Title: Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age

Authors: Barbara Mühlemann[1,2,†], Lasse Vinner[3, †], Ashot Margaryan[3,4], Helene Wilhelmson[5,6], Constanza de la Fuente Castro[7], Morten E. Allentoft[3,8], Peter de Barros Damgaard[3], Anders Johannes Hansen[3], Sofie Holtsmark Nielsen[3], Lisa Mariann Strand[9], Jan Bill[10], Alexandra Buzhilova[11], Tamara Pushkina[12], Ceri Falys[13], Valeri Khartanovich[14], Vyacheslav Moiseyev[14], Marie Louise Schjellerup Jørkov[15], Palle Østergaard Sørensen[16], Yvonne Magnusson[17], Ingrid Gustin[5], Hannes Schroeder[18], Gerd Sutter[19], Geoffrey L. Smith[20], Christian Drosten[2], Ron A. M. Fouchier[21], Derek J. Smith[1], Eske Willerslev[3,22,23,24*], Terry C. Jones[1,2,*], Martin Sikora[3,*]

Affiliations

[1]Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Downing St., Cambridge, CB2 3EJ, U.K.

[2]Institute of Virology, Charité, Universitätsmedizin Berlin and Deutsches Zentrum für Infektionsforschung, Associated partner site Berlin, Charitéplatz 1, 10117 Berlin, Germany.

[3]Lundbeck Foundation GeoGenetics Center, Globe Institute, University of Copenhagen, Øster Voldgade 5–7, 1350, Copenhagen, Denmark.

[4]Institute of Molecular Biology, National Academy of Sciences, 7 Hasratian St., 0014, Yerevan, Armenia.

[5]Department of Archaeology and Ancient History, Lund University, Pb 192, Lund 221 00, SE, Sweden.

[6]Sydsvensk arkeologi AB, Pb 134, Kristianstad 291 22, SE, Sweden.

[7]Department of Human Genetics, University of Chicago, 920 E 58th St., Chicago, IL 60637, USA.

[8]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Kent Street, 6102 Perth, Australia.

[9]Department of Archaeology and Cultural History, Norwegian University of Science and Technology University Museum, 7491 Trondheim, Norway.

[10]Museum of Cultural History, University of Oslo, P.O. Box 6762 St. Olavs plass, 0130 Oslo,
30 Norway.

[11]Research Institute and Museum of Anthropology, Lomonosov Moscow State University, Mokhovaya St. 11, Moscow 125009, Russian Federation.

[12]Department of Archaeology, Faculty of History, Lomonosov Moscow State University, Moscow 119992, GSP-1, Lomonosovsky Prospekt, 27–4, Russian Federation.

35 [13]Thames Valley Archaeological Services, 47–49 de Beauvoir Rd, Reading, RG1 5NR, U.K.

[14]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, 199034 St. Petersburg, Russian Federation.

[15]Laboratory of Biological Anthropology, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Frederik V's vej 11, DK-2100 Copenhagen, Denmark.

40 [16]Roskilde Museum, Frederikssund Museum, Færgelundsvej 1, 3630 Jægerspris, Denmark.

[17]Malmö Museum, Pb-406, 201 24 Malmö, Sweden.

[18]Section for Evolutionary Genomics, The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Øster Farimagsgade 5, 1353, Copenhagen, Denmark.

[19]Institute for Infectious Diseases and Zoonoses, LMU University of Munich, and Deutsches
45 Zentrum für Infektionsforschung, partner site Munich, Veterinärstr. 13, 80539 Munich, Germany.

[20]Department of Pathology, University of Cambridge, Tennis Court Rd., Cambridge, CB2 1QP, U.K.

[21]Department of Viroscience, Erasmus Medical Centre, Wytemaweg 80, Rotterdam, 3015 CN,
50 Netherlands.

[22]Lundbeck Foundation GeoGenetics Center, Department of Zoology, Downing St., University of Cambridge, Downing St., Cambridge, CB2 3EJ, U.K.

[23]Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, U.K.

[24]Danish Institute for Advanced Study, University of Southern Denmark, Campusvej 55, 5230
55 Odense M, Denmark.

* Correspondence to: ewillerslev@sund.ku.dk, tcj25@cam.ac.uk, martin.sikora@sund.ku.dk

† Contributed equally.

Abstract

Smallpox, one of the most devastating human diseases, killed 300–500 million people in the 20th
60 century alone. We recovered viral sequences from thirteen northern European individuals, including eleven from ~600–1050 Common Era, overlapping the Viking Age, and reconstructed near-complete variola virus genomes for four of them. The samples pre-date the earliest confirmed smallpox cases by ~1000 years and the sequences reveal a now-extinct sister clade to the modern variola viruses in circulation prior to the eradication of smallpox. We date the most
65 recent common ancestor of variola virus to ~1700 years ago. Distinct patterns of gene inactivation in the four near-complete sequences show that different evolutionary paths of genotypic host adaptation resulted in variola viruses that circulated widely among humans.

One Sentence Summary

Ancient human DNA sequences reveal a new chapter in the history of smallpox.

70 **Main Text**

Variola virus, the causative agent of smallpox, is estimated to have killed 300–500 million people in the 20th century and was responsible for widespread mortality and terror for at least several preceding centuries (*1*) (Supplementary Text). Smallpox drove the 1796 development of vaccination by Edward Jenner, who later declared that "*it now becomes too manifest to admit of*
75 *controversy, that the annihilation of the Small-pox, the most dreadful scourge of the human species, must be the result of this practice*" (*2*). His prophecy was fulfilled in 1980, when the disease was declared eradicated by the World Health Organisation, following a world-wide vaccination campaign. The virus is now exclusively stored in laboratories in the United States and Russia. Despite the eradication of smallpox, there are ongoing concerns regarding the re-
80 emergence of a smallpox-like disease via accidental or deliberate re-introduction of variola virus, adaptation of monkeypox virus to humans, or zoonosis or genetic engineering of another

orthopoxvirus (*3–5*). Specific information regarding the evolutionary history of variola virus is therefore of substantial interest.

85 The orthopoxviruses are a genus of the *Poxviridae* and have large, linear, double-stranded DNA genomes (*6*). They differ in the range of mammalian host species they infect and the severity of the disease they cause: variola virus (VARV) and camelpox virus (CMLV) have a narrow host range and can be highly virulent, while taterapox virus (TATV), infecting gerbils, does not cause significant morbidity (*7*, *8*). Cowpox virus (CPXV), monkeypox virus (MPXV), and vaccinia
90 virus (VACV) infect several host species, with varying severity of disease. VACV, whose origin and reservoir host are unknown (*9*), generally has low virulence in humans, elicits an immune response that is protective against VARV, and was used as the vaccine during the smallpox eradication campaign (*8*). VACV infection of mice is a useful model for studying orthopoxvirus gene function *in vivo*. Orthopoxvirus genomes are typically between ~186,000 and ~228,000
95 nucleotides (nt) long (*6*). Conserved genes, necessary for virus transcription and replication, are located in the central ~100,000 nt region of the genome, and are commonly used in phylogenetic inference (*10*). *In vitro* and *in vivo* studies show that orthopoxvirus genomes also contain many genes important for modulating host innate immunity and determining host range, but whose deletion does not prevent virus replication (*11*, *12*). Those genes are predominantly located near
100 the genome termini, and vary between orthopoxvirus species. The factors governing host range and virulence of orthopoxviruses are complex and not fully understood. For example, some genes that promote VACV virulence are inactivated in the more virulent VARV (*13–15*), while in other cases the loss or inactivation of host immune system-modulating genes in VACV can result in increased virulence (*12*). In general, orthopoxvirus species with a narrow host range
105 have fewer genes, such as VARV with ~162 functional genes, than those with broad host ranges, such as viruses of the CPXV species with ~209 (*6*). No orthopoxvirus species has a gene unique to that species, and all genes present across all orthopoxvirus species are also present in CPXV. It has therefore been proposed that the different orthopoxvirus species arose from a common ancestor containing a full set of CPXV genes, by a process of gene inactivation and loss (*6*).

110

The timeline of the emergence of smallpox in humans is unclear. Analysis of sequences from a 17th century Lithuanian mummy (VARV-VD21) (*16*), two Czech museum specimens from the 19th and 20th century (*17*), and VARV from the 20th century sampled during the eradication campaign, dates their most recent common ancestor (MRCA) to the 17th or 18th century (*10*, *16*–
115 *18*). However, written records of possible smallpox infections date back to at least 3000 years ago (ya), and mummified remains suggestive of smallpox date to 3570 ya (*1*, *19*). Ancient virus sequences recovered from archaeological remains provide direct molecular evidence for past infections and can reconcile the discrepancy between the written historical record of possible early infections and the time to the oldest available genetic sequences.

120

Recovering ancient variola virus DNA

To investigate the evolutionary history of VARV, we screened high-throughput shotgun sequencing data from skeletal and dental remains of 1867 humans living in Eurasia and the Americas between ~31,630 and ~150 ya. Using Kraken (*20*) on the microbial, protozoan, and
125 viral genomes from NCBI RefSeq databases (*21*–*23*), we identified a total of 26 samples with between 1 and 730 reads assigned to VARV (Fig. S1, Table S1). Based on the availability of additional sampling material, 13 of those were then enriched for viral sequences by a targeted capture approach, as described (*24*). Eleven of the 13 individuals are dated to 603–1050 Common Era (CE), overlapping the Viking Age (793–1066 CE) and are from northern Europe,
130 western Russia, or the UK, and two are historical 19th century from western Russia (Fig. 1A, Table 1, Supplementary Text). All of these 13 samples showed evidence of VARV infection after capture, with the following characteristics supporting authenticity of the recovered reads: elevated rates of mismatching nucleotides (due to post-mortem DNA damage) towards the read termini when aligned against a reference sequence (*25*), consistent with their respective ages and
135 sequencing depth (Fig. S2); mapped reads distributed evenly across the reference genome (Fig. S3); lowest edit distances (the number of nucleotide differences in a read alignment) to VARV (19th century samples) or CMLV/TATV (Viking age samples) when comparing read mapping against eight diverse orthopoxvirus reference genomes (Fig. S4); and 100% orthopoxvirus specificity when mapping reads against the NCBI ‘nt’ database (Table S2, Supplementary Text).

140

In four samples (VK281, VK382, VK388, and VK470, hereafter 'higher-coverage' samples), alignment against VARV (Accession number NC_001611.1) resulted in a mean coverage depth from 5.01X to 45.19X (Table 1, Fig. S3) and consensus sequences covering 95.89% to 99.56% of the genome (Tables 1, S3). Based on nucleotide sequence identity, the consensus sequences
145 for the four higher-coverage samples were collectively most similar to each other, and then to TATV (Table S4). For the remaining seven 'lower-coverage' Viking age samples, read edit distances were lowest against the assembled VK382 consensus sequence, further supporting their authenticity (Fig. S5).

150 ***Phylogenetic analysis reveals an unknown variola virus clade***

Before inferring phylogenetic trees based on the conserved central region of the genomes from the four higher-coverage sequences, we tested whether modern VARVs show evidence of genomic regions that descend from a virus population containing sequences resembling these four higher-coverage ancient VARV (aVARV) sequences and whether those sequences are
155 themselves recombinants. We examined the higher-coverage sequences and seven representative sequences from related modern orthopoxvirus species for evidence of recombination using seven algorithms from the RDP4 toolset (*26*). A 125 nt non-coding region of aVARV-VK470 was flagged as possibly due to recombination with an unknown sequence, but the evidence of recombination was uncertain. No other recombination event was suggested in the higher-
160 coverage aVARV sequences and no modern VARV sequence in the selection was suggested to be a recombinant derived from any aVARV sequence (Materials and Methods, Table S5).

To establish the phylogenetic placement of the four sequences from the higher-coverage samples, we inferred a maximum likelihood (ML) tree including 84 sequences chosen to
165 represent the full orthopoxvirus diversity. In the resulting tree (Figs. 1B, S6), the sequences form a now-extinct monophyletic clade in sister relationship with the clade consisting of all modern VARV and VARV-VD21 (collectively referred to as mVARV). We then placed partial consensus sequences of the lower-coverage samples onto the inferred tree using the evolutionary

placement algorithm, EPA-ng (*27*). Four of the seven Viking age samples (VK108, VK168,
170 VK515, and VK533) are confidently placed within the aVARV clade (Figs. 1B, S7–S12). Their placement position suggests a common ancestor with the more recent VK281 and VK470 lineages, supporting the existence of distinct Viking age lineages (Fig. S8). The placements of the other samples show some uncertainty, but are consistent with their respective sample ages: the Viking age samples (VK138, VK255, VK443) in the vicinity of the aVARV clade and the
175 19th century samples (FIN1 and KHA1) within the mVARV clade. This provides further evidence of the authenticity of these sequences and independently confirms the presence of a separate aVARV clade, as already shown by the higher-coverage samples alone.

A prior condition for making dated phylogenetic trees is the existence of a temporal signal in the
180 sequence data (*28*). A linear regression of sample ages and root-to-tip distances from an ML tree that includes sequences from the four higher-coverage samples and mVARV (Fig. S13) shows clear evidence of clock-like evolution (Fig. S14), in agreement with earlier results (*16*, *18*). Consequently, we proceeded to infer dated coalescent trees using BEAST2 (*29*) under six different evolutionary models (Table S6). With the best-fitting model, the MRCA of the
185 combined mVARV and aVARV sequences is dated to 1.7 thousand years ago (kya) (95% highest priority density interval (HPD95): 2.2–1.4 kya) and the MRCA of the aVARV clade alone is dated to 1.5 kya (HPD95: 1.7–1.4 kya). The MRCA of the mVARV clade is dated to 0.44 kya (HPD95: 0.6–0.35 kya), and to 0.31 kya (HPD95: 0.43–0.21 kya) for the mVARV clade without VARV-VD21, both of which are in line with previous molecular dating analyses (*10*,
190 *16*–*18*) (Table S7, Fig. S15).

The ancient variola viruses have a diverse pattern of gene inactivation

Many genes are inactivated (truncated, fragmented, or completely absent) in mVARV compared to other orthopoxviruses, possibly due to host adaptation (*6*, *13*). We therefore compared and
195 contrasted the gene status of sequences from the four higher-coverage aVARV samples, mVARV, plus CMLV and TATV (the orthopoxviruses most closely related to aVARV) (Figs. 2, S16, Table S8, Supplementary Text). Gene status analysis allows a detailed comparison of inter-

and intra-clade variation at the level of possible phenotypic differences. Gene differences are of interest individually and in combination, and the overall pattern of differences in gene status
200 complements the whole-genome diversity summarized by phylogenetic analysis.

Inter-clade differences in gene status can be logically organized into four groups (Fig. S17): A) inactive in mVARV and aVARV clades; B) inactive in mVARV but active in at least one aVARV sequence; C) active in mVARV but inactive in at least one aVARV sequence; and D)
205 active in both clades. Figure 2 shows details of groups A-C, highlighting gene differences between clades.

In group A (genes inactive in both clades), comparing the exact position of gene-inactivating mutations allows us to infer the likely status of the genes in the common ancestors of mVARV
210 and aVARV. Five genes in group A (*B23R;B24R/C17L;C18L*, 26765, *A25L*, *A37R*, and *B16R*) have identical inactivating mutations in both clades (Fig. 2, Supplementary Text). Identical inactivating mutations among orthopoxviruses not in sister relationship are rare: The 17 orthopoxviruses included in our gene-loss analysis (Materials and Methods) have 53 genes with gene-inactivating mutations in viruses not in sister relationship, but only two of these have the
215 mutation in the same location. This suggests that the five genes listed above were already inactivated in the common ancestor of mVARV and aVARV (Fig. 1B). A further 10 genes (20088, *A39R*, *A44L*, *A53R*, *B2R;B3R,* 215231*, C9L, A9L, A52R,* and 202711) are inactivated in both clades, but in different ways, and are thus projected to be active in the ancestor of mVARV and aVARV. This suggests parallel virus evolution following the divergence of the clades ~1.7
220 kya (Fig. S15).

A gene inactivated in both mVARV and aVARV may have induced a similar outcome after infection of humans. The VACV gene encoding a soluble IL-1β receptor is inactive in VACV strain Copenhagen (VACV-COP) (gene *B16R*) but functional in VACV strain Western Reserve
225 (VACV-WR, where it is denoted *B15R) (30)*. Infection of mice with either VACV-COP or VACV-WR in which *B16R* was inactivated, resulted in a febrile response (*15*). Conversely,

infection with either virus expressing a functional IL-1β receptor prevented the induction of fever (*15*). Human infections with VARV, which all have a disrupted *B16R* gene, are accompanied by high fever (*8*). The identical inactivation of *B16R* in the aVARV clade may
230 suggest that the symptoms of the disease caused by those viruses also included high fever and that the inactivation occurred prior to the divergence of the clades.

Group B (genes inactivated in mVARV but present or with undetermined status in some or all of the aVARV sequences) contains 14 genes. Eight of these (*B7R*, *CrmE*, *C2L*, *K1L*, *F3L*, *A35R*,
235 A40R, and *A55R*) encode known virulence factors or immunomodulators in VACV, three of which (*C2L*, *F3L*, and *A55R*) are members of the kelch-like family (Supplementary Text) (*12*). Deletion of the kelch-like proteins *C2L* or *A55R* in VACV-WR leads to increased lesion size in intradermally infected mice (*31*, *32*), while the virulence of VACV lacking *F3L* is reduced in the same model (*33*). Deletion of kelch-like proteins has also been proposed to reduce the host range
240 and virulence of CPXV-GRI90 *in vitro (34)*. Also in contrast with mVARV, aVARV-VK382 and aVARV-VK388 encode a predicted functional guanylate kinase (*A57R*). This gene is otherwise only active in CPXV, and appears to have been lost independently in CMLV, TATV, mVARV, aVARV-VK281, and aVARV-VK470 (Supplementary Text), suggesting that the gene was functional in their common ancestors.

245

Finally, group C (genes inactivated in at least one aVARV but in no mVARV sequence) contains seven genes (*CrmB*, *C1L, E5R, F10L, I8R, B20R,* and 210863, though only the first three have unequivocal gene-inactivating mutations). The effect of the inactivation of *C1L* and *E5R* in VACV is unknown. *CrmB* is a tumor necrosis factor receptor homolog and chemokine binding
250 protein (*35*). Interestingly, mVARVs have a functional *CrmB*, while it is identically inactivated in aVARV-VK281 and -VK470, and the start codon in aVARV-VK388 is replaced by a valine. However, in contrast to mVARV, *CrmE,* a tumor necrosis factor receptor homolog, and the chemokine binding protein *B7R* (*35*)*,* are functional in aVARV-VK281, -VK470, and -VK388 and together may provide the same functionality as *CrmB*. The older aVARV-VK382 has
255 functional versions of *CrmB*, *CrmE*, and *B7R*. The seven group C genes, already inactivated in

the aVARV clade, show that these viruses had taken a different genetic (though perhaps not functional) evolutionary path than the mVARV sequences, in which none of these genes were inactivated during an additional ~1000 years of evolution.

260 The gene status analysis and molecular clock dating of the internal nodes allow us to investigate the temporal trend of gene inactivation within the combined mVARV and aVARV clades. We observe an inactivation of genes over time (Fig. 3), supporting suggestions that orthopoxvirus species may derive from a common ancestor containing all genes found in orthopoxviruses today (*6*). A simple linear model of gene loss suggests that an ancestor of aVARV and mVARV with a
265 gene content similar to CPXV would have existed ~4 kya (Fig. 3). But any such model will necessarily involve unsupported and currently untestable assumptions, some of which will almost certainly be invalid, so caution is warranted (Supplementary Text). Loss of host-range genes has been correlated with an increase in host specificity (*11*), suggesting that the aVARVs and the common ancestor of mVARV and aVARV may have had a broader host range than
270 mVARV. However, host range in poxviruses is likely to be a multigenic trait that depends on the cooperation of several genes, as is the case in myxoma virus (*36*) and VACV (*37–41*). The higher number of active genes thus does not necessarily imply that the aVARVs had a host tropism for animals other than humans, although this is a possibility.

275 Within the aVARV clade, we observe a diverse pattern of gene inactivation (Figs. 2 and 3). Eleven genes are unequivocally inactivated in some, but not all, of the sequences from the four higher-coverage samples (in group B: 22682*, C2L, K1L, F3L, A35R, A40R,* 168145*, A57R;* in group C: *CrmB, C1L, E5R*). An additional four genes are inactivated in all four viruses, but with differing gene-inactivating mutations (in group A: *A39R,* 193435*, *C13L;C14L,* 202711),
280 suggesting that these genes were lost independently (Supplementary Text). By comparison, among 50 available mVARV sequences, the median number of pairwise differences in gene activation status is zero (mean: 0.6). Based on gene status, aVARV-VK281 and aVARV-VK470 can be grouped together, as can aVARV-VK382 and aVARV-VK388, differing unequivocally in status in one (*A35R*) and five (*A35R, A40R, CrmB, C1L, E5R*) genes, respectively (Fig. S18), in

285 agreement with the grouping based on their overall gene inactivation counts (Fig. 3), sample ages (Table 1, Fig. 3), and phylogenetic positions (Figs. 1B, S13-15). Our data suggest that VARV followed at least two independent trajectories of gene inactivation during its evolution: one leading to aVARV-VK281 and aVARV-VK470 and one to mVARV (Fig. 3). The geographical and temporal spread of the viruses in the aVARV clade (Figs. 1A, S8, S15) indicates that variola 290 viruses with a gene composition pattern different from mVARV existed over a period of at least ~450 years and circulated widely among humans during the Viking Age.

Historical perspective

We found evidence of VARV in 11 of 525 individuals pre-dating and overlapping the Viking 295 Age (~2%). If aVARV, like mVARV, did not lead to persistent infections (*8*) and if no viral particles remained in teeth or bone following an infection, these individuals all died during an acute infection. The 11 samples with reads matching VARV are likely not the only infections amongst those 525 individuals because: 1) The presence of viral DNA in ancient teeth and bones is certainly an imperfect diagnostic test of viral infection, with an elevated false negative rate due 300 to differences in DNA preservation, source tissue, or sequencing depth in the screening; 2) VARV is not thought to be viremic for the entire duration of the illness (*8*); 3) Based on mVARV case fatality rates (up to ~30%), many individuals survived infection; 4) Five Viking age samples with reads assigned to VARV by Kraken were not available for capture, due to lack of sample material. Moreover, the 525 individuals are drawn from a ~450 year period, during 305 which the overall population of Scandinavia rose from ~750K to ~1M (*42*). If the overall population increased linearly during this period and life expectancy at birth was ~30 years throughout, a total of ~13.1M people would have lived in Scandinavia during this period, so a sample of 525 individuals is only ~0.004% of the estimated total population. Given this considerable uncertainty, it would not be prudent to use the 2% detection rate to estimate 310 smallpox prevalence or case fatality rates during the Viking Age. The likely inaccurate (low) detection rate and the very small sample size argue against regarding the figure as a prevalence. Even if all other sources of uncertainty were resolved, it would still not be clear how the figure relates to case fatality rates during that period because we cannot be sure that the individuals died as a result of their infections.

315

Hypotheses regarding the earliest history of VARV have been derived exclusively from often ambiguous historical accounts, and the visual examination of mummies dating from as early as 3570 ya (*1*, *19*). The dates of the earliest likely presence of smallpox in specific regions are difficult to determine from written records and have in some cases been a matter of dispute (*1*, *8*,
320 *19*, *43*). Ancient DNA sequences can resolve such questions by providing clear evidence of VARV presence in time and space. The dating of the aVARV samples, from as early as 603 CE, matches that of multiple written accounts of likely smallpox infections in southern and western Europe from the late 6th century onwards (*1*, *8*, *19*, *44*, *45*). Our finding of the virus in northern Europe at these times disproves various suggestions of first introductions involving later dates.
325 For example, the introduction (or later establishment) of smallpox into Europe via returning Crusaders (*8*, *9*), into Europe via Spain during the Moorish invasion of 710 CE (*9*), or into England by Norman invaders in 1241–2 CE (*46*), as well as the claim that the virus had not reached northern Europe by 1000 CE (*8*). We did not find evidence for infections in samples from earlier time periods, despite screening 619 individuals pre-dating the Viking Age (Fig. S1).
330 Taken together, the written record and the confirmed aVARV infections suggest a pan-European presence of the virus by the end of the Viking Age at the latest, as suggested for parts of southern and western Europe based purely on written records (*1*, *19*, *44*).

The Viking age sequences reported here push the definitive date of the earliest VARV infection
335 in humans back by ~1000 years and reveal the existence of a previously unknown, now-extinct virus clade. The ancient viruses were following a genotypic evolutionary path that differs from modern VARV. This is highlighted by at least three genes inactivated in some aVARV sequences but still active in the mVARV clade, and by 10 genes inactivated in both clades, but with gene-inactivating mutations that differ between clades. The ancient viral genomes show
340 reduction of gene content during the evolution of VARV, and that multiple combinations of gene inactivations have led to viruses able to circulate widely within the human population.

Materials and Methods

Software package versions

345 The following software package versions were used, with default arguments, unless otherwise noted. BEAST2 (*29*): 2.4.7, BLASTn (*49*): 2.2.29+, bModelTest (*50*): 1.0.4, Bowtie2 (*51*): 2.3.5.1, BWA (*52*): 0.7.15-r1140, DIAMOND (*53*): 0.9.23, EPA-ng (*27*): 0.3.6, Gappa (*54*): 0.5.0, gargammel (*55*): released without a version number, October 2017, Geneious (*56*): 9, ggtree (*57*): 1.16.6, Kraken (*20*): 1.0, MAFFT (*58*): 7.419, mapDamage (*59*): 2.0.8-2, Mosdepth
350 (*60*): 0.2.6, PastML (*48*): 1.9.20, Picard (*51*): 2.20.2, R (*61*): 3.6.0 , RAxML-NG (*62*): 0.7.0 BETA, RDP4 (*26*): 4.9.5, Samtools (*63*): 1.9, SciPy (*64*): 1.4.1, TempEst (*65*): 1.5, Tracer (*66*): 1.6.0, TreeAnnotator (*29*): 2.2.4 pre-release.

Screening and sample preparation

355 Using Kraken on a database of all microbial, protozoan, and viral genomes in the NCBI RefSeq databases (*21–23*), we analysed whole-genome shotgun sequencing data from 1867 individuals for the presence of reads that best matched VARV. We identified a total of 26 samples with between 1 and 730 reads assigned to VARV (Fig. S1) as candidates for enrichment of viral target sequences in library by capture. Of these, sufficient sample materials were available for capture
360 processing for 13 samples. Uniquely dual-indexed sequencing libraries (2-6 libraries/sample) were prepared from 1-3 ancient DNA (aDNA) extracts/sample. Libraries were pooled into capture reactions containing 1183-1880 ng/rxn. Hybridisation and sequencing on HiSeq4000 were performed as described previously (*24*). An additional round of sequencing was performed on pools of enriched libraries, using the Illumina NovaSeq6000 (S4 flowcell, XP workflow,
365 2x100 cycles). Read length and insert size distributions are shown in Fig. S19.

Authenticity

The following points argue for the authenticity of the recovered DNA sequences.

370 1) Standard precautions (*67*) and recommended methodologies (*68*) for working with ancient DNA were applied.

2) Clear damage patterns characteristic of DNA sequences recovered from archaeological remains are present in all samples with sufficient viral reads available. Note that the absence of a damage pattern in a sample with few reads does not (by itself) indicate that
375 the sample reads are not ancient. The human reads all show clear damage patterns in all samples, showing that the nucleic acids present in the samples are indeed ancient.
3) The ancient samples included in the study have reads matching across the entire TATV reference genome (Fig. S3). This is in stark contrast to false positives that regularly occur with human DNA sequences in which there are matches of only one or two poxvirus
380 proteins out of ~160-220, in which cases the reads in fact are better matches for similar-but-unrelated sequences, e.g., human mRNA.
4) Reads matching TATV were matched against the entire NCBI 'nt' database using BLASTn. 118,049 of 118,118 (99.94%) of reads had matches in the 'nt' database, all of which had a poxvirus as the best match, ignoring matches against artificial constructs
385 (Table S2). A small number of reads with no BLASTn match in the 'nt' database is expected (Materials and Methods).
5) The edit distances of the reads are fully consistent with the placement of (full or partial) consensus sequences in phylogenetic trees. Reads from all nine aVARV clade samples very clearly have lowest edit distances against TATV and CMLV, the intermediate
390 samples (VK138 and VK255) have read edit distances that are closest to VARV/TATV/CMLV, and the two modern samples (FIN1 and KHA1) have read edit distances very clearly closest to VARV. As would also be expected, edit distances against our aVARV-VK382 sequence are the lowest of all for the Viking age samples.
6) The fact that the almost-complete genome consensus sequences from the four higher-
395 coverage ancient samples form a self-consistent separate basal clade indicates that these sequences are very unlikely to be due to four independent modern contaminations. If, for example, 20th century smallpox sampling entirely missed a still-circulating clade that these samples actually came from, those lineages would need to have evolved with a far lower substitution rate than other variola viruses due to the relatively very short branch
400 lengths leading to the ancient sequences in the Maximum Likelihood tree (Fig. S6).
7) The DNA sequenced in these samples was drilled out of the interior of well-preserved teeth (11 samples) and petrous bone (2 samples). Pox virus DNA was therefore very

likely circulating in some form in the bloodstream of the individual at the moment of death. Subsequent contamination of any form of virus or viral DNA prior to aDNA
405 extraction in our dedicated clean laboratory is unlikely.

8) The EPA-ng placements from the lower-coverage samples provide important information via the phylogenetic placement of the (incomplete) consensus sequences from those samples. The fact that all ancient samples are placed consistently (according to sample ages) within or basal to the aVARV or mVARV clades is further evidence of the overall
410 authenticity and consistency of the reads. We also show that reads simulated *in silico* from a different orthopoxvirus genome which was not included in the tree are consistently placed on their expected branch, and not with the aVARV or mVARV clades (Fig. S12).

9) Present-day analysis of purported VARV reads has the advantage that contamination by a
415 modern variola virus can be definitively ruled out, given the nature of the associated disease. No member of the analysis team has any poxvirus infection and the facilities where the sample extraction, library preparation, and sequencing was done is not a virology laboratory (rather it is focused on ancient non-microbial host genome recovery) and has never been involved in processing modern viral material. The paper author (LV)
420 who carried out all the targeted laboratory work received a routine VARV vaccination prior to 1980.

Read poxvirus specificity

To check that sample reads found to match poxvirus sequences were not in fact better matches
425 for other sequences, reads matching TATV (GenBank accession no.: NC_008291.1) using bowtie2 --end-to-end matching and with MAPQ >= 30 were checked for specificity by matching them against the NCBI 'nt' nucleotide database (downloaded during the week of January 13, 2020) using BLASTn (-task blastn -evalue 0.01). The best-matching database sequence for each read was checked to see if it corresponded to a poxvirus. Matches against bacterial artificial
430 chromosomes and synthetic constructs (matching the case-insensitive regular expression (BAC cloning vector|synthetic construct)) were ignored. Table S2 shows that apart from 69 reads (from a total of 118,118) that did not match anything in the 'nt' database, all reads had a poxvirus as

their best match, as determined by the 'nt' database subject title matching the case-insensitive regular expression (pox|ectromelia|variola|akhmeta|abatino|vaccinia).

435

The --end-to-end option to Bowtie2 and the MAPQ >=30 filtering were applied to examine only reads that mapped in full and with a high confidence regarding their genome match position. This is necessary because orthopox viruses contain repetitive low-complexity regions near their temini and such sequences often match many hundreds of host chromosome sequences from a

440 diverse range of unrelated organisms. These matches sometimes exceed the default 500 matches returned by BLASTn, meaning no orthopox match may be reported at all, despite a perfect match with an orthopox genome. The stricter matching (in particular, only considering reads matching with MAPQ >=30) eliminates low-complexity viral reads with these ambiguous placements from consideration.

445

It is expected that some orthopox reads matched by Bowtie2 will not match anything with BLASTn. This is due to search algorithm variations in seeding, match scoring values and thresholds, and reporting thresholds. For example, despite nucleotide identity exceeding 90%, BLASTn (-task blastn) will not match a sequence against a copy of itself with every 11th

450 nucleotide mutated because a perfectly-matching seed region of 11 nucleotides is required before a region is further considered as a potential match. A different search algorithm will identify such matches. These algorithmic idiosyncrasies are the reason we also examined read matches found by DIAMOND and bwa as part of the consensus and gene-level analysis, with both analyses subject to manual visual inspection of aligned reads.

455

Investigation of gene-inactivating mutations

To assemble the reads used for the generation of consensus sequences and the investigation of gene-inactivating mutations, reads from the capture were mapped against orthopoxvirus reference genomes using DIAMOND, matching against version 13.0 of the RVDB protein

460 database (*69*), and BWA (using both the mem and aln (with and without the -l option set to 1000) algorithms, with other arguments left at their defaults), matching against a collection of orthopox genomes from NCBI (GenBank accession numbers AY009089.1, AF438165.1,

AF482758.2, X94355.2, AF012825.2, AF380138.1, DQ437594.1, AM501482.1, M35027.1, AY243312.1, L22579.1, X69198.1, Y16780.1, KY358055.1). Reads from all mapping
465 algorithms were combined and de-duplicated using Picard after mapping to a TATV genome (GenBank accession no.: NC_008291.1) using Bowtie2.

In order to assess gene content of the higher-coverage aVARV samples in the absence of a reference genome, we investigated the presence of gene-inactivating mutations in the ancient
470 samples by comparison of alignments made to 17 orthopoxviruses. Information about the offset locations of 219 homologous genes in 17 orthopoxvirus reference genomes (Dataset 3) is presented in Supplementary Table 1 of Hendrickson *et al.*, 2010 (*6*). These offsets were used to extract the sequences of the homologous genes in the 17 reference sequences. Excluding cases where a gene is absent in a reference sequence, this leads to 3422 sequences for the 219 genes.
475 Supplementary Table 1 in Hendrickson *et al.*, 2010 (*6*) was also used to acquire information on status (presence, absence, fragmentation, and truncation) of genes in the 17 reference sequences. We aligned reads independently at the gene level with sequences from this broad set of references, to make it possible to correctly deal with cases where, for example, two genes from an ancient sequence matched one gene best in one modern reference (e.g., CMLV) and another
480 gene best in a different modern reference (e.g., TATV), or where a gene is present in the ancient virus, but completely absent in the modern reference. Aligning against just one of the modern full-genome references would necessarily risk potential error. We considered this the only defensible approach, given the fact that the gene composition of the ancient sequences was unknown.
485

For each gene, all reads from an ancient sample were aligned to one or more of the 17 reference gene sequences using Geneious, medium sensitivity, fast parameters, and then visually inspected for the presence of gene-inactivating mutations. The reference(s) to align against were chosen in the following way: reads for each ancient sample were matched against each reference gene
490 sequence using BLASTn, and a consensus sequence for each gene was generated using BWA (aln algorithm) and samtools. The reference gene sequence with over 50% coverage of the gene sequence and the highest median bit score was chosen as the reference. If this was inconclusive, additional references were used, usually one of CPXV-Gri, -Ger, or -BR. For each ancient

sample, gene status, coverage, and reference sequence are given in Table S8. Gene-inactivating
495 mutations in the ancient sequences were classified as either *certain*, *uncertain*, or *uncertain if functional* according to the following criteria:

A *certain* gene-inactivating mutation: There is a mutation that leads to a stop codon (insertion / deletion / point mutation) which is covered by at least three reads, and the reference against
500 which the ancient reads are aligned is also truncated or fragmented in an identical fashion.
An *uncertain* gene-inactivating mutation: There is a mutation that leads to a stop codon which is covered by less than three reads.
An *uncertain if functional* gene: A gene of intermediate length: shorter than a gene that is functional in some references, but substantially longer than inactivated versions of the same gene
505 in others (see gene *A9L*).

The ancient gene status was further confirmed by mapping the reads, consensus sequences resulting from the Geneious alignments against the gene reference(s), and the contigs from a *de novo* assembly of the reads (also performed in Geneious, default parameters) against the CPXV-
510 Ger reference genome (GenBank accession no.: DQ437593.1) and inspecting the resulting consensus for the expected genes. The status of each gene in the resulting consensus sequence agreed with the analysis performed above.

Ten genes (1393 and 222274 (*C23L/B29R*), 2286 and 221381 (*CrmB*), 3420 and 220247
515 (*C19;C20L/B25/B27R*), 5392 and 218275 (*C17L;C18L/B23R;B24R*), 7652 and 216015 (*C16L/B22R*)) in the orthopoxvirus genome are occasionally diploid. While both versions of these genes are present in CPXV-Ger, -Gri, and -BR, only one version of the gene is present in VARV. Given low genome coverage and short DNA template fragments, it is not clear how to determine whether these genes are present at the 3' or 5' (or both) ends of the aVARV genomes.
520 The (potentially) two copies of the gene were thus treated as one in the analyses described above.

To infer the gene status at internal nodes of the phylogeny of mVARV, aVARV, CMLV, and TATV, we used the DELTRAN algorithm implemented in PastML, to perform a maximum parsimony ancestral state reconstruction. Gene-inactivating mutations were determined to be

525 identical in two tips if the same mutation leading to an inactivation of the gene was present, whether or not a different upstream mutation may have previously ablated the gene in one of them.

We compared differences in gene activation status using a larger set of modern VARV
530 sequences. The Virus Orthologous Clusters database contains (among other orthopoxviruses) the gene annotations of 50 modern VARV (*70*). Each gene is assigned to a cluster based on BLASTp results and manual curation, genes within each cluster are orthologous. We determined correspondence between the assignments in Hendrickson et al., 2010 (*6*) and the orthologous clusters based on sequence identity. Only clusters which had corresponding genes assigned by
535 Hendrickson et al., 2010 were included. Genes in the Virus Orthologous Cluster database were deemed functional if they were at least 80% of the length of the corresponding functional gene in Hendrickson et al., 2010. To compare the gene content of the 50 modern VARV sequences, we counted the number of differences in gene status between two sequences. Gene status was considered different if the gene was present in one sequence and had a gene-inactivating
540 mutation or was absent in the other.

Generation of consensus sequences

Reads from aVARV-VK382, the sample with the highest coverage, were mapped against VARV strain SLN68_258 (GenBank accession no.: DQ441437.1) in Geneious using Medium sensitivity
545 / fast and iterate up to five times. The alignment was visually inspected and sections not present in the reference sequence were determined by iteratively mapping the reads against the new consensus sequence from either side of the insertion, until the consensus sequences from the mapped reads overlapped. The resulting consensus had 41 undetermined positions out of 192,255 nucleotides. Gene status and their position in the genome order were both fully consistent with
550 the analysis performed in the previous section. Consensuses for the remaining samples were generated by mapping the reads against the aVARV-VK382 consensus using Geneious using Medium sensitivity / fast and iterate up to five times. Manual inspection showed that all expected genes were present in the ancient consensus sequences, consistent with the gene loss analysis (Table S3). The consensus sequences were also consistent with contiguous sequences ("contigs")

555 resulting from a *de novo* assembly of the reads (performed in Geneious, using default parameters). We called majority rule consensus sequences at a depth of one read, taking into account residue quality.

Recombination analyses

560 Seven algorithms, as implemented in the RDP4 package, were used to search for evidence of recombination in a selection of 11 orthopoxvirus sequences. The algorithms were: 3Seq (*71*), BootScan (*72*), Chimaera (*73*), GENECONV (*74*), MaxChi (*75*), RDP (*76*), and SiScan (*77*). The sequences (with GenBank accession numbers in parentheses) were: Akhmeta (MH607143), CMLV (AY009089), CPXV (KC813492), CPXV (KC813508), TATV (NC_008291), VARV-
565 VD21 (BK010317), VARV (NC_001611), aVARV-VK281, aVARV-VK382, aVARV-VK388, and aVARV-VK470. A region of 125 nucleotides of aVARV-VK470 was identified as being of possible recombinant origin by three of the seven RDP4 algorithms (3Seq, BootScan, GENECONV), with aVARV-VK281 as the major parent and the minor contribution from an unknown parent. The RDP4 program warned that "*the proposed recombination signal may be*
570 *attributable to a process other than recombination*" and of a "*possible misidentification of recombinant*". No recombination signals were detected with aVARV-VK281, aVARV-VK382, or aVARV-VK388 as recombinant sequences. No modern sequence was identified as being the possible descendant of a recombination involving any ancient sequence. Details on the algorithms, p-values, and the possible breakpoints for aVARV-470 are shown in Table S5.
575

Phylogenetic analyses

1. Maximum likelihood trees: Separate trees were made from Datasets 1 and 2. The sequences were aligned using MAFFT. Maximum likelihood (ML) trees were generated using RAxML-NG with a K81uf+G+I (selected using bModelTest) substitution model, and an ML estimate of tree
580 topology, branch lengths, substitution rates, and nucleotide frequencies, with support estimated using 1000 bootstrap replicates.

2. Lower-coverage samples trees: The EPA-ng algorithm was used to infer the positions of consensus sequences from the nine lower-coverage samples (VK108, VK138, VK168, VK255, VK443, VK515, VK533, FIN1, and KHA1) on the ML tree made from Dataset 2, including the

585 sequences from the four higher-coverage samples. For all lower-coverage samples, consensus sequences were obtained by extracting the majority allele from BAM files mapped against the TATV reference genome (GenBank accession no.: NC_008291.1) using the local alignment approach, to reduce the impact of genotyping errors due to post-mortem DNA damage (Fig. S2). The resulting consensus sequence was aligned to the multi-sequence alignment of Dataset 2 from 590 which the original ML tree was inferred (see 'Maximum likelihood trees', above) using MAFFT (version 7.407) with the '--add' and '--keeplength' options. EPA-ng was run using the same model parameters as used for the original ML tree (K81uf+G+I). Placements of lower-coverage samples were further analyzed using gappa, and the uncertainty in placements for each sample was visualized by plotting the inferred likelihood weight ratios (LWRs) for each branch on the 595 reference tree, using the ggtree package in R (Figs. S7, S8).

In order to assess the EPA-ng placements of the lower-coverage samples, BAM files with mapped reads from the higher-coverage sample VK388 (aVARV) and the published sample VD21 (mVARV) were randomly subsampled to sets containing 30, 50, 100, 500, 1000, and 2000 reads, with 100 replicates each. Consensus sequences for each sample and replicate were made 600 and placed back on the reference tree, using the same approach as for the lower-coverage samples (described above). Placement uncertainty was visualized using the normalized placement edge mass (*78*) across all 100 replicates, for each branch on the reference tree (Fig. S9).

To further assess the reliability of the placements, two additional metrics were used (*79*). First, 605 the distributions of LWRs for the first, second, and third most-likely placement for each subsample set were inferred. If many replicates of the placements for a particular subsample set have high confidence, a large fraction of the most-likely placements will show high LWRs (close to the maximum LWR of 1), and if second or third most-likely placements are observed they will have very low LWRs (Fig. S11A). Second, the 'Expected Distance between Placement 610 Locations' (EDPL) was calculated. EDPL is defined as the average distance on the reference tree between different placements for a query sequence, weighted by their respective LWR. A particular query sequence can have low confidence in the actual placement (i.e., many placements with low LWR), but high confidence for a local neighbourhood of the reference tree, for example in the case of sets of highly similar sequences within a sub-clade of the reference

615 tree. The EDPL thus gives a measure of the locality of the query placements across the reference tree, with low values indicating high locality (Fig. S11B).

We also tested the ability of EPA-ng to correctly place a sequence not currently in the tree. For that we chose the Abatino orthopoxvirus (GenBank accession no.: MH816996.1) sequence, which diverges basal to Ectromelia virus (*80*). We simulated a total of 1,000,000 sequencing
620 reads from the Abatino reference genome using gargammel. Sequencing reads were simulated from HiSeq 2500 Illumina single-end runs of length 81 bp, using an empirical read length distribution derived from sample VK388. Following removal of sequencing adapters, read mapping, consensus sequence building, and phylogenetic placement was carried out as previously for random subsamples of 30, 50, 100, 500, 1000, and 2000 reads (Fig. S12). As
625 expected, the placement edge masses were concentrated on the branch leading to the Ectromelia virus clade, even for as few as 30 simulated reads.

To complement and confirm the EPA-ng placements, branch-defining single nucleotide polymorphisms (SNP) were inferred from the reference alignment. A SNP is considered branch-defining if all sequences descending from the branch share the same nucleotide at that location,
630 and all other sequences (on the alternate branches) do not. For each lower-coverage sample and each branch of the reference tree, the number and fraction of SNPs where the allele observed in the sample matched the branch-defining allele is shown in Fig. S20.

3: Dated coalescent trees using BEAST2. We performed a linear regression of root-to-tip dates against sampling dates. Root-to-tip distances were extracted using TempEst from the ML tree
635 inferred from Dataset 1 and the sequences from the four higher-coverage samples in Fig. S6 (Dataset 1). The regression analysis was performed in SciPy. Dated coalescent trees were inferred using BEAST2. Forty-eight modern VARV sequences were used, as well as two published ancient sequences from Czech Republic, and VARV-VD21 (Dataset 1). Using bModelTest, we selected a TPM1 substitution model with unequal base frequencies, invariant
640 sites, and gamma distributed rate heterogeneity among sites. The clock rate was constrained using a uniform($1x10^{-9}$–$1x10^{-3}$ substitutions / site / year (s/s/y)) prior. Proper priors were used throughout. Trees were inferred using strict and relaxed log-normal molecular clocks, as well as constant and exponential coalescent and Bayesian skyline population priors. The Markov chain Monte Carlo analysis was run for 100M (for the strict clock model and coalescent constant and
645 exponential population priors) or 200M (for the strict clock model and Bayesian skyline

population prior, and the log-normal relaxed clock model and coalescent constant, coalescent exponential, and Bayesian skyline population priors) generations, sampling every 2000 generations. Convergence and mixing were assessed using Tracer, and an effective sample size >200 was reached for all relevant parameters. Final tree files were sub-sampled to contain 10,000
650 trees and Maximum Clade Credibility trees were summarised using TreeAnnotator. Path sampling, as implemented in BEAST2, was used to select the best-fitting molecular clock model and population prior. Per path sampling run, 50 steps with a chain length of 1,000,000 generations were run. Likelihood values were compared using a Bayes factor test (*81*).

655 **References and Notes**

1. D. R. Hopkins, *Princes and Peasants: Smallpox in History* (The University of Chicago Press, Chicago, 1983).

2. E. Jenner, On the Origin of the Vaccine Inoculation. *Med Phys J.* 5, 505–508 (1801).

3. G. L. Smith, G. McFadden, Smallpox: anything to declare? *Nat. Rev. Immunol.* 2, 521–
660 527 (2002).

4. C. B. Coyne, Horsepox: Framing a dual use research of concern debate. *PLoS Pathog.* 14, e1007344 (2018).

5. N. Sklenovská, M. Van Ranst, Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. *Front Public Health*. 6, 241 (2018).

665 6. R. C. Hendrickson, C. Wang, E. L. Hatcher, E. J. Lefkowitz, Orthopoxvirus genome evolution: the role of gene loss. *Viruses*. 2, 1933–1967 (2010).

7. S. Parker, R. Crump, H. Hartzler, R. M. Buller, Evaluation of Taterapox Virus in Small Animals. *Viruses*. 9 (2017), doi:10.3390/v9080203.

8. F. Fenner, D. A. Henderson, I. Arita, Z. Jezek, I. D. Ladnyi, *Smallpox and its Eradication*
670 (World Health Organization, Geneva, 1988).

9. D. Baxby, *Jenner's Smallpox Vaccine: The Riddle of Vaccinia Virus and Its Origin* (Heinemann Educational Publishers, 1981).

10. C. Smithson, J. Imbery, C. Upton, Re-Assembly and Analysis of an Ancient Variola Virus Genome. *Viruses*. 9 (2017), doi:10.3390/v9090253.

675 11. K. A. Bratke, A. McLysaght, S. Rothenburg, A survey of host range genes in poxvirus genomes. *Infect. Genet. Evol.* 14, 406–425 (2013).

12. G. L. Smith, C. T. O. Benfield, C. Maluquer de Motes, M. Mazzon, S. W. J. Ember, B. J. Ferguson, R. P. Sumner, Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity. *J. Gen. Virol.* 94, 2367–2392 (2013).

680 13. B. Aguado, I. P. Selmes, G. L. Smith, Nucleotide sequence of 21.8 kbp of variola major virus strain Harvey and comparison with vaccinia virus. *J. Gen. Virol.* 73 (Pt 11), 2887–2902 (1992).

14. J. B. Moore, G. L. Smith, Steroid hormone synthesis by a vaccinia enzyme: a new type of virus virulence factor. *EMBO J.* 11, 3490 (1992).

685 15. A. Alcami, G. L. Smith, A mechanism for the inhibition of fever by a virus. *Proceedings of the National Academy of Sciences*. 93, 11029–11034 (1996).

16. A. T. Duggan, M. F. Perdomo, D. Piombino-Mascali, S. Marciniak, D. Poinar, M. V. Emery, J. P. Buchmann, S. Duchêne, R. Jankauskas, M. Humphreys, G. B. Golding, J. Southon, A. Devault, J.-M. Rouillard, J. W. Sahl, O. Dutour, K. Hedman, A. Sajantila, G. L. Smith, E. C.
690 Holmes, H. N. Poinar, 17 Century Variola Virus Reveals the Recent History of Smallpox. *Curr. Biol.* 26, 3407–3412 (2016).

17. P. Pajer, J. Dresler, H. Kabíckova, L. Písa, P. Aganov, K. Fucik, D. Elleder, T. Hron, V. Kuzelka, P. Velemínsky, J. Klimentova, A. Fucikova, J. Pejchal, R. Hrabakova, V. Benes, T. Rausch, P. Dundr, A. Pilin, R. Cabala, M. Hubalek, J. Stríbrny, M. H. Antwerpen, H. Meyer,
695 Characterization of Two Historic Smallpox Specimens from a Czech Museum. *Viruses*. 9 (2017), doi:10.3390/v9080200.

18. A. F. Porter, A. T. Duggan, H. N. Poinar, E. C. Holmes, Comment: Characterization of Two Historic Smallpox Specimens from a Czech Museum. *Viruses*. 9 (2017), doi:10.3390/v9100276.

700 19. C. W. Dixon, *Smallpox* (J & A Churchill Ltd, 104 Cloucester Place, London, 1962).

20. D. E. Wood, S. L. Salzberg, Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46 (2014).

21. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova,
705 V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D.

Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–45 (2016).

22. T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A. Lomsadze, K. D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*. 44 (2016), pp. 6614–6624.

23. J. R. Brister, D. Ako-Adjei, Y. Bao, O. Blinkova, NCBI viral genomes resource. *Nucleic Acids Res.* 43, D571–7 (2015).

24. B. Mühlemann, T. C. Jones, P. de B. Damgaard, M. E. Allentoft, I. Shevnina, A. Logvin, E. Usmanova, I. P. Panyushkina, B. Boldgiv, T. Bazartseren, K. Tashbaeva, V. Merz, N. Lau, V. Smrčka, D. Voyakin, E. Kitov, A. Epimakhov, D. Pokutta, M. Vicze, T. D. Price, V. Moiseyev, A. J. Hansen, L. Orlando, S. Rasmussen, M. Sikora, L. Vinner, A. D. M. E. Osterhaus, D. J. Smith, D. Glebe, R. A. M. Fouchier, C. Drosten, K.-G. Sjögren, K. Kristiansen, E. Willerslev, Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature*. 557, 418–423 (2018).

25. E. Cappellini, A. Prohaska, F. Racimo, F. Welker, M. W. Pedersen, M. E. Allentoft, P. de Barros Damgaard, P. Gutenbrunner, J. Dunne, S. Hammann, M. Roffet-Salque, M. Ilardo, J. V.

Moreno-Mayar, Y. Wang, M. Sikora, L. Vinner, J. Cox, R. P. Evershed, E. Willerslev, Ancient Biomolecules and Evolutionary Inference. *Annu. Rev. Biochem.* 87, 1029–1060 (2018).

26. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and
730 analysis of recombination patterns in virus genomes. *Virus Evolution*. 1 (2015), doi:10.1093/ve/vev003.

27. P. Barbera, A. M. Kozlov, L. Czech, B. Morel, D. Darriba, T. Flouri, A. Stamatakis, EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst. Biol.* (2018), doi:10.1093/sysbio/syy054.

735 28. A. J. Drummond, O. G. Pybus, A. Rambaut, R. Forsberg, A. G. Rodrigo, Measurably evolving populations. *Trends in Ecology & Evolution*. 18 (2003), pp. 481–488.

29. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537 (2014).

740 30. A. Alcamí, G. L. Smith, A soluble receptor for interleukin-1 beta encoded by vaccinia virus: a novel mechanism of virus modulation of the host response to infection. *Cell*. 71, 153–167 (1992).

31. M. Pires de Miranda, P. C. Reading, D. C. Tscharke, B. J. Murphy, G. L. Smith, The vaccinia virus kelch-like protein C2L affects calcium-independent adhesion to the extracellular
745 matrix and inflammation in a murine intradermal model. *J. Gen. Virol.* 84, 2459–2471 (2003).

32. P. M. Beard, Vaccinia virus kelch protein A55 is a 64 kDa intracellular factor that affects virus-induced cytopathic effect and the outcome of infection in a murine intradermal model. *J. Gen. Virol.* 87, 1521–1529 (2006).

33. G. C. Froggatt, G. L. Smith, P. M. Beard, Vaccinia virus gene F3L encodes an
750 intracellular protein that affects the innate immune response. *J. Gen. Virol.* 88, 1917–1921 (2007).

34. G. Kochneva, I. Kolosova, T. Maksyutova, E. Ryabchikova, S. Shchelkunov, Effects of deletions of kelch-like genes on cowpox virus biological properties. *Arch. Virol.* 150, 1857–1870 (2005).

755 35. A. Alejo, M. B. Ruiz-Argüello, Y. Ho, V. P. Smith, M. Saraiva, A. Alcami, A chemokine-binding domain in the tumor necrosis factor receptor from variola (smallpox) virus. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5995–6000 (2006).

36. J. Liu, S. Wennier, L. Zhang, G. McFadden, M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. *J.*
760 *Virol.* 85, 3270–3282 (2011).

37. G. Sivan, P. Ormanoglu, E. C. Buehler, S. E. Martin, B. Moss, Identification of Restriction Factors by Human Genome-Wide RNA Interference Screening of Viral Host Range

Mutants Exemplified by Discovery of SAMD9 and WDR6 as Inhibitors of the Vaccinia Virus K1L-C7L- Mutant. *MBio*. 6, e01122 (2015).

765 38. S. Gillard, D. Spehner, R. Drillien, A. Kirn, Localization and sequence of a vaccinia virus gene required for multiplication in human cells. *Proceedings of the National Academy of Sciences*. 83, 5573–5577 (1986).

39. M. E. Perkus, S. J. Goebel, S. W. Davis, G. P. Johnson, K. Limbach, E. K. Norton, E. Paoletti, Vaccinia virus host range genes. *Virology*. 179, 276–286 (1990).

770 40. J. O. Langland, B. L. Jacobs, The role of the PKR-inhibitory genes, E3L and K3L, in determining vaccinia virus host range. *Virology*. 299, 133–141 (2002).

41. B. Liu, D. Panda, J. D. Mendez-Rios, S. Ganesan, L. S. Wyatt, B. Moss, Identification of Poxvirus Genome Uncoating and DNA Replication Factors with Mutually Redundant Roles. *J. Virol.* 92 (2018), doi:10.1128/JVI.02152-17.

775 42. C. McEvedy, R. Jones, *Atlas of world population history* (Puffin, 1978).

43. C. Creighton, *A History of Epidemics in Britain: From A. D. 664 to the extinction of plague* (1891).

44. R. Willan, *Miscellaneous Works of the Late Robert Willan, M.D* (T. Cadell, London, 1821).

780 45. J. Moore, *The History of the Small Pox* (Longman, Hurst, Rees, Orme, and Brown, Paternaster Row, London, 1815).

46. E. Paschen, in *Jochmann's Lehrbuch der Infektionskrankheiten* (Springer Verlag, ed. 2, 1924).

47. P. J. Reimer, E. Bard, A. Bayliss, J. Warren Beck, P. G. Blackwell, C. B. Ramsey, C. E.
785 Buck, H. Cheng, R. Lawrence Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Haflidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. Felix Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. Marian Scott, J. R. Southon, Richard A Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13

Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*. 55 (2013), pp. 1869–
790 1887.

48. S. A. Ishikawa, A. Zhukova, W. Iwasaki, O. Gascuel, A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evol.* 36, 2069–2085 (2019).

49. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. 10, 421 (2009).

795 50. R. R. Bouckaert, A. J. Drummond, bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* 17, 42 (2017).

51. Broad Institute, Picard Toolkit, (available at http://broadinstitute.github.io/picard/).

52. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26, 589–595 (2010).

800 53. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. 12, 59–60 (2015).

54. L. Czech, P. Barbera, A. Stamatakis, Genesis and Gappa: Processing, Analyzing and Visualizing Phylogenetic (Placement) Data, , doi:10.1101/647958.

55. G. Renaud, K. Hanghøj, E. Willerslev, L. Orlando, gargammel: a sequence simulator for
805 ancient DNA. *Bioinformatics*. 33, 577–579 (2017).

56. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious

Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28, 1647–1649 (2012).

810 57. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36 (2016).

58. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013).

815 59. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 29, 1682–1684 (2013).

60. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 34, 867–868 (2018).

820 61. R Development Core Team, 'R: A Language and Environment for Statistical Computing' (R Foundation for Statistical Computing, Vienna, Austria, 2008), (available at http://www.R-project.org).

62. A. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference (2018), ,
825 doi:10.1101/447110.

63. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25, 2078–2079 (2009).

64. E. Jones, T. Oliphant, P. Peterson, SciPy: Open Source Scientific Tools for Python
830 (2001), (available at http://www.scipy.org/).

65. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2, vew007 (2016).

66. A. Rambaut, M. Suchard, A. Drummond, Tracer, (available at
835 http://tree.bio.ed.ac.uk/software/tracer/).

67. E. Willerslev, A. Cooper, Ancient DNA. *Proc. Biol. Sci.* 272, 3–16 (2005).

68. C. Warinner, A. Herbig, A. Mann, J. A. Fellows Yates, C. L. Weiß, H. A. Burbano, L. Orlando, J. Krause, A Robust Framework for Microbial Archaeology. *Annu. Rev. Genomics Hum. Genet.* 18, 321–356 (2017).

840 69. T. Bigot, S. Temmam, P. Pérot, M. Eloit, RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Research*. 8 (2019), p. 530.

70. C. Upton, S. Slack, A. L. Hunter, A. Ehlers, R. L. Roper, Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J. Virol.* 77, 7590–7600 (2003).

71. M. F. Boni, D. Posada, M. W. Feldman, An exact nonparametric method for inferring 845 mosaic structure in sequence triplets. *Genetics*. 176, 1035–1047 (2007).

72. D. P. Martin, D. Posada, K. A. Crandall, C. Williamson, A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retroviruses*. 21, 98–102 (2005).

73. D. Posada, K. A. Crandall, Evaluation of methods for detecting recombination from DNA 850 sequences: computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13757–13762 (2001).

74. M. Padidam, S. Sawyer, C. M. Fauquet, Possible emergence of new geminiviruses by frequent recombination. *Virology*. 265, 218–225 (1999).

75. J. M. Smith, Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129 (1992).

76. D. Martin, E. Rybicki, RDP: detection of recombination amongst aligned sequences. 855 *Bioinformatics*. 16, 562–563 (2000).

77. M. J. Gibbs, J. S. Armstrong, A. J. Gibbs, Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 16, 573–582 (2000).

78. L. Czech, A. Stamatakis, Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples. *PLoS One*. 14, e0217050 (2019).

860 79. F. Mahé, C. de Vargas, D. Bass, L. Czech, A. Stamatakis, E. Lara, D. Singer, J. Mayor, J. Bunge, S. Sernaker, T. Siemensmeyer, I. Trautmann, S. Romac, C. Berney, A. Kozlov, E. A. D. Mitchell, C. V. W. Seppey, E. Egge, G. Lentendu, R. Wirth, G. Trueba, M. Dunthorn, Parasites

dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol*. 1, 91 (2017).

865 80. C. Gruber, E. Giombini, M. Selleri, S. Tausch, A. Andrusch, A. Tyshaieva, G. Cardeti, R. Lorenzetti, L. De Marco, F. Carletti, A. Nitsche, M. Capobianchi, G. Ippolito, G. Autorino, C.

Castilletti, Whole Genome Characterization of Orthopoxvirus (OPV) Abatino, a Zoonotic Virus Representing a Putative Novel Clade of Old World Orthopoxviruses. *Viruses*. 10 (2018), p. 546.

81. R. E. Kass, A. E. Raftery, Bayes Factors. *J. Am. Stat. Assoc.* 90, 773 (1995).

870 82. W. M. Bass, *Human Osteology: A Laboratory and Field Manual* (Special Publications of the Missouri Archaeological Society, No. 2, ed. 4, 1995).

83. M. B. Brickley, Cribra orbitalia and porotic hyperostosis: A biological approach to diagnosis. *Am. J. Phys. Anthropol.* 167, 896–902 (2018).

84. P. L. Walker, R. R. Bathurst, R. Richman, T. Gjerdrum, V. A. Andrushko, The causes of
875 porotic hyperostosis and cribra orbitalia: a reappraisal of the iron-deficiency-anemia hypothesis. *Am. J. Phys. Anthropol.* 139, 109–125 (2009).

85. E. Naumann, T. D. Price, M. P. Richards, Changes in dietary practices and social organization during the pivotal late iron age period in Norway (AD 550-1030): isotope analyses of Merovingian and Viking Age human remains. *Am. J. Phys. Anthropol.* 155, 322–331 (2014).

880 86. D. Avdussin, Gnezdovo–der Nachbar von Smolensk. *Zeitschrift für Archdologie*, 263–290 (1977).

87. T. Puškina, V. Muraševa, N. Eniosova, Die Rus' im 9. – 10. Jahrhundert. Ein archaologisches Panorama. *Wachholtz Murmann, 2017*, 250–281.

88. M. Müller-Wille, Boat-graves in northern Europe. *Int. J. Naut. Archaeol.* 3, 187–204
885 (1974).

89. M. H. Beskow-Sjöberg, K.-H. & Arnell, *Ölands järnåldersgravfält. Volym I, Alböke, Köpings, Räpplinge, Löts, Egby, Bredsätra och Gärdslösa socknar* (Riksantikvarieämbetet och Statens historiska museer, Stockholm, 1987).

90. M. Beskow-Sjöberg, U. E. Hagberg, *Ölands järnåldersgravfält, Vol. II* (Högskolan i
890 Kalmar, Kalmar, 1991).

91. U. E. Hagberg, M. Beskow-Sjöberg, *Ölands järnåldersgravfält. Vol. III* (Riksantikvarieämbetet och Statens historiska museer, Stockholm, 1996).

92. J.-H. Fallgren, M. Rasch, *Ölands järnåldersgravfält. Vol. 4* (Riksantikvarieämbetet och Statens historiska museer., Stockholm, 2001).

895 93. J. E. Anderbjörk, Brev daterat 17/2 1939. Till Riksantikvarieämbetet. Dnr 0975 (1939).

94. T. D. Price, T. Douglas Price, J. Peets, R. Allmäe, L. Maldre, E. Oras, Isotopic provenancing of the Salme ship burials in Pre-Viking Age Estonia. *Antiquity*. 90, 1022–1037 (2016).

95. J. E. Anderbjörk, Nabberör i Böda-undersökningen av en märklig öländsk båtgrav.
900 *Kalmar Läns Fornminnesförening, meddelanden XXVII.* (1939).

96. H. Wilhelmson, *Perspectives from a Human-centred Archaeology: Iron Age People and Society on Öland* (2017).

97. H. Wilhelmson, T. Douglas Price, Migration and integration on the Baltic island of Öland in the Iron Age. *Journal of Archaeological Science: Reports*. 12, 183–196 (2017).

905 98. H. Wilhelmson, Shifting diet, shifting culture? A bioarchaeological approach to island dietary development on Iron-Age Öland, Baltic Sea. *Am. J. Phys. Anthropol.* 163, 264–284 (2017).

99. A. Margaryan, thesis, University of Copenhagen (2017).

100. P. Ø. Sørensen, Ed., in *Markildegård – En tidligneolitisk samlingsplads* (Kulturhistoriske
910 Studier I Sydsjællands Museum, Vordingborg, 1995), pp. 13–45.

101. P. Ø. Sørensen, Ed., in *Markildegård – mødested gennem årtusinder. 5000 år under motorvejen* (Vejdirektoratet og Rigsantikvarens Arkæologiske Sekretariat, København, 1994), pp. 32–33.

102. S. Wallis, *The Oxford Henge and Late Saxon Massacre: With Medieval and Later*
915 *Occupation at St John's College, Oxford* (2014).

103. T. D. Price, T. Douglas Price, M. Naum, P. Bennike, N. Lynnerup, K. M. Frei, H. Wagnkilde, T. Pind, F. O. Nielsen, Isotopic investigation of human provenience at the eleventh

century cemetery of Ndr. Grødbygård, Bornholm, Denmark. *Danish Journal of Archaeology*. 1 (2012), pp. 93–112.

920 104. M. Naum, *Homelands Lost and Gained: Slavic Migration and Settlement on Bornholm in the Early Middle Ages* (2008).

105. Y. Bäckström, T. Douglas Price, Social identity and mobility at a pre-industrial mining complex, Sweden. *J. Archaeol. Sci.* 66, 154–168 (2016).

106. I. Gustin, Contacts, Identity, and Hybridity: Objects from South-western Finland in the
925 Birka Graves. *Identity Formation and Diversity in the Early Medieval Baltic and Beyond* (2017), doi:10.1163/9789004328471_010.

107. S. Brooks, J. M. Suchey, Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* 5, 227–238 (1990).

108. Developmental Juvenile Osteology (2016), , doi:10.1016/c2009-0-63841-1.

930 109. U. E. Hagberg, *Archive report of excavations during 1965 in Sörby-Störlinge gravefield, fornl 74,Gärdslösa sn, Öland* (1966).

110. M. H. Beskow-Sjöberg, K.-H. Arnell, *Ölands järnåldersgravfält. Volym I* (Riksantikvarieämbetet och Statens historiska museer, Stockholm, 1987).

111. H. Clarke, Eketorp. Fortification and Settlement on Öland/Sweden. The Setting. Edited
935 by Ulf Näsman and Erik Wegraeus. 30 × 22 cm. Pp. 127 88 figs. Stockholm: Royal Academy of

Letters, History and Antiquities, 1979. Sw. Kr. 54. *The Antiquaries Journal*. 60 (1980), pp. 120–121.

112. B. Arrhenius, Knivar från Helgö och Birka. *Fornvännen*, 40–51 (1970).

113. T. Waldron, *Paleopathology* (Cambridge University Press, Cambridge, UK, 2008).

940 114. N. N. Gutsol, S. N. Vinogradova, A. G. Samorukova, *Kola Saami relocated groups* (Apatity: Kola Science Center RAS, 2007).

115. V. I. Khartanovich, in *Sbornik Muzeya antropologii i ehtnografii*, I. Gokhman, Ed. (Leningrad, 1980), vol. 36, pp. 35–47.

116. M. Krzewińska, G. Bjørnstad, P. Skoglund, P. I. Olason, J. Bill, A. Götherström, E.
945 Hagelberg, Mitochondrial DNA variation in the Viking age population of Norway. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130384 (2015).

117. S. Niinimäki, What do muscle marker ruggedness scores actually tell us? *International Journal of Osteoarchaeology*. 21, 292–299 (2009).

118. W. K. Seow, Developmental defects of enamel and dentine: challenges for basic science
950 research and clinical management. *Aust. Dent. J.* 59 Suppl 1, 143–154 (2014).

119. O. Murashko, N. Krenke, Aboriginal culture of the Obdorsk North in the XIX century (according to the archaeological and ethnographic collections of the Museum of Anthropology of Moscow State University). *Nauka*, 155 (2001).

120. D. J. Ortner, *Identification of Pathological Conditions in Human Skeletal Remains*
955 (Academic Press, 2003).

121. B. Å. Samuelsson, 'Ljungbacka, Lockarp 8 archive report' (Malmö museum, 1976).

122. B. Å. Samuelsson, Ljungbacka, a Late Iron Age Cemetery in Southwest Scania. *Lund Archaeological Review*, 89–108 (2001).

123. *Standards for Data Collection from Human Skeletal Remains* (Archaeological Survey
960 Research Series, Arkansas, U.S.A., 1994), vol. 44.

124. M. K. Moore, in *Research Methods in Human Skeletal Biology* (2013), pp. 91–116.

125. M. A. Kelley, Phenice's visual sexing technique for the os pubis: A critique. *Am. J. Phys. Anthropol.* 48, 121–122 (1978).

126. D. H. Ubelaker, C. G. Volk, A test of the phenice method for the estimation of sex. *J.*
965 *Forensic Sci.* 47, 19–24 (2002).

127. N. C. Lovell, Test of Phenice's technique for determining sex from the os pubis. *Am. J. Phys. Anthropol.* 79, 117–120 (1989).

128. J. Bruzek, A method for visual determination of sex, using the human hip bone. *Am. J. Phys. Anthropol.* 117, 157–168 (2002).

970 129. P. L. Walker, Greater sciatic notch morphology: sex, age, and population differences. *Am. J. Phys. Anthropol.* 127, 385–391 (2005).

130. J. L. Buckberry, A. T. Chamberlain, Age estimation from the auricular surface of the ilium: a revised method. *Am. J. Phys. Anthropol.* 119, 231–239 (2002).

131. S. M. Hens, M. G. Belcastro, Auricular surface aging: a blind test of the revised method
975 on historic Italians from Sardinia. *Forensic Sci. Int.* 214, 209.e1–5 (2012).

132. C. E. Merritt, The influence of body size on adult skeletal age estimation methods. *Am. J. Phys. Anthropol.* 156, 35–57 (2015).

133. K. Moraitis, E. Zorba, C. Eliopoulos, S. C. Fox, A test of the revised auricular surface aging method on a modern European population. *J. Forensic Sci.* 59, 188–194 (2014).

980 134. C. Rissech, J. Wilson, A. P. Winburn, D. Turbón, D. Steadman, A comparison of three established age estimation methods on an adult Spanish sample. *Int. J. Legal Med.* 126, 145–155 (2012).

135. M. San Millán, M. S. Millán, C. Rissech, D. Turbón, A test of Suchey–Brooks (pubic symphysis) and Buckberry–Chamberlain (auricular surface) methods on an identified Spanish
985 sample: paleodemographic implications. *J. Archaeol. Sci.* 40, 1743–1751 (2013).

136. C. A. Kunos, S. W. Simpson, K. F. Russell, I. Hershkovitz, First rib metamorphosis: its possible utility for human age-at-death estimation. *Am. J. Phys. Anthropol.* 110, 303–323 (1999).

137. L. Nilsson, 'Osteologisk Analys, Ljungbacka, Lockarp 8' (Malmö Museum, 1976).

138. J. M. Suchey, Instructions for use of the Suchey-Brooks system for age determination of
990 the female os pubis. Instructional materials accompanying female pubic symphysial models of the Suchey-Brooks system (1988).

139. C. Arcini, Prone burials. *Current archaeology*. 231, 30–35 (2008).

140. M. S. Toplak, Prone Burials and Modified Teeth at the Viking Age Cemetery of Kopparsvik: The Changing of Social Identities at the Threshold of the Christian Middle Ages.
995 *Analecta Archaeologica Ressoviensia*. 10 (2015).

141. E. Arwill-Nordbladh, Nyckelsymbolik i järnålderns kvinnogravar. *Fornvännen*. 85 (1990).

142. T. Zachrisson, in *I Att befolka det förflutna. Fem artiklar om hur vi kan synliggöra människan och hennes handlingar i arkeologiskt material*, A. Carlie, Ed. (Riksantikvarieämbetet,
1000 2014), pp. 73–91.

143. T. D. Price, T. Douglas Price, K. Prangsgaard, M. Kanstrup, P. Bennike, K. M. Frei, Galgedil: isotopic studies of a Viking cemetery on the Danish island of Funen, AD 800–1050. *Danish Journal of Archaeology*. 3, 129–144 (2014).

144. M. Kanstrup, in *Forhistorisk Arkeologi* (Aarhus Universitet, 2006).

1005 145. L. Melchior, T. Kivisild, N. Lynnerup, J. Dissing, Evidence of authentic DNA from Danish Viking Age skeletons untouched by humans for 1,000 years. *PLoS One*. 3, e2214 (2008).

146. J. D. Gardner, G. L. Smith, D. C. Tscharke, P. C. Reading, Vaccinia virus semaphorin A39R is a 50–55 kDa secreted glycoprotein that affects the outcome of infection in a murine intradermal model. *J. Gen. Virol.* 82, 2083–2093 (2001).

1010 147. J. B. Moore, G. L. Smith, Steroid hormone synthesis by a vaccinia enzyme: a new type of virus virulence factor. *EMBO J.* 11, 3490–3490 (1992).

148. M. P. Holgado, J. Falivene, C. Maeto, M. Amigo, M. F. Pascutti, M. B. Vecchione, A. Bruttomesso, G. Calamante, M. P. Del Médico-Zajac, M. M. Gherardi, Deletion of A44L, A46R and C12L Vaccinia Virus Genes from the MVA Genome Improved the Vector Immunogenicity

1015 by Modifying the Innate Immune Response Generating Enhanced and Optimized Specific T-Cell Responses. *Viruses*. 8 (2016), doi:10.3390/v8050139.

149. P. C. Reading, J. B. Moore, G. L. Smith, Steroid hormone synthesis by vaccinia virus suppresses the inflammatory response to infection. *J. Exp. Med.* 197, 1269–1278 (2003).

150. P. C. Reading, A. Khanna, G. L. Smith, Vaccinia virus CrmE encodes a soluble and cell
1020 surface tumor necrosis factor receptor that contributes to virus virulence. *Virology*. 292, 285–298 (2002).

151. K. Dai, Y. Liu, M. Liu, J. Xu, W. Huang, X. Huang, L. Liu, Y. Wan, Y. Hao, Y. Shao, Pathogenicity and immunogenicity of recombinant Tiantan Vaccinia Virus with deleted C12L and A53R genes. *Vaccine*. 26, 5062–5071 (2008).

1025 152. J. B. Eaglesham, Y. Pan, T. S. Kupper, P. J. Kranzusch, Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS-STING signalling. *Nature*. 566, 259–263 (2019).

153. A. H. Banham, G. L. Smith, Characterization of vaccinia virus gene B12R. *J. Gen. Virol.* 74 (Pt 12), 2807–2812 (1993).

154. A. Alcami, G. L. Smith, A soluble receptor for interleukin-1β encoded by vaccinia virus:
1030 A novel mechanism of virus modulation of the host response to infection. *Cell*. 71, 153–167 (1992).

155. R. Liu, B. Moss, Vaccinia Virus C9 Ankyrin Repeat/F-Box Protein Is a Newly Identified Antagonist of the Type I Interferon-Induced Antiviral State. *J. Virol.* 92 (2018), doi:10.1128/JVI.00053-18.

1035 156. M. T. Harte, I. R. Haga, G. Maloney, P. Gray, P. C. Reading, N. W. Bartlett, G. L. Smith, A. Bowie, L. A. J. O'Neill, The poxvirus protein A52R targets Toll-like receptor signaling complexes to suppress host defense. *J. Exp. Med.* 197, 343–351 (2003).

157. A. Bowie, E. Kiss-Toth, J. A. Symons, G. L. Smith, S. K. Dower, L. A. O'Neill, A46R and A52R from vaccinia virus are antagonists of host IL-1 and toll-like receptor signaling. *Proc.*
1040 *Natl. Acad. Sci. U. S. A.* 97, 10162–10167 (2000).

158. S. C. Graham, M. W. Bahar, S. Cooray, R. A. -J. Chen, D. M. Whalen, N. G. A. Abrescia, D. Alderton, R. J. Owens, D. I. Stuart, G. L. Smith, J. M. Grimes, Vaccinia Virus

Proteins A52 and B14 Share a Bcl-2–Like Fold but Have Evolved to Inhibit NF-κB rather than Apoptosis. *PLoS Pathog.* 4, e1000128 (2008).

1045 159. K. H. Martin, D. W. Grosenbach, C. A. Franke, D. E. Hruby, Identification and analysis of three myristylated vaccinia virus late proteins. *J. Virol.* 71, 5218–5226 (1997).

160. V. I. Chernos, T. S. Vovk, O. N. Ivanova, T. P. Antonova, V. N. Loparev, [Insertion mutants of the vaccinia virus. The effect of inactivating E7R and D8L genes on the biological properties of the virus]. *Mol. Gen. Mikrobiol. Virusol.*, 30–34 (1993).

1050 161. N. Price, D. C. Tscharke, M. Hollinshead, G. L. Smith, Vaccinia virus gene B7R encodes an 18-kDa protein that is resident in the endoplasmic reticulum and affects virus virulence. *Virology*. 267, 65–79 (2000).

162. S. N. Shchelkunov, Orthopoxvirus genes that mediate disease virulence and host tropism. *Adv. Virol.* 2012, 524743 (2012).

1055 163. J. L. Shisler, J. Xiao-Lu, The Vaccinia Virus K1L Gene Product Inhibits Host NF-κB Activation by Preventing I B Degradation. *J. Virol.* 78, 3553–3560 (2004).

164. K. E. Rehm, R. F. Connor, G. J. B. Jones, K. Yimbu, R. L. Roper, Vaccinia virus A35R inhibits MHC class II antigen presentation. *Virology*. 397, 176–186 (2010).

165. D. Wilcock, P. Traktman, G. L. Smith, S. A. Duncan, W.-H. Zhang, The vaccinia virus
1060 A40R gene product is a nonstructural, type II membrane glycoprotein that is expressed at the cell surface. *J. Gen. Virol.* 80, 2137–2148 (1999).

166. D. C. Tscharke, P. C. Reading, G. L. Smith, Dermal infection with vaccinia virus reveals roles for virus proteins not seen using other inoculation routes. *J. Gen. Virol.* 83, 1977–1986 (2002).

1065 167. M. A. Pallett, H. Ren, R.-Y. Zhang, S. R. Scutts, L. Gonzalez, Z. Zhu, C. M. de Motes, G. L. Smith, Vaccinia Virus BBK E3 Ligase Adaptor A55 Targets Importin-Dependent NF-κB

Activation and Inhibits CD8 T-Cell Memory. *Journal of Virology*. 93 (2019), , doi:10.1128/jvi.00051-19.

168. G. L. Smith, Y. S. Chan, S. T. Howard, Nucleotide sequence of 42 kbp of vaccinia virus
1070 strain WR from near the right inverted terminal repeat. *J. Gen. Virol.* 72 (Pt 6), 1349–1376 (1991).

169. A. Murcia-Nicolas, G. Bolbach, J. C. Blais, G. Beaud, Identification by mass spectroscopy of three major early proteins associated with virosomes in vaccinia virus-infected cells. *Virus Res.* 59, 1–12 (1999).

1075 170. WHO smallpox fact sheet (2012), (available at https://web.archive.org/web/20120901000000*/http://www.who.int/mediacentre/factsheets/smallpox/en/index.html).

171. WHO, WHO Hepatitis B fact sheet (2019), (available at https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b).

1080 172. A. G. Carmichael, A. M. Silverstein, Smallpox in Europe before the seventeenth century: virulent killer or benign disease? *J. Hist. Med. Allied Sci.* 42, 147–168 (1987).

173. Y. Li, D. S. Carroll, S. N. Gardner, M. C. Walsh, E. A. Vitalis, I. K. Damon, On the origin of smallpox: correlating variola phylogenics with historical smallpox records. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15787–15792 (2007).

1085 174. B. Mühlemann, T. C. Jones, M. Sikora, Processed data for phylogenetic analysis of ancient variola virus samples. FigShare dataset https://doi.org/10.6084/m9.figshare.12185466 (2020).

1090 **Acknowledgements**

EW thanks St. John's College, Cambridge, for providing an excellent environment for scientific discussions. We thank Stuart Rankin and the staff of the University of Cambridge High Performance Computing service, the National High-throughput DNA Sequencing Centre (Copenhagen, seqcenter.ku.dk), Leena Drenzel (Curator at the Historical Museum in Stockholm,
1095 Sweden), and Samuel Thelin and Catharina Ödman (archaeologists and curators, Malmö Museum, Sweden). We thank the three reviewers for their suggestions and constructive criticism. **Funding**: The Danish National Research Foundation; The Danish National Advanced Technology Foundation (The Genome Denmark platform, Grant 019-2011-2); The Lundbeck Foundation; The Novo Nordic Foundation; The Carlsberg Foundation; The Villum Kann
1100 Rasmussen Foundation, KU2016; Deutsches Zentrum für Infektionsforschung; LMS is funded by the Norwegian University of Science and Technology, the Royal Norwegian Society of Sciences and Letters and Sparebank 1 SMN; GLS is a Wellcome Trust Principal Research Fellow; EW was supported by the Lundbeck Foundation, the Danish National Research Foundation, the Wellcome Trust [214300/B/18/Z], and KU2016; BM, CD, RAMF, DJS, and
1105 TCJ were supported by the European Union Horizon 2020 Research and Innovation Programme, COMPARE (Grant agreement 643476); TCJ was supported by the Wellcome Trust [214300/B/18/Z]. **Author contributions:** BM: Initiated the study, sample identification and data processing, phylogenetic analysis, gene inactivation analysis, visualisations, consensus sequences, wrote manuscript (initial draft, reviewing, and editing); LV: Design of capture array.
1110 All virus sequence-targeted data production. Contribution to data analysis and manuscript; AM: Conducted sampling and generated sequence data, reviewing and editing of manuscript; HW: Osteological analysis, reviewing and editing of manuscript, wrote site descriptions; CFC: Conducted sampling and generated sequence data, reviewing and editing of manuscript; MEA: Supervising data production, reviewing and editing of manuscript; PBD: Conducted sampling
1115 and generated sequence data, reviewing and editing of manuscript; AJH: Design of capture array; SHN: Initial data collection, processing, and identification of positive samples; LMS: Osteological analysis, wrote site descriptions; JB: selection of material, reviewing and editing of manuscript, wrote site description; AB, TP, CF, VK, VM, MLSJ, PØS: Osteological analysis, reviewing and editing of manuscript, wrote site description; YM: Contribution to site

description; IG: Contribution to site descriptions; HS: Conducted sampling and generated sequence data, reviewing and editing of manuscript; GS: Analysis of results, reviewing and editing of manuscript; GLS: Analysis of results, reviewing and editing of manuscript; CD: Analysis of results, reviewing and editing of manuscript; RAMF: Analysis of results, reviewing and editing of manuscript; DJS: Analysis of results, reviewing and editing of manuscript; EW: Initiated the study, led efforts related to sampling and data production, participated in discussion and interpretation of results and writing; TCJ: Initiated the study, sample identification and data processing, phylogenetic analysis, recombination analysis, read specificity analysis, wrote manuscript (initial draft, reviewing, and editing); MS: Initiated the study, sample identification and data processing, ancient DNA data authentication, lower-coverage sample phylogenetic analyses, edit distance analysis, coverage, damage figures, reviewing and editing of manuscript. **Competing interests:** Eske Willerslev is involved in scientific collaboration with Illumina, Inc. **Data and materials availability:** Sequencing data and consensus genomes are available in the European Nucleotide Archive under project number PRJEB38129. Data for sequence alignments, BEAST2 XML files, RAxML-NG output for the ML tree of the higher-coverage sequences, consensus sequences of the lower-coverage samples used in EPA analysis, Python code and data for producing figures and gene loss analysis, and SNP tables for each lower-coverage sample, showing status (derived/ancestral/missing) at all branch informative SNPs are available via DOI 10.6084/m9.figshare.12185466 (*174*).

Supplementary Materials

Supplementary Text

Figs. S1 to S27

Tables S1 to S17 (S1, S3, S8, and S14 are data tables)

References 82 to 173

Tables and Figures

Fig. 1. Geographic location and phylogenetic placement of samples with reads matching VARV. A) Map showing geographic locations of samples. The locations of the higher-
1150 coverage and lower-coverage samples are indicated with solid and open red circles, respectively. The location of the published VD21 sample is shown in blue. Grey dots indicate locations of screened samples in which VARV was not detected. **B) Condensed maximum likelihood tree showing the higher-coverage samples and the subsequent EPA placement of the lower-coverage samples.** Lower-coverage samples were placed onto the tree containing the higher-
1155 coverage samples using EPA-ng (*27*). Clades that do not have any lower-coverage samples placed on them are collapsed and indicated by black circles. Colour is used to indicate the branch with the highest likelihood weight ratio assigned by EPA-ng for each lower-coverage sample. The full tree is shown in Fig. S6. Complete phylogenetic trees, showing the placement of each sample, individually and combined, are shown in Figs. S7 and S8, respectively.

1160 **Fig. 2. Gene-inactivating mutations in mVARV, aVARV, CMLV, and TATV sequences, with inferred gene status for internal nodes.** The cladogram on the left shows the topology of the phylogenetic relationship between mVARV, the aVARV sequences we report in this paper, TATV, and CMLV (see trees in Figs. 1B, S13, and S15). Each row to the right indicates gene status (present, inactivated) in the virus at that level in the cladogram (bottom eight rows), or in
1165 an inferred internal node or the root (top seven rows). The 40 columns represent genes that are either absent or that have an inactivating mutation in at least one of the mVARV or aVARV sequences, excluding 19 genes that are absent in both mVARV and aVARV. Genes are sorted left-to-right by category and then by position in the genome. Empty circles indicate genes assumed to be present and functional, while coloured circles indicate genes with a gene-
1170 inactivating mutation, genes that are absent, or genes that do not have coverage in the ancient sequences (also indicated by an 'X'). Within a column, genes with identical gene-inactivating mutations are shown in the same colour. Colour correspondences between columns carry no meaning. Genes with partial coverage (in the aVARV sequences) are marked with a black dot. A horizontal bar across a filled circle indicates a gene-inactivating mutation considered uncertain,
1175 due to less than three reads covering the position of the mutation. A horizontal bar across an empty circle indicates a gene which may be functional (with a length intermediate between its

length in viruses where it is functional and viruses where it is not). Gene labels indicate either the name of the VACV-COP homolog or the final offset of the gene in CPXV-Gri/GER (*6*). We inferred gene status at internal nodes using the DELTRAN maximum parsimony algorithm
1180 implemented in PastML (*48*). Genes are organised into four groups: A) inactivated in all mVARV and all aVARV sequences; B) inactivated in mVARV, but present in some or all aVARV sequences; C) present in mVARV, but inactivated in some aVARV sequences; D) present in mVARV and aVARV (not shown in the figure). See Supplementary Text for further information and the description of sub-groups.

1185 **Fig. 3. Gene inactivation over time.** The number of absent genes and genes with inactivating mutations is shown on the y-axis, plotted against time (in years into the past) on the x-axis. The evolutionary relationships between internal nodes (blue dots) and the tips of the phylogenetic tree (blue squares with white outline), as given by the ML trees in Figs. 1B and S13, are indicated by blue lines. Point estimates for dates for the aVARV sequences and the internal
1190 nodes are taken from the dated coalescent tree in Fig. S15, as estimated using BEAST2 (25) with a relaxed log-normal clock and a Bayesian skyline population prior. Gene counts for internal nodes are inferred based on gene status in their descendant viruses (see Fig. 2). Blue squares plotted vertically below tips show the number of gene-inactivating mutations that can be identified with certainty for the corresponding tip sample, as well as genes with no coverage in
1195 the ancient samples, and genes where the presence of a gene-inactivating mutation is uncertain (Materials and Methods). The dashed dark-grey line and surrounding grey area show a linear least squares regression and 95% confidence region, with gene counts from the inferred internal nodes included in the calculation. The x-axis intercept (not shown) is 4017 years into the past. The other parameters of the regression are: slope: 0.014, R: 0.91, R^2: 0.84, y-Intercept: 58.8,
1200 Stderr: 0.002, P-value: 0.0001. Sample names have been shortened as follows for conciseness: ‘aVARV-’ omitted from the Viking age names (aVARV-VK281, aVARV-VK382, aVARV-VK388, and aVARV-VK470) and ‘VARV-’ omitted from VARV-VD21. See Supplementary Text for comments regarding modelling gene inactivation over time.

1205

Sample	Date (CE)	Site	Sex	Age (years)	Sample type	TATV reads	TATV genome coverage (%)	TATV mean coverage depth	VARV reads	VARV genome coverage (%)	VARV mean coverage depth
VK382	640–770	Öland, SWE	M	Over 15	Tooth	85,036	91.079	40.723	88,324	99.564	45.188
VK388	603–653	Nordland, NOR	M	12–17	Tooth	20,119	90.396	7.082	20,629	98.999	7.755
VK470	950–1000	Gnezdovo, RUS	F	25–35	Tooth	18,894	89.522	6.944	19,117	97.114	7.500
VK281	885–990	Zealand, DEN	M	30–35	Tooth	10,801	88.358	4.594	11,006	95.885	5.012
VK168	880–1000	Oxford, GBR	M	16–25	Petrous	8876	77.065	2.397	8798	84.008	2.553
VK533	800–1050	Öland, SWE	*	50–60	Tooth	590	18.694	0.211	610	20.712	0.233
VK515	664–768	Nordland, NOR	M	18–22	Tooth	648	12.143	0.149	639	12.975	0.157
VK255	950–1000	Gnezdovo, RUS	F	20–25	Tooth	109	2.572	0.027	101	2.583	0.027
FIN1	1800–1900	Varzino, RUS	F	40–45	Tooth	83	1.869	0.021	95	2.275	0.025
KHA1	1800–1900	Yamalo-Nenetskiy, RUS	N/D	3–4	Petrous	58	1.763	0.018	58	1.932	0.019
VK443	800–1050	Öland, SWE	M	20–23	Tooth	88	1.719	0.018	85	1.767	0.018
VK108	800–1000	Malmö, SWE	F	55-75	Tooth	69	1.413	0.015	70	1.537	0.017
VK138	~1000	Fyn, DEN	M	25-35	Tooth	49	1.327	0.013	49	1.389	0.014

Table 1: Overview of samples with reads matching orthopoxviruses. From left to right, columns give the sample name, approximate sample date based on either carbon-14 (14C) dating (VK382, VK388, VK515) or archaeological context (all other samples), the site where the

sample was found (DEN: Denmark, GBR: Great Britain, NOR: Norway, RUS: Russia, SWE: Sweden), the sex of the individual, the age of the individual at the time of death, the type of tissue that was sequenced, the number of reads that aligned to TATV (GenBank accession number NC_008291.1) using bowtie2 --local and with mapping quality of at least 30, the percentage of the TATV genome covered (Fig. S3), the mean TATV genome coverage depth, the number of reads that aligned to VARV (GenBank accession number NC_001611.1) using bowtie2 --local with mapping quality of at least 30, the percentage of the VARV genome covered, and the mean VARV genome coverage depth (Fig. S3). Rows are ordered by decreasing mean coverage depth (Table S1). Note that the aVARV sequences are neither VARV (185,578 nt) nor TATV (198,050 nt), these are just the most logical similar modern reference sequences to align against (Table S4). In fact, the aVARV-VK382 consensus contains 192,255 nt and is 99.98% complete, with only 41 unresolved nucleotides. Eleven of the samples date from the Viking Age (with sample names starting with 'VK') or just prior to it and are from Scandinavian countries, western Russia, or the UK. The 14C (2 sigma) age ranges for VK382, VK388, and VK515 were calibrated with IntCal 13 (February 2020) (*47*) with no correction made for possible marine reservoir effect. The horizontal line separates the four higher-coverage samples from the nine lower-coverage. Site descriptions, with additional sample information, are given in the Supplementary Text. *VK533 was identified as female by osteology, but is genetically male. N/D: not determined.