a spatial projection (http://earth.google.com).

3. Results

3.1 Evolutionary rates and tMRCA estimations

The mean substitution rate on WG sequences was $9.41x10^{-6}$ substitutions/site/year (95%HPD: $8.3\text{-}10.6x10^{-6}$) considering the entire dataset. When the SNPs were used, it was $1.12x10^{-3}$ (95%HPD: $0.97\text{-}1.28x10^{-3}$) corresponding to a WG evolutionary rates of $9.89x10^{-6}$ (95%HPD: $8.5\text{-}11.3x10^{-6}$). Comparable results were obtained considering only the P1 dataset.

The Bayesian dated tree of the WG sequences (Figure 1) showed two main significant clades with a high posterior probability (pp) corresponding to the P1 and P2 lineages (pp=1 for both). The P1 lineage included isolates from all over the Old World (Africa, Asia and Europe) whereas the P2 lineage only included strains from America (alastrim) and western Africa. The tMRCA estimation showed that the root of tree, corresponding to the divergence between P1 and P2, went back to the late 1700s, while the radiation of the extant P1 and P2 lineages was the first decade of the 1900 and the second half of 1800, respectively. The same tree topology and similar tMRCA estimations were obtained using SNPs dataset (see Table S1).

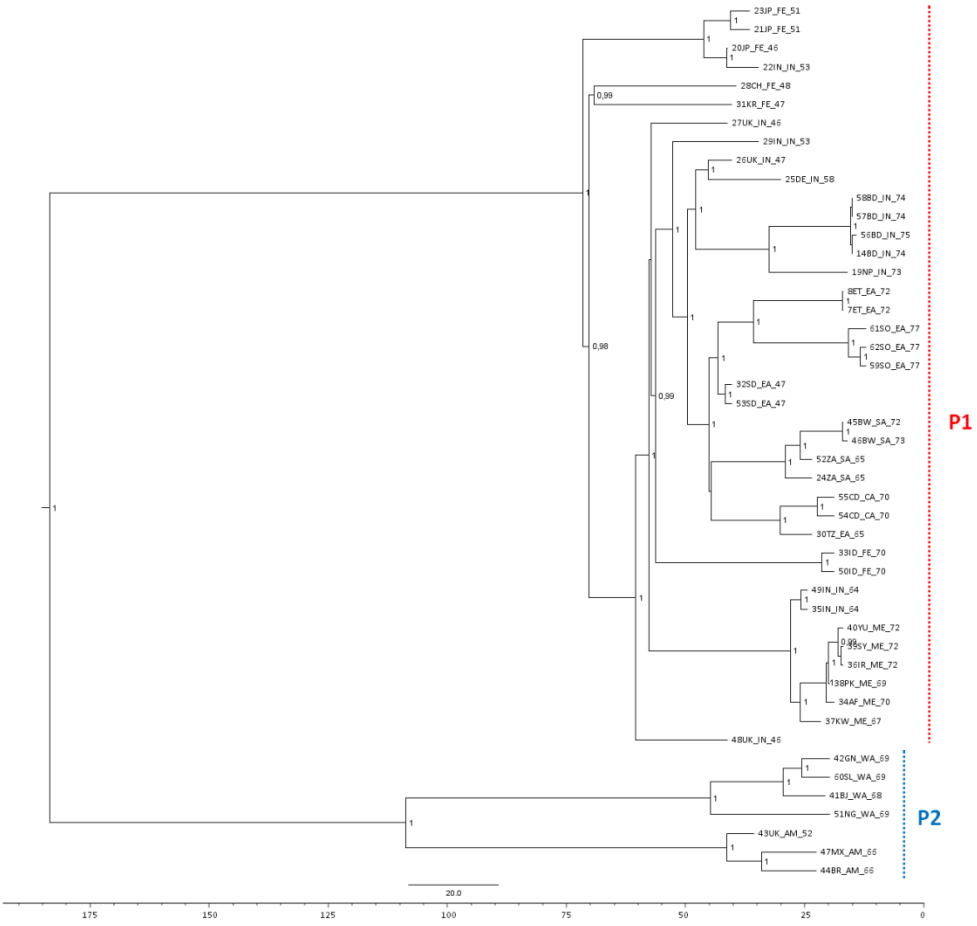

Figure 1. Dated tree, including all VARV whole genome sequences available (P1=VARV *major*; P2=VARV *minor*,alastrim). Significant statistical support for clades along the branches (posterior probability >0.7) are reported. The scale axis below of the tree shows the years before the last sampling time, corresponding to 1977.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

3.2 Phylogeographical reconstruction

Given the scarcity of sequences of the P2 clade, we limited the phylogeographical analysis to the lineage P1, including only Old World VARV genomes.

The Bayesian phylogeographical maximum credibility tree, shown in Figure 2, of the P1 clade indicated the Far East as the most probable location of the root (which included sequences from China, Korea and Japan) with a state posterior probability (spp) of 1.

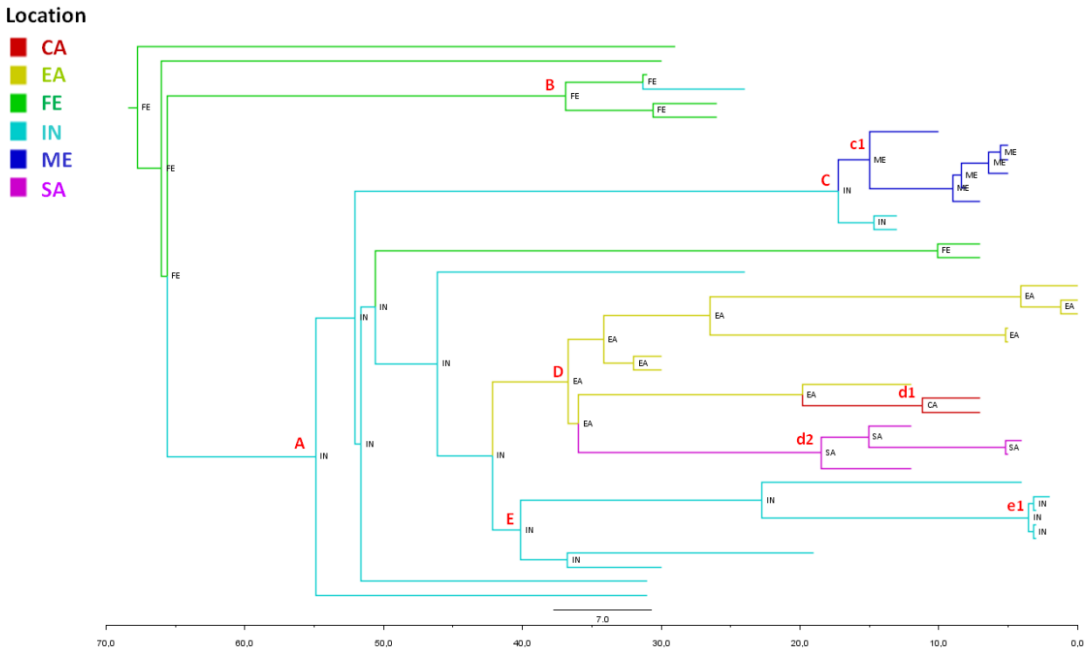

Figure 2. Bayesian phylogeographical tree of Single Nucleotide Polymorphisms P1 sequences. Branches are coloured on the basis of the most probable location of the descendent nodes that is indicated. The colour legend is shown in the panel (bottom left). (CA= central Africa; EA= eastern Africa; FE= Far East; IN= India; ME= Middle East; SA= southern Africa). The scale axis below the tree shows the years before the last sampling time (1977).

The tree showed several nested clades (see Table 1), the largest of which (Node A) have India as a most probable place of origin (spp=1), as well as the deepest nodes of the clade (nodes C and E). The terminal subclades, on the other hand, included isolates from homogeneous geographic areas segregating significantly and showing different MRCA-locations (node c1: Middle East, node d1: Central Africa, node d2: South Africa, Node D: East Africa, and node B: Far East).

Table 1. MRCA-locations and tMRCA estimations in different clades.

Clade	Location	spp[a]	tMRCA[b]	95%HPDL[c]	95%HPDH[d]	Year	95%HPDL	95%HPDH
Root	FE	1	67.8	61	76.11	1909.2	1901	1916
A	IN	1	54.9	49.5	60.9	1922.1	1916.1	1927.5
B	FE	1	36.9	34.1	40	1940.1	1937	1942.9
C	IN	1	17.2	15.2	19.4	1959.8	1957.6	1961.8
c1	ME	0.99	15	12.8	17.4	1962	1959.6	1964.2
D	EA	1	36.7	33.8	40	1940.3	1937	1943.2
d1	CA	1	11.2	9.3	13.2	1965.8	1963.8	1967.7
d2	SA	1	18.5	15.7	21.2	1958.5	1955.8	1961.3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

E	IN	0.96	40.1	36.3	44.4	1936.9	1932.6	1940.7
e1	IN*	1	3.5	3.1	4.1	1973.5	1972.9	1973.9

[a] spp, state posterior probability
[b] tMRCA, time of the Most Recent Common Ancestor
[c] 95%HPDL, Highest Posterior Density Low Interval
[d] 95%HPDH, Highest Posterior Density High Interval
* Bangladesh

The Bayesian symmetrical and asymmetrical phylogeographic diffusion models indicated five highly supported linkages from eastern to southern Africa (BF=26.4) and central Africa (BF=64.5), from the Far East to India (BF=427.5), and from India to the Middle East (BF=99.6) and eastern Africa (BF=34.5).

Phylogeographical fluxes (Figure 3) indicated that the tMRCA of the P1 clade existed in the Far East in 1909 and moved to India in 1920-1930 (1922) from where it arrived in eastern Africa in about 1940. from which the virus spread to all the continent, reaching southern Africa in the 1950s-1960s (1959) and central Africa in the 1960s-1970s (1966). Finally the virus reached the Middle East starting always from India in the 1960s (1962).

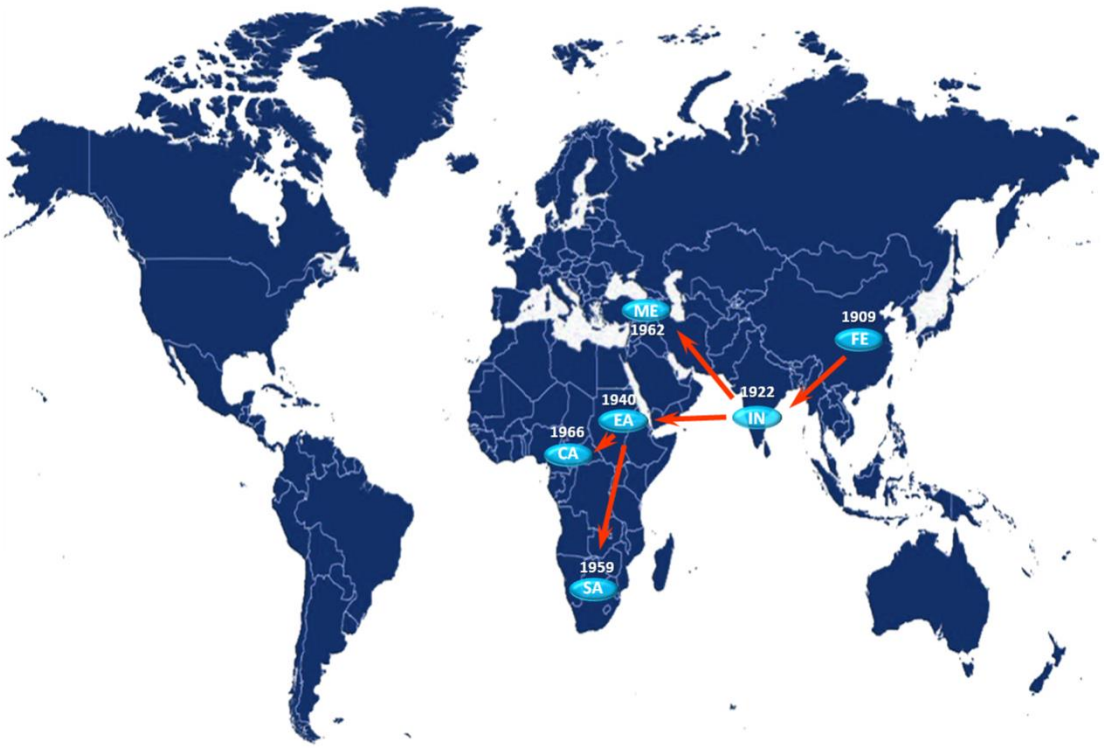

Figure 3. Significant non-zero migration rates of VARV P1 worldwide supported by a Bayes Factor >4. The migrations were calculated using SPREAD program. The dates in which the virus entered the area and the direction of fluxes are shown (CA= central Africa; EA= eastern Africa; FE= Far East; IN= India; ME= Middle East; SA= southern Africa).

3.3 Phylodynamic analyses

The Bayesian Skyline plot showed that the effective number of infections grew exponentially between the 1920s and 1945, and then sharply decreased significantly from about 1970 until reaching almost zero in the second half of 1970s (Figure 4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

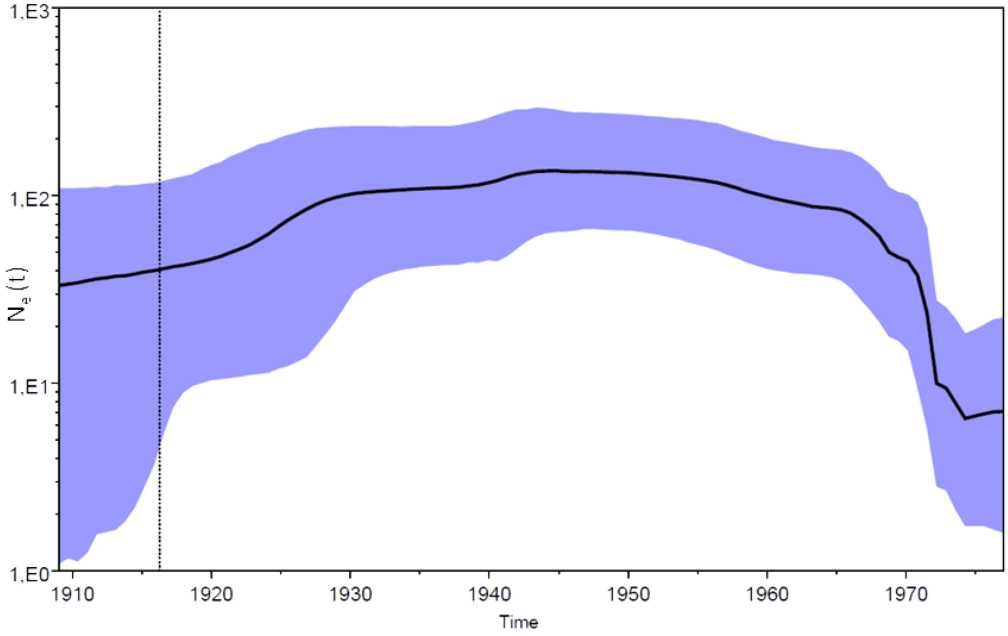

Figure 4. Skyline plot obtained by analysing the Single Nucleotide Polymorphisms dataset including only P1 sequences. Ordinate: the number of effective infections at time t ($N_e(t)$); abscissa: time in years before the present. The thick solid line represents the median value of the $N_e(t)$ estimates, and the grey area the 95%HPD. The vertical lines indicate the 95% lower HPD (dotted) and the mean tMRCA estimate (bold) of the tree root.

4. Discussion

In the late 1970s, the global collaborative programme of mass vaccination obtained the worldwide eradication of smallpox causing the extinction of VARV in nature. Today, genomic studies of VARV are possible only on sequences available in public databases, obtained from the strains circulating between the 1940s and late 1970s.

To date, several authors have attempted to reconstruct the evolutionary history of the human smallpox virus on a time scaled phylogeny, obtaining similar estimates of the origin of virus, roughly between 700 and 1,000 years ago [6, 7], regardless of whether they used whole genomes, partial sequences or different genes.

Our recent study aimed at reconstructing the evolutionary history of the entire *Orthopoxvirus* genus led to an estimated tMRCA of about 10,000 YA for the entire animal tree root and a mean tMRCA of 720 YA for the VARV species [1].

Our preliminary analysis was based on a single gene (B7R encoding for the viral protein hemagglutinin) and an external historical calibration that assumed the post-Colombian divergence of the South American and western Africa *variola minor* clades, and indicated a mean evolutionary rate of the entire tree of about $6x10^{-6}$ substitutions/site/year. The calibration used in the present study included "heterochronous" sequences with known sampling dates that allowed us to reconstruct the recent phylogeographical history of the last *variola major* virus, which circulated between the 1940s and 1970s, in the years immediately preceding its extinction. This short timescale covering a time span of only 250 years, it is very different from that of our previous study, which covered up to 10,000 years.

The newly estimated mean evolutionary rate on the entire viral genomes and SNPs is $9.4x10^{-6}$ sub/site/year, which is 30% higher than that previously estimated on hemagglutinin gene (although not very different in terms of the 95% IC). Under these conditions, we estimate that the divergence of P1 and P2 occurred in the XVIII century, and that the time of radiation of P1 and P2 was respectively the first decade of 1900s and the second half of 1800s. Very similar results have recently been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 December 2020 doi:10.20944/preprints202012.0310.v1

obtained by Duggan *et al*., who studied an ancient strain of VARV obtained from a Lithuanian mummy [6].

This apparently recent history of VARV may be due to the fact that we can now only see the history of extant lineages, whereas basal lineages can disappear due to vaccination, as is also suggested by the findings of studies of the very few available ancient sequences showing that they are always basal to all the XIX century sequences [15]. The oldest sequences in our analysis were isolated in the 1940s, when endemic smallpox had already been eliminated in a number of countries, including Europe. The number of reported deaths due to smallpox decreased in the first decade of the 1900s in all European countries except Russia, Germany and Finland [2] (see Figure S1): for example, the last major outbreaks of *variola major* in England and Wales occurred in 1902-1905 [2, 16].

The long branch connecting P1 and P2 in the complete VARV phylogeny found by us and Duggan *et al.* [6] suggest the existence of major bottlenecks in both clades during the XIX and XX centuries, probably because the increase in global smallpox vaccination led to the extinction of various lineages [6].

However, the deleterious effect of the First World War on the health conditions of civilians led to large smallpox epidemics affecting tens of thousands of people in Italy, Austria, Spain, Portugal, Russia [2], and these contrasting phenomena may explain the phylogenetic bottleneck observed in these years in the P1 clade.

Although smallpox had been eliminated from almost all European countries by the late 1930s, isolated incidents occurred during and after the Second World War in Italy, Spain and Greece causing small outbreaks due to cases imported from colonial possessions or neighboring countries in which VARV was still endemic [2].

Furthermore, smallpox remained endemic in various other countries until the late 1960s, when it started disappearing around the world due to the WHO's international strategic action plan for intensifying its eradication, which was introduced in 1967 and led to the complete eradication of the disease by 1980.

The skyline plot of the P1 clade seems to reflect these epidemiological data as it shows a gradual increase in the effective number of infections between the root of the tree (1909) and the 1940s, when the curve reached a plateau before rapidly falling between 1965 and the 1970s. The consistency between our phylodynamic analysis and the known epidemiological data supports the reliability of our evolutionary rate and tMRCA estimates.

Given the scarcity of *variola minor* sequences, we limited our phylogeographical reconstruction to the P1 *variola major* clade. On the basis of this analysis, we estimated that the common ancestor of the *variola major* virus circulating in the Old World between the 1940s and the 1970s, originated in the Far East in the first two decades of the XX century, and then spread to Indian subcontinent in the 1920s.

Smallpox was highly endemic in the Far East in the first decades of the XX century, and caused mainly seasonal epidemics [2]. Although the virus was eliminated in some countries as early as the 1930s, this did not happen until the period between the 1950s and the 1970s in China (where the vaccination was not introduced until the 1950s), Japan and Indonesia [2].

The migration of the virus to the Indian sub-continent in the 1920s may have been due to the close relationships between former British colonies, (particularly between British India and Hong Kong), before and during the First World War. Although the introduction of smallpox vaccinations in British India in the late XIX century led to a decrease in the number of deaths, smallpox remained endemic even after the partition of British India in western and eastern Pakistan in 1947 (Pakistan and Bangladesh, after 1971). Particularly, a major epidemic in Indian port cities in the 1930s required the restriction of regional maritime trading. With the sole exception of Sri Lanka, smallpox was definitively eliminated from the Indian subcontinent only after the 1970s [2].

In our reconstruction, India acted as an important center of spreading of VARV, exporting the infection to East Africa in the 1940s and to the Middle East in the '60s.

During the second WW several British Indian divisions were engaged in the East African campaign against Italian colonies possibly justifying the exportation of Asian smallpox to East Africa.

After the Second World War, cases of smallpox outbreak occurred in England and Wales due to the repatriation of troops from endemic areas [2, 17]. Although smallpox vaccinations were introduced in eastern Africa in the 1940s, the disease remained endemic until the 1970s, and Somalia suffered the very last *variola major* epidemic in 1977. In the phylogeographic reconstruction, the virus spread to other African countries, in the central and southern part of the continent between 1950s and 1960s, in particular in the former British colonies where important communities of South Asian descendent people were living [2].

Finally, Muslim pilgrimages to the holy cities and the presence of migrant workers from endemic areas in the Gulf States contributed to the resurgence of smallpox in Iran, Iraq and Syria in the early 1960s, justifying our observation of a flux of virus from India to the Middle East in the '60s [2]. The return from a pilgrimage to Mecca is also associated with the last smallpox outbreak in Europe, which occurred in Yugoslavia in 1972 [18].

Our results may help to explain the controversies concerning the phylogenetic reconstruction of the history of VARV using different calibration scales and clarify the epidemiology of the last circulating strains.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Table S1**: tMRCA estimates comparison among different datasets (WG/SNPs); **Figure S1**: Number of reported deaths due to smallpox in the first decade of the 1900s in all European countries except Russia, Germany and Finland.

Author Contributions: AL, AB, GZ, and MG conceived and designed the study. AB, AL, CDV, GZ, MC and RM made the phylogenetic analyses. AL, AB, GZ, and MG wrote the first draft of the manuscript. All of the authors contributed to revising the manuscript, and read and approved the submitted version.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zehender, G.; Lai, A.; Veo, C.; Bergna, A.; Ciccozzi, M.; Galli, M., Bayesian reconstruction of the evolutionary history and cross-species transition of variola virus and orthopoxviruses. *Journal of medical virology* **2018,** 90, 1134-1141.
2. Fenner, F., *Smallpox and its eradication*. World Health Organization: 1988.
3. Sanchez-Sampedro, L.; Perdiguero, B.; Mejias-Perez, E.; Garcia-Arriaza, J.; Di Pilato, M.; Esteban, M., The evolution of poxvirus vaccines. *Viruses* **2015,** 7, 1726-803.
4. Shchelkunova, G. A.; Shchelkunov, S. N., 40 Years without Smallpox. *Acta naturae* **2017,** 9, 4-12.
5. Li, Y.; Carroll, D. S.; Gardner, S. N.; Walsh, M. C.; Vitalis, E. A.; Damon, I. K., On the origin of smallpox: correlating variola phylogenics with historical smallpox records. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, 15787-92.
6. Duggan, A. T.; Perdomo, M. F.; Piombino-Mascali, D.; Marciniak, S.; Poinar, D.; Emery, M. V.; Buchmann, J. P.; Duchene, S.; Jankauskas, R.; Humphreys, M.; Golding, G. B.; Southon, J.; Devault, A.; Rouillard, J. M.; Sahl, J. W.; Dutour, O.; Hedman, K.; Sajantila, A.; Smith, G. L.; Holmes, E. C.; Poinar, H. N., 17(th) Century Variola Virus Reveals the Recent History of Smallpox. *Current biology : CB* **2016,** 26, 3407-3412.
7. Biagini, P.; Theves, C.; Balaresque, P.; Geraut, A.; Cannet, C.; Keyser, C.; Nikolaeva, D.; Gerard, P.; Duchesne, S.; Orlando, L.; Willerslev, E.; Alekseev, A. N.; de Micco, P.; Ludes, B.; Crubezy, E., Variola virus in a 300-year-old Siberian mummy. *The New England journal of medicine* **2012,** 367, 2057-9.
8. Hughes, A. L.; Irausquin, S.; Friedman, R., The evolutionary biology of poxviruses. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **2010,** 10, 50-9.
9. Darling, A. C.; Mau, B.; Blattner, F. R.; Perna, N. T., Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* **2004,** 14, 1394-403.
10. Posada, D., jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **2008,** 25, 1253-6.
11. Drummond, A. J.; Suchard, M. A.; Xie, D.; Rambaut, A., Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* **2012,** 29, 1969-73.

 doi:10.20944/preprints202012.0310.v1

12. Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **2010,** 59, 307-21.
13. Kass, R. E.; Raftery, A. E., Bayes factors. *Journal of the american statistical association* **1995,** 90, 773-795.
14. Lemey, P.; Rambaut, A.; Drummond, A. J.; Suchard, M. A., Bayesian phylogeography finds its roots. *PLoS computational biology* **2009,** 5, e1000520.
15. Wertheim, J. O., Viral Evolution: Mummy Virus Challenges Presumed History of Smallpox. *Current biology : CB* **2017,** 27, R119-R120.
16. Geddes, A. M., The history of smallpox. *Clinics in dermatology* **2006,** 24, 152-7.
17. Millward, G., *Vaccinating Britain*. Manchester University Press: 2019.
18. Ristanovic, E.; Gligic, A.; Atanasievska, S.; Protic-Djokic, V.; Jovanovic, D.; Radunovic, M., Smallpox as an actual biothreat: lessons learned from its outbreak in ex-Yugoslavia in 1972. *Annali dell'Istituto superiore di sanita* **2016,** 52, 587-597.