On the origin of smallpox: Correlating variola phylogenics with historical smallpox records

Yu Li[†], Darin S. Carroll[†], Shea N. Gardner[‡], Matthew C. Walsh[†§], Elizabeth A. Vitalis[‡], and Inger K. Damon[†¶]

[†]Poxvirus and Rabies Branch, Division of Viral and Rickettsial Diseases, National Center for Zoonotic, Vector-Borne, and Enteric Diseases, Coordinating Centers for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30329; and [‡]Pathogen Bio-Informatics, Lawrence Livermore National Laboratory, Livermore, CA 94550

Edited by Bernard Moss, National Institutes of Health, Bethesda, MD, and approved August 15, 2007 (received for review October 19, 2006)

Human disease likely attributable to variola virus (VARV), the etiologic agent of smallpox, has been reported in human populations for >2,000 years. VARV is unique among orthopoxviruses in that it is an exclusively human pathogen. Because VARV has a large, slowly evolving DNA genome, we were able to construct a robust phylogeny of VARV by analyzing concatenated single nucleotide polymorphisms (SNPs) from genome sequences of 47 VARV isolates with broad geographic distributions. Our results show two primary VARV clades, which likely diverged from an ancestral African rodent-borne variola-like virus either ≈16,000 or ≈68,000 years before present (YBP), depending on which historical records (East Asian or African) are used to calibrate the molecular clock. One primary clade was represented by the Asian VARV major strains, the more clinically severe form of smallpox, which spread from Asia either 400 or 1,600 YBP. Another primary clade included both alastrim minor, a phenotypically mild smallpox described from the American continents, and isolates from West Africa. This clade diverged from an ancestral VARV either 1,400 or 6,300 YBP, and then further diverged into two subclades at least 800 YBP. All of these analyses indicate that the divergence of alastrim and variola major occurred earlier than previously believed.

evolution | SNPs | variola virus

Observations of smallpox-typical skin rashes on Egyptian mummies dating from 1100 to 1580 B.C. (1–3) gave early credibility to theories that ancient Egypt was an early (and perhaps the earliest) smallpox endemic region. However, smallpox researchers such as Dixon (1) noted that "The most striking thing about smallpox is its absence from the books of the Old and New Testaments, and also from the literature of the Greeks and Romans. Such a serious disease as *variola* major is very unlikely to have escaped a description by Hippocrates if it existed." Historical records from Asia describe evidence of smallpox-like disease in medical writings from ancient China (1122 B.C.) and India (as early as 1500 B.C.) (1, 3). The earliest unmistakable description of smallpox first appears in the 4th century A.D. in China, the 7th century A.D. in India and the Mediterranean, and the 10th century A.D. in southwestern Asia (2). These latter texts describe a well known local/endemic disease. These earlier Asian descriptions could indicate that pandemic smallpox originated in East Asia.

Declared eradicated in 1980, smallpox was an acute, self-limited human illness with no known non-human reservoir. No latent state existed, and illness outcome was either death or recovery with life-long immunity. Clinically, the severe form of smallpox [variola virus (VARV) major] had a long descriptive history in the Old World medical literature, with symptoms including pustular rash, severe morbidity, and significant case fatality rates (CFRs) (up to 30%) in unvaccinated populations. There are no credible descriptions of VARV major-like disease on the American continents or in sub-Saharan Africa before the westward exploration in the 15th century A.D., which exported the virulent smallpox to aboriginal populations in those regions. During the 19th century, mild outbreaks of smallpox disease with low CFRs (<1%) were described in both Central America and South Africa. These less virulent viruses found in Central America had distinct biological properties and were named alastrim minor (2). The puzzle of smallpox evolution is ambiguous because there are significant gaps in historic medical records and competing hypotheses regarding an ancient vs. more recent origin of various forms of smallpox, such as alastrim minor. To gain better understanding of smallpox evolution and address the discrepancies surrounding the origins of the various types of smallpox, we carried out systematic analyses of SNPs of VARV isolates and applied data-processing algorithms to build phylograms. Additionally, we combined historic records and observations from the SNP-based phylograms to place the puzzle pieces and thereby gain a better understanding of the origin and subsequent evolution of smallpox variants as human pathogens.

Results and Discussion

VARV Genome Stability and Phylogenetic Analysis. The availability of genome sequences from a collection of VARV isolates with a broad geographic distribution (Fig. 1) was used for the evolutionary study of VARV (4). VARV, a member of the genus *Orthopoxvirus*, contains a single linear double-stranded DNA genome of 186 kb that encodes most enzymes for its propagation (5). The VARV isolates studied demonstrated a low mutation rate because epidemiologically linked isolates showed little or no sequence changes in samples with collection times up to 1 year apart [supporting information (SI) Table 3]. This low mutation rate can be clearly seen in our genetically identical samples from Bangladesh collected between 1974 and 1975. These samples certainly represent long complex transmission chains from one of the most densely populated regions in the world but have identical or nearly identical nucleotide sequences. The related Orthopoxvirus monkeypox virus similarly shows single or no nucleotide changes among isolates from three different hosts in a recent United States outbreak (6). Within and across species of orthopoxviruses, the genomes possess nucleotide identities of >99.6% among VARV isolates and 98% among VARV and congeneric taxa such as taterapox virus (TATV) (4) (West Africa) and camelpox virus (CMLV) (7) (Central Asia) (Table 1), two species with the highest sequence similarity to VARV (4). The large size and slow nucleotide substitution rate allow the VARV genomic DNA to retain phylogenetically informative mutations while lowering the incidence of homoplastic

Author contributions: Y.L., S.N.G., and I.K.D. designed research; Y.L., D.S.C., S.N.G., M.C.W., and E.A.V. performed research; Y.L., D.S.C., and S.N.G. contributed new reagents/analytic tools; Y.L., D.S.C., and I.K.D. analyzed data; and Y.L., D.S.C., and I.K.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: CFR, case fatality rate; CMLV, camelpox virus; cSNP, concatenated SNP; P-I, primary clade I; P-II, primary clade II; TATV, taterapox virus; VARV, variola virus; YBP, years before present.

[§]Present address: Department of Population Health Sciences, University of Wisconsin, Madison, WI 53726.

[¶]To whom correspondence should be addressed. E-mail: idamon@cdc.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0609268104/DC1.

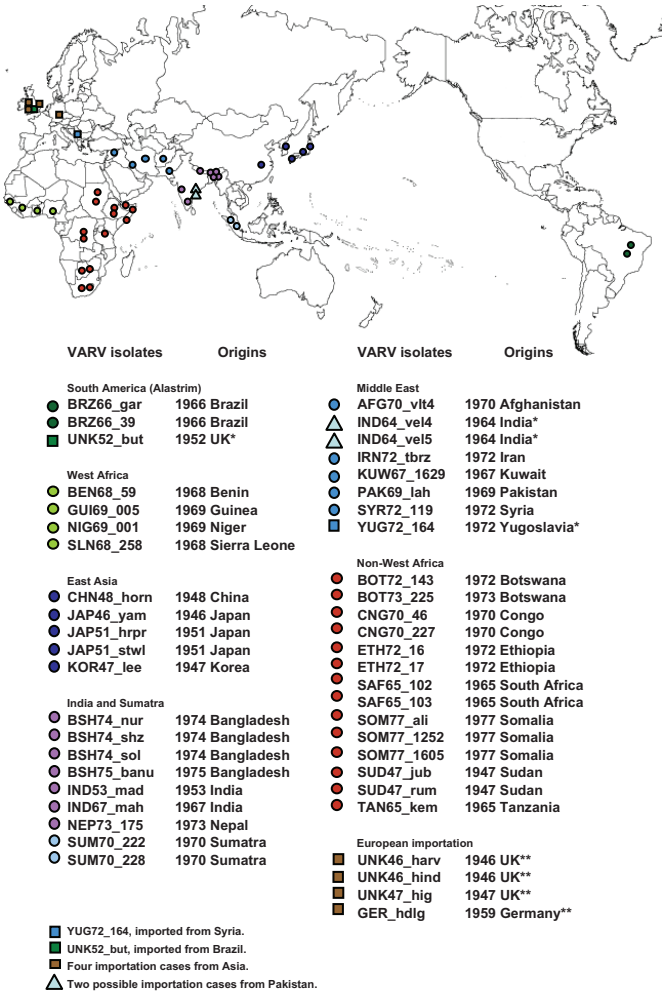

Fig. 1. Geographic location of VARV isolates. The VARV genome sequences used in this study are from ref. 4 and GenBank (accession nos. DQ 441438, DQ 441439, and DQ 441441); the isolates are derived from geographically distinct smallpox outbreaks between the mid-1940s and 1977. Outbreak locations are grouped and color keyed by regions; indigenous country cases are labeled as circles, and imported cases are labeled as squares. All the European isolates are imported cases from endemic regions, of which two had known origins (YUG72_164 originated in the Middle East, and UNK52_but originated in South America). The high similarity between IND64_vel4 and IND64_vel5 (4) to Central Asia/Middle East isolates suggests that the two India 1964 VARV isolates probably were imported cases that were isolated from a Christian Mission Hospital in Vellore, India (41).

mutations (i.e., fewer character-state reversals). The diversity of 47 recently sequenced VARV genomes was largely associated with the geographic origin of the isolates (ref. 4 and this study), and mutations had minimal association with temporal variation across isolates. These observations demonstrate that there was not a single pandemic strain (*SI Text*, *Note 1*) but rather that the genetic clades of these analyses are representative of old and perhaps ancient locally endemic strains.

We developed a unique method of using concatenated SNP (cSNP) matrices for VARV phylogenetic analysis. The cSNPs, defined as a polymorphic nucleotide flanked by a fixed number of conserved nucleotides, provided a less complex representation of the whole genomic sequence by excluding insertion/deletions (indels) (*SI Text*, *Note 2*), and the topology generated in this study (Fig. 2) agreed with the analyses of the viral genome in ref. 4. SNPs constitute ≈89% of the total genomic sequence diversity among the most diverse VARV isolates (Table 1). The cSNPs of VARV isolates alone and VARV isolates combined with CMLV and TATV captured ≈79% and 42% of the total SNP diversity of VARV, respectively (Table 1). Phylogenetic analyses of cSNPs within VARV and among VARV, CMLV, and TATV resulted in a cladogram with a very low homoplasy index (0.039) (see *Methods*) and well supported clades (Fig. 2*B*). The low incidence of homoplastic events indicated that the resultant VARV and VARV–CMLV–TATV phylogenies were not likely influenced by convergent evolution (8, 9). The VARV cSNPs, using different numbers of conserved flanking nucleotides (*SI Text*, *Note 3*), generated different numbers of SNPs but resulted in congruent topologies (Fig. 2 and SI Fig. 4). The low homoplasy indices (9, 10) suggest that DNA recombination was not likely influencing the topology generated by these analyses, although recombination occurs in laboratory conditions (11).

Major Variola Clades and the Origins of African minor and Alastrim minor Strains. Our rooted analyses of VARV resulted in the description of two primary clades that diverged from a common ancestral VARV (Fig. 2). Primary clade I (P-I) comprises Asian VARV major and East, Central, and South African VARV isolates. Asian VARV was traditionally known as VARV major with a CFR of 20–30% (2). The non-West African (African) isolates examined in this study had a 30-year temporal representation from South, Central, and Eastern Africa and were collected from regions with both low (1%) and high (10–15%) CFR smallpox (4), including the last naturally occurring smallpox case in Somalia (African minor). Excluding isolates with poorly characterized origins of importation into Europe (Fig. 2), the South, Central, and East African (African) isolates formed a related monophyletic cluster with isolates of VARV major from Bangladesh. This analysis suggests that African minor is likely a variant of Asia VARV major, which is consistent with their similar biological properties (12). Thus, the traditional nomenclature separating smallpox into strains of variola major and variola minor based exclusively on CFRs is not supported by our phylogeny.

VARV primary clade II (P-II) comprises two subclades. One encompasses the biologically distinctive South American isolates traditionally known as alastrim minor, and the other is composed of West African isolates. Limited historical medical documents indicate that alastrim was described by indigenous people in central Brazil as a disease with pustular rash similar to smallpox but with a markedly low fatality rate (13–15). The first published description of this clinically distinct mild type of smallpox was in the Caribbean region by Anderson in the 1860s (16), and the disease was later exported to Florida and subsequently elsewhere in North America (17). During the smallpox eradication program, the CFR of smallpox in West Africa was described to be intermediate (≈10%) between variola major and alastrim based on age-adjusted CFRs, and the isolates in this region were described with intermediate biological characteristics with respect to the alastrim minor and VARV major isolates (12, 18). Our analyses show that West Africa VARV shared a relatively recent common ancestor with alastrim minor (compared with its relationship to P-I) (Fig. 2).

Recent or Ancient Origin of VARV in This Study. We calculated estimates of evolutionary rates and time to most recent common ancestry on multiple evolutionary clocks. Initially, the year of isolation for each isolate was used under both strict and relaxed molecular clocks as in Bird 2007 (19, 20) (Table 2). These analyses result in a prediction that the time to the most recent common ancestor (t_{MRCA}) of the two primary VARV clades is 207/231 years before present (YBP), and the radiation of P-I VARV major is 83/92 YBP, using strict/relaxed clocks, respectively (Table 2 and SI Datasets 1 and 2), which implies a very recent origin for the VARV isolates in this study. Both of these calculations would suggest that the VARV major isolates are from a recent pandemic sweep from

Table 1. Sequence similarity and SNPs representative of selected VARV and orthopoxviruses

Isolates compared*	Genome similarity†	All (genome) SNPs (% of mutations)†	VARV cSNPs (% of all SNPs)	VARV-TATV-CMLV cSNPs (% of all SNPs)	Genome indels (% of mutations)†
CHN48_horn vs. SOM77_ali	0.9967	204 (84%)	164 (80%)	91 (45%)	40 (16%)
BRZ66_39 vs. CHN48_horn	0.9965	566 (88%)	436 (77%)	244 (43%)	79 (12%)
BRZ66_39 vs. SOM77_ali	0.9962	642 (90%)	496 (77%)	281 (44%)	75 (10%)
BEN68_59 vs. TATV_dah68	0.981	3,108 (88%)	nd	1,299 (42%)	424 (12%)
IRN72_tbrz vs. CMLV_cms70	0.98	3,442 (91%)	nd	1,542 (46%)	342 (9%)
CNG70_46 vs. MPXV_zai79	0.957	7,459 (93%)	nd	nd	617 (7%)

nd, not available.

*The number of SNPs, indels, VARV cSNPs, and VARV–TATV–CMLV cSNPs from ref. 4 and SI Figs. 8 and 9 and SI Table 5 . VARV genomes are ≈186 kb, and MPXV and TATV, CMLV are ≈200 kb.

†Indels were counted as a fifth character state.

East Asia (the basal subclade), which replaced all original local endemic VARV strains within the last ≈90 years. However, this contradicts the geographically (as opposed to temporal) associated topology of VARV isolates in this study. Additionally, VARV of South Africa (African minor), which was endemic in that region in the late 19th century, was first officially reported in 1904 (21, 22), and our isolates were obtained in 1965. These historical records would not agree with a 90-year (very recent) diversification of the isolates used in our phylogeny. Additionally, this analysis estimates the divergence of VARV from the rodent isolates (TATV) of West Africa at ≈1,374 (relaxed clock)/1,899 (strict clock) YBP, which is the approximate time that endemic smallpox was initially recorded in East Asia. The large geographic separation of Asian VARV and the West African enzootic TATV (occurring in a West African naked-soled gerbil) would not support this timeline. The slowly evolving VARV ancestor would likely need more time to become

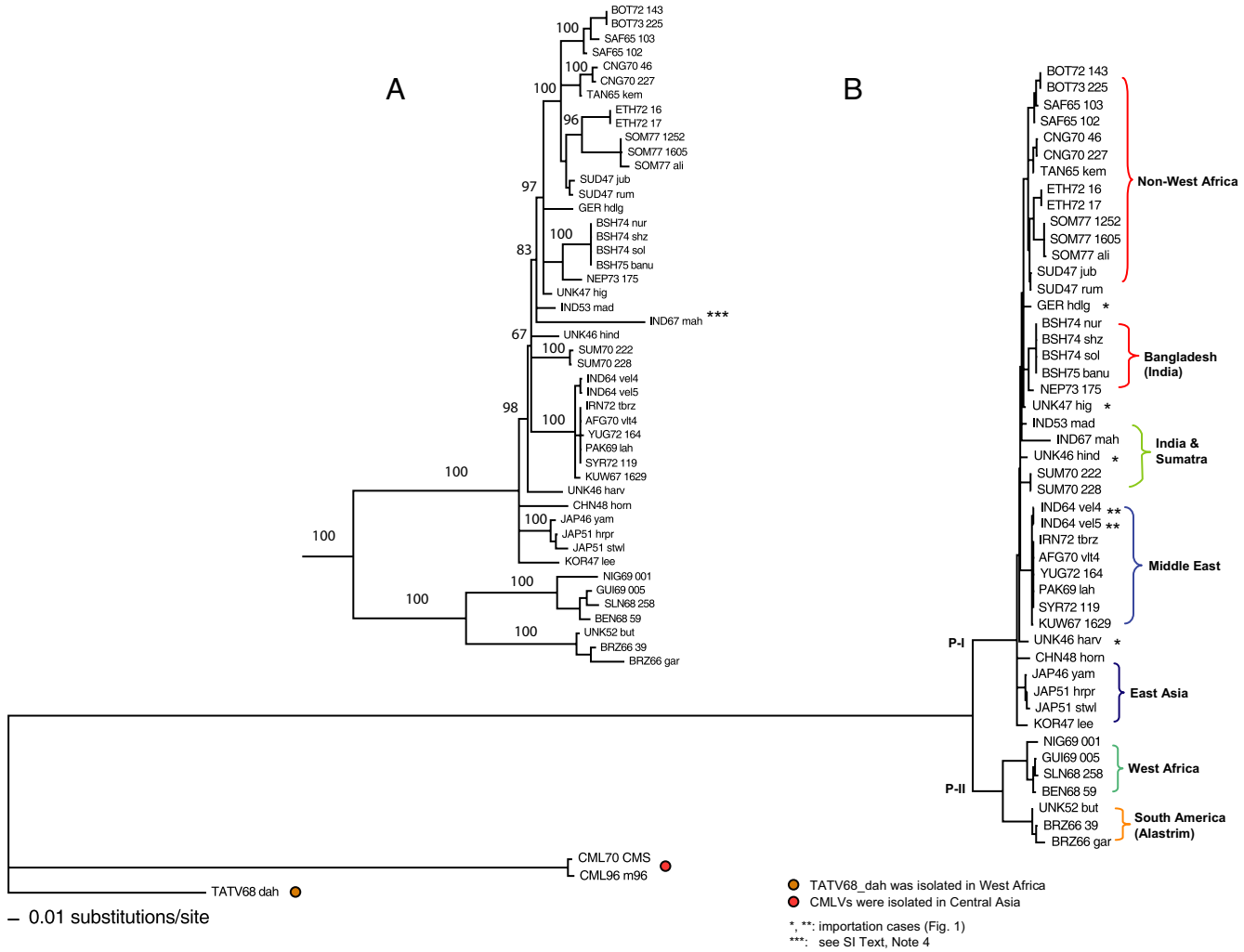

Fig. 2. Tree topologies generated by using maximum likelihood analysis of VARV plus CMLV and TATV SNP matrices. (*A*) The VARV portion of rooted VARV phylogram was expanded horizontally to display the details of VARV subclades and Bootstrap values. (*B*) The VARV phylogram generated from concatenated VARV plus CMLV and TATV SNP matrices, using TATV as an outgroup. The major clades and subclades were designated by name. The consistency index (CI) is 0.97, and the retention index (RI) is 0.98.

Table 2. Coalescent analyses-posterior probability densities of VARV, TATV, and CMLV

Data set*	Parameters		Calibration node (95%HPD)‡			
	Clock model	Calibration†	VARV P-I radiation	Alastrim/West Africa	VARV P-I/P-II	VARV–TATV
1. VARV–TATV–CMLV	Strict	Isolation date†	83 (74, 93)	126 (103, 150)	207 (170, 246)	1,899 (1,513, 2,318)
2. VARV–TATV–CMLV	Relax	Isolation date†	92 (73, 118)	133 (75, 203)	231 (135, 360)	1,374 (657, 2,242)
3. VARV–TATV–CMLV	Strict	African clade§	373 (331, 422)	778 (634, 926)	1,425 (1,218, 1,648)	$1.55\ (1.34, 1.75) \times 10^4$
4. VARV–TATV–CMLV	Strict	East Asia clade¶	1,600¶ (1,598, 1,602)	3,502 (2,845, 4,198)	6,266 (5,330, 7,204)	$6.82\ (5.85, 7.86) \times 10^4$

HPD, highest posterior density.

*The VARV–TATV–CMLV cSNP matrix includes 2,436 SNP sites (768 from VARV).

†The isolation date of each isolate was used as the date in the BEAST calculation.

‡The mean and 95% HPD intervals (in years) of the posterior probability distribution for the nodes TMRCA.

§The calculation was calibrated by setting the prior of African VARV clades (excluding West Africa) at 250 years based on smallpox epidemic history.

¶The calculation was calibrated by setting the prior of VARV P-I radiation, starting at 1,600 years ago (earliest smallpox historic records)

an efficient and exclusively human pathogen; therefore, the above results based on the entire set of isolation dates most likely underestimate the evolutionary history of VARV.

The historical smallpox records provide two calibration points for the coalescent analyses. The radiation of VARV P-I appears to begin in East Asia (basal node), which is congruent with the location of the earliest recorded smallpox cases. The basal position of the Asian isolates is supported by the presence of 42 pleisomorphic (maintained from an ancestral virus) SNPs between these isolates and members of P-II (SI Fig. 5). If the radiation of P-I represented the gradual establishment of smallpox endemics within the Old World following the smallpox historical records, then the VARV P-I radiation began as early as 1,600 YBP (400 A.D.) as described in Fenner *et al.*'s discussion (2) of the earliest (Chinese) descriptions of smallpox. Using this date as a "prior" in the Bayesian Evolutionary Analysis Sampling Trees (BEAST) software (http://beast.bio.ed.ac.uk) analyses under a strict molecular clock yields a divergence date of ≈6,300 YBP for P-I/P-II divergence (Table 2 and SI Dataset 3). These results support an ancient origin of VARV and would suggest that smallpox endemics started in East Asia and spread through the Middle East and India and then to Africa. This hypothesis agrees with both the resultant tree topology and the early smallpox history in East Asia and India and historic records that describe the importations of smallpox to South Africa in 1713, when highly virulent VARV strains were imported from the Indian continent (23) and later spread to Central Africa (24). As an alternate possibility, we used the smallpox historic records of South Africa as a calibration point, ≈250 YBP, as the "prior" divergence date (BEAST) for the South, Central, and East African (African) isolates subclade. The entire topology yields a divergence date of P-I/P-II at ≈1,400 (Table 2 and SI Dataset 4) to 2,300 YBP (SI Table 4).

If the P-I radiation represents the smallpox endemic history in the Old World, it explains the historic puzzle regarding the absence of a description of smallpox in the literature of the ancient Greek and Roman civilizations (1). It is probable that the smallpox epidemics in this region occurred at a time later than those in ancient China and India. The "Athens plague" in 430 B.C. described by Thucydides, once considered to be smallpox (3), would contradict this theory, but it has recently been determined to be typhoid fever (25). During the reign of Ramses V, Egypt was in a civil war and was attacked by enemies from the north (3); if the pustular eruption of Ramses V was from smallpox, it could represent a smallpox outbreak from imported cases because of war rather than regional endemic disease. This hypothesis is supported by the fact that only three mummies in that period had similar lesions (2).

Diversification of Alastrim minor. The smallpox history/VARV topology-based analyses show that the divergence of alastrim minor and West African variola began at least 800 years ago (Table 2 and SI Table 4), predating the previous hypothesis that this divergence coincides with the beginning of the slave trade. Under the more ancient origin hypothesis, this divergence could have occurred after a speciation event in either the New World (based on the calibration from the earliest smallpox historic records; Fig. 3*A*) or West Africa (based on the calibration from the South African smallpox historic records; Fig. 3*B*). These scenarios would suggest that there were unidentified variola minor/alastrim isolates that existed in either Africa or the New World much earlier than the discovery of alastrim minor (P-II); thus, this taxon has a longer history than has been described.

Although the well documented importation of VARV major into the Americas had a devastating effect on many of the native populations (2), the records of rapid disease spread with great mortality might, in part, be attributable to the combined effect of the high population density, lack of basic care for victims, and potential coinfections with other Old World diseases (26, 27). Historically, there are few records of smallpox in West Africa, and Henige (28) argued that the first VARV major smallpox outbreak in the New World did not originate from West Africa, which would refute the hypothesis of an early smallpox history in West Africa. It is possible that after an ancient introduction, alastrim evolved independently in isolated regions within the large geographic area of the New World. In fact, Phelan's study showed that the native populations of the Ecuador Highlands did not decrease during the initial Spanish rule (attributed to have introduced VARV major to the New World) (29), suggesting that these populations could have been more resistant to smallpox, perhaps by previous contact with alastrim. In Trinidad, a mild smallpox outbreak in 1902 was documented to have originated from this region (30). If this diversification occurred in the New World, then the primary isolation event would likely have resulted from the movement of an ancestral VARV with early humans into that region (Fig. 4*A*). The absence of SNP signatures suggestive of mixing between P-I and P-II (SI Figs. 6 and 7), the consistent and significantly different clinical characteristics of alastrim from traditional VARV major of P-I (31), could be explained by such a significant geographical separation.

If the split of P-I/P-II occurred within the Old World, P-II likely originated in West Africa. The two P-II subclades would then have diverged ≈800–1,100 YBP (Table 2 and SI Table 4) on the African continent. The geographic distributions of these events are difficult to determine because of a lack of written history in these regions. However, the presence and diversification of other closely related Orthopoxviruses such as monkeypox and TATV could support an African origin for ancestral VARVs. If the isolation and diversification of P-II occurred in Africa, then, based on our phylogeny, there is an undescribed African representative from the P-II subclade, which was introduced into the New World perhaps hundreds of years after the diversification event, possibly during the time of the slave trade between countries in the Old and New World (Fig. 3*B*).

An inadequate understanding of the history of these cultures and the relatively milder clinical symptoms (with respect to variola

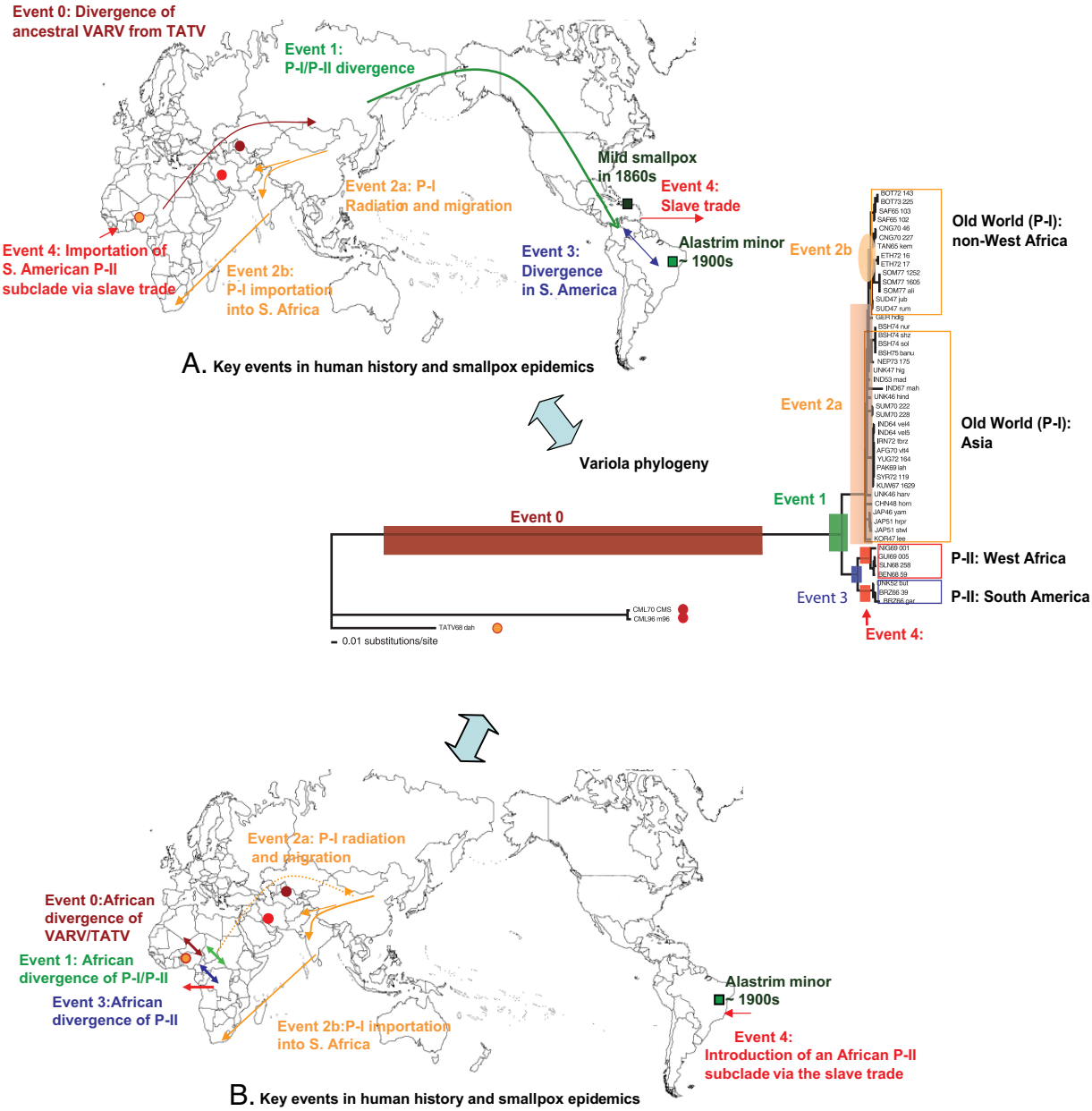

Fig. 3. Hypotheses of the origin of VARV. Event 0, The high similarity of VARV to CMLV and TATV genome sequences suggests that TATV and CMLV share a more recent common ancestor with VARV than do other known *Orthopoxvirus* species. Event 1, hypothesis A: An ancestral VARV diverged into two primary clades (P-I and P-II) and evolved independently in the Old World and the New World by following its human hosts. Event 1, hypothesis B: The divergence of P-I/P-II in Africa. Event 2, hypotheses A and B: The long single branch and subsequent radiation of P-I suggest that the ancestral VARV of P-I probably originated in Northeast Asia and that the smallpox endemics of P-I started to spread to East Asia, the Middle East, and India as local populations became sufficiently large. Event 2a, The diversification and migration of P-I throughout Asia. Event 2b, Western exploration brought VARV major to South Africa from the Indian subcontinent, which then spread northward to Central Africa and the Horn of Africa (Somali Peninsula). Events 3 and 4, hypothesis A: The diversification of P-II in the New World followed by its reintroduction into Africa; hypothesis B: The diversification of P-II in Africa, followed by the later introduction of one subclade into the New World.

major) contribute to the absence of a more extensive historic record of smallpox disease(s) in either the New World or sub-Saharan Africa. The historical human population densities in both of these regions were lower than the regions with endemic variola major. In cases of milder disease, individuals would have been more mobile when infectious and would have had a greater number of contacts and thus perpetuated the disease's spread. In support of this hypothesis, the mild smallpox outbreaks in Somalia persisted for long periods among small nomadic bands and showed a significantly different transmission pattern compared with that of VARV major (2). Although it is unknown whether the practice of variolation/inoculation was used by the indigenous peoples of Africa or the Americas, the longevity of VARV within convalescent scabs (up to a decade) (32–34) would provide another mechanism for the repeated introduction of a smallpox strain within a geographic community either by inadvertent contact with scab materials or directly through variolation practices.

Origin of VARV. Possible clues regarding the adaptation of VARV to humans can be found in the close relationship between VARV and TATV/CMLV (4). TATV is associated with a terrestrial rodent native to West Africa (35). Our coalescent analyses indicate that the

divergence between VARV and TATV occurred from 16,000 YBP (Table 2 and SI Dataset 4, based on the smallpox historic records of South Africa) to 68,000 YBP (Table 2 and SI Dataset 3, based on the earliest recorded smallpox history in East Asia). Thus, like the related zoonotic orthopoxviruses with rodent reservoirs (6, 36, 37), VARV may have evolved from an enzootic pathogen of African rodents and subsequently spread out of Africa.

The genomic sequences of VARV isolates selected for this study were taken from the World Health Organization Collaborating Center repository and, although they display a low level of sequence divergence, are representative of the greatest geographic and temporal diversity within the collection. The cSNP-derived phylogeny of these isolates produces a topology consistent with historical human migratory patterns, population expansion, and documented smallpox epidemics. The depicted topology is supportive of a hypothesis (Fig. 3) that two clades of variola diverged from an ancestral virus and subsequently evolved in geographically discrete human populations. Although the reason for the different human pathogenesis of alastrim and VARV major is unclear, the genetic structure of alastrim minor and its topology with the other VARV isolates add critical pieces to the smallpox evolutionary puzzle. This proposed phylogeny of VARV extends the evolutionary time of smallpox relative to that of previous theories.

Methods

The genomic sequences were aligned with multiple genome alignment methods (38), and the alignment was screened for SNPs by using software described in ref. 39. The SNP bases in a given genome were concatenated and compiled into an SNP matrix, so the number of rows is the same as the number of genomes. Each column in the matrix represents a homologous SNP position. The SNP matrices in this study were formed by concatenating the SNPs that meet the requirement of seven nucleotides surrounding and conserved on both sides of the SNP (39). The resultant VARV SNP matrix contains 1,347 SNPs (SI Fig. 8). These represent ≈79% of total genomic SNPs (Table 1) (4), because some of the clustered SNPs were omitted as a result of the requirement for sequence conservation surrounding the SNP position. An SNP matrix of VARV, CMLV, and TATV was generated by adding CMLV and TATV genomic sequences to the VARV sequence alignment. Using the same criteria, the VARV–TATV–CMLV SNP matrix included 2,436 SNPs (SI Fig. 9), which contained ≈42% of their total SNPs (Table 1) (4). The requirement that SNPs be flanked by a string of conserved nucleotides has the disadvantage that some SNPs with surrounding sequence variation are ignored. The advantage, however, is that an SNP matrix can be found even in cases where attempts to generate a multiple sequence alignment fail, for example with large data sets or when there are genome rearrangements. In such situations, the software (40) can locate SNPs based on a consensus sequence that could be generated without an actual alignment, for example, by using BLAST relative to a reference sequence. This was required in other SNP analyses (data not shown) across all *Orthopoxvirus* genomes.

Both data sets were analyzed by using maximum likelihood methods (Fig. 2 and *SI Text*, *Note 5*). For the maximum likelihood analyses, MODELTEST (40) was used to evaluate 56 nucleotide substitution models for our data matrix. The results showed that the most appropriate model was the transversion model, which assumes variable base frequencies, variable transversions, and equal transitions across the data set. These assumptions were used in all subsequent likelihood analyses, including that depicted in Fig. 2. Bootstrap values were calculated to indicate support for each node (Fig. 2*B*) (8) (PAUP*). The homoplasy, consistency, and retention indices were calculated by using PAUP* software, Version 4.0b10. The VARV SNP profiles for major VARV subclades (SI Figs. 6 and 7) were generated manually by stratifying the VARV cSNP matrix. Nexus files were imported into the Bayesian Evolutionary Analysis Utility (BEAUti), and coalescent analyses of the converted cSNPs matrix were conducted by using BEAST software to determine the t_{MRCA} for VARV primary and major subclades. All divergence estimates were calculated by using BEAST analyses as described in refs. 19 and 20. Three sets of input dates were used. The first analysis was conducted by using the dates of isolation for each isolate in the phylogeny and both relaxed and strict molecular clocks. The second analysis was conducted by setting the "prior" of the African VARV subclade (excluding West Africa isolates) at 250 YBP. The third analysis used the earliest historical records for the root height (prior) of P-I (1,600 YBP) and, similarly to the previous analysis, used a strict clock across the entire topology based on the evolutionary rate obtained for P-I. The BEAST parameters for each node are listed in Table 2. For each analysis, the output was visualized by using Tracer software, Version 1.3 (http://beast.bio.ed.ac.uk).

We thank J. Esposito, M. Reynolds, R. Regnery, S. Reeder, R. Biek, V. Olson, S. Smith, S. Nichol, B. Bird, and K. Karem for comments on the manuscript; S. Sammons and V. Olson helped with sequence alignments; and T. Slezak for contributing to discussions about the bioinformatics analyses. We thank the Centers for Disease Control and Prevention (CDC) Biotech center for providing variola genomic sequences and the CDC Information Center for assistance in searching smallpox literature. This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

1. Dixon CW (1962) in *Smallpox* (Churchill, London).
2. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID (1988) in *Smallpox and Its Eradication*, ed Frank Fenner (WHO, Geneva).
3. Hopkins DR (1983) in *The Greatest Killer – Smallpox in History* (Univ of Chicago Press, Chicago).
4. Esposito JJ, Sammons SA, Frace AM, Osborne JD, Olsen-Rasmussen M, Zhang M, Govil D, Damon IK, Kline R, Laker M, *et al.* (2006) *Science* 313:807–812.
5. Moss B (2000) in *Fields Virology* (Lippincott Williams & Wilkins, Philadelphia), pp 2849–2883.
6. Likos AM, Sammons SA, Olson VA, Frace AM, Li Y, Olsen-Rasmussen M, Davidson W, Galloway R, Khristova ML, Reynolds MG, *et al.* (2005) *J Gen Virol* 86:2661–2672.
7. Gubser C, Hue S, Kellam P, Smith GL (2004) *J Gen Virol* 85:105–117.
8. Swofford DL (2002) *PAUP* Phylogenetic Analysis Using Parsimony (*And Other Methods), Version 4.0b10* (Sinauer, Sunderland, MA).
9. Maynard SJ, Smith NH (1998) *Mol Biol Evol* 15:590–599.
10. Farris JS (1969) *Syst Zool* 18:374–385.
11. Dumbell KR, Bedson HS (1964) *J Hyg* 62:133–140.
12. Dumbell KR, Huq F (1986) *Am J Epidemiol* 123:403–415.
13. Downie AW, Dumbell KR, Ayroza Galvao PA, Zatz I (1963) *Trop Geogr Med* 15:25–28.
14. Jorge R (1924) *Lancet* 204:1317–1322.
15. Ribas E (1911) *Trans R Soc Trop Med Hyg* 4:224–232.
16. Anderson I (1866) *Trans Epidemiol Soc Lond* 2:414–418.
17. Chapin CV, Smith J (1932) *J Prev Med* 6:273–320.
18. Shafa E (1972) *WHO Surveillance Epidemiol* 72.35.
19. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST (2007) *J Virol* 81:2805–2816.
20. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) *PLoS Biol* 4:e88.
21. De Korte WE (1904) *Lancet* 163:1273–1276.
22. De Korte WE (1923) *Lancet* 201:101.
23. Burrows EH (1958) in *A History of Medicine In South Africa-Up to the End of the Nineteenth Century* (AA Balkema, Cape Town, South Africa), pp 64–65.
24. Roberts CJ (1967) *Cent Afr J Med* 13:31–33.
25. Papagrigorakis MJ, Yapijakis C, Synodinos PN, Baziotopoulou-Valavani E (2006) *Int J Infect Dis* 10:206–214.
26. Smith AB (1989) *S Afr Med J* 75:25–26.
27. Stannard DE (1993) in *The Cambridge World History of Human Disease*, ed Kiple KF (Cambridge Univ Press, Cambridge, UK), pp 35–44.
28. Henige D (1986) in *Africans in Bondage: Studies in Slavery and the Slave Trade: Essays in Honor of Philip D. Curtin on the Occasion of the Twenty-Fifth Anniversary of African Studies at the University of Wisconsin*, ed Lovejoy PE (Univ of Wisconsin Press, Madison, WI), pp 1–26.
29. Phelan JL (1967) in *The Kingdom of Quito in The Seventeenth Century–Bureaucratic Politics in the Spanish Empire* (Univ of Wisconsin Press, Madison, WI), pp 43–65.
30. Dickson JR, Lassalle CF (1903) *Br Med J* ii:711–713.
31. Jorge R (1924) *Lancet* 204:1366–1370.
32. Walther BA, Ewald PW (2004) *Biol Rev Camb Philos Soc* 79:849–869.
33. Wolff HL, Croon JJ (1968) *Bull WHO* 38:492–493.
34. Downie AW, Dumbell KR (1947) *Lancet* 249:550–553.
35. Lourie B, Nakano JH, Kemp GE, Setzer HW (1975) *J Infect Dis* 132:677–681.
36. Fonseca FG, Lanna MCS, Campos MAS, Kitajima EW, Peres JN, Golgher RR, Ferreira PCP, Kroon EG (1998) *Arch Virol* 143:1171–1186.
37. Meyer H, Schay C, Mahnel H, Pfeffer M (1999) *Arch Virol* 144:491–501.
38. Hohl M, Kurtz S, Ohlebusch E (2002) *Bioinformatics* 18(Suppl 1):S312–S320.
39. Gardner SN, Wagner MC (2005) *BMC Genomics* 6:73.
40. Posada D, Crandall KA (1998) *Bioinformatics* 14:817–818.
41. Dumbell KR, Harper L, Buchan A, Douglass NJ, Bedson HS (1999) *Epidemiol Infect* 122:287–290.