

Research



Cite this article: Ferrari G, Neukamm J, Baalsrud HT, Breidenstein AM, Ravinet M, Phillips C, Rühli F, Bouwman A, Schuenemann VJ. 2020 Variola virus genome sequenced from an eighteenth-century museum specimen supports the recent origin of smallpox. *Phil. Trans. R. Soc. B* **375**: 20190572. <http://dx.doi.org/10.1098/rstb.2019.0572>

Accepted: 10 June 2020

One contribution of 14 to a theme issue 'Insights into health and disease from ancient biomolecules'.

Subject Areas:

evolution, genetics, health and disease and epidemiology, genomics

Keywords:

virus evolution, smallpox, ancient DNA, metagenomics, museum specimens

Authors for correspondence:

Giada Ferrari
e-mail: giada.ferrari@ibv.uio.no
Judith Neukamm
e-mail: judith.neukamm@uzh.ch
Abigail Bouwman
e-mail: abigail.bouwman@uzh.ch
Verena J. Schuenemann
e-mail: verena.schuenemann@iem.uzh.ch

†These authors contributed equally to this study.

‡These authors jointly supervised this study.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5092246>.

Variola virus genome sequenced from an eighteenth-century museum specimen supports the recent origin of smallpox

Giada Ferrari^{1,2,†}, Judith Neukamm^{2,3,†}, Helle T. Baalsrud¹, Abigail M. Breidenstein², Mark Ravinet^{1,4}, Carina Phillips⁵, Frank Rühli², Abigail Bouwman^{2,‡} and Verena J. Schuenemann^{2,‡}

¹Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316, Oslo, Norway

²Institute of Evolutionary Medicine, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

³Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany

⁴School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

⁵The Royal College of Surgeons of England, 35-43 Lincoln's Inn Fields, London WC2A 3PE, UK

id GF, 0000-0002-0850-1518; JN, 0000-0001-8141-566X; HTB, 0000-0002-4161-3247; AMB, 0000-0002-1675-8399; MR, 0000-0002-2841-1798; VJS, 0000-0002-8593-3672

Smallpox, caused by the variola virus (VARV), was a highly virulent disease with high mortality rates causing a major threat for global human health until its successful eradication in 1980. Despite previously published historic and modern VARV genomes, its past dissemination and diversity remain debated. To understand the evolutionary history of VARV with respect to historic and modern VARV genetic variation in Europe, we sequenced a VARV genome from a well-described eighteenth-century case from England (specimen P328). In our phylogenetic analysis, the new genome falls between the modern strains and another historic strain from Lithuania, supporting previous claims of larger diversity in early modern Europe compared to the twentieth century. Our analyses also resolve a previous controversy regarding the common ancestor between modern and historic strains by confirming a later date around the seventeenth century. Overall, our results point to the benefit of historic genomes for better resolution of past VARV diversity and highlight the value of such historic genomes from around the world to further understand the evolutionary history of smallpox as well as related diseases.

This article is part of the theme issue 'Insights into health and disease from ancient biomolecules'.

1. Introduction

Smallpox was a highly contagious and lethal disease [1,2]. Before its eradication—declared in 1980 AD [1]—smallpox caused several large-scale epidemics that spanned centuries with remarkably high death rates [1,2]. For instance, between 1900 and 1980 AD, smallpox was responsible for an estimated 300–500 million deaths, or one in ten global deaths [3]. The causative agent of smallpox was the variola virus (VARV), a member of the genus *Orthopoxvirus* [1,4,5]. The variola virus is thought to have emerged fairly recently, around 3000–4000 years ago [6–9]. Historically, possible accounts of smallpox-like diseases have been recorded in 1122 BC China and 1500 BC India and rashes consistent with a smallpox infection have been observed in ancient Egyptian mummies dating to 1580–1100 BC [1,2,10]. The earliest unmistakable descriptions of smallpox, however, can first be found in the fourth century AD China, seventh century AD India and the Mediterranean, and tenth century AD southwestern Asia [1,2,11]. Despite the insights provided by historical accounts and molecular studies, the emergence of VARV and its subsequent evolution remain contested.

Ancient DNA (aDNA) studies can provide a new understanding of the emergence and evolution of diseases via the reconstruction of ancient pathogen genomes [12,13]. However, whereas several historical and ancient bacterial genomes are currently available, particularly for the causative agent of plague *Yersinia pestis* (e.g. [14–16]) or environmentally resistant mycobacteria (e.g. [17–21]), arguably less effort has been devoted to viral sequences. Early studies on historic viruses include, for example, the post-mortem detection of dolphin morbillivirus [22] and of the 1918 AD Spanish influenza virus [23,24]. Increased genomic resolution power later allowed for more in-depth phylogenetic analyses, e.g. of human immunodeficiency virus type 1 [25,26], and the reconstruction of complete RNA genomes, such as the 1918 AD pandemic influenza virus, [27,28] and barley stripe mosaic virus [29], as well as complete DNA genomes, such as hepatitis B virus [30–33].

Several attempts at obtaining VARV DNA sequences from historic tissues and materials have been undertaken [34] but only a few have been successful. Polymerase chain reaction (PCR) fragments isolated from seventeenth to eighteenth century AD Siberian mummies suggest a VARV origin approximately 2000 years ago [35]; however, the phylogenetic resolution that can be obtained with such short DNA sequences is limited. More recently, whole VARV genomes have been reconstructed from a seventeenth century AD Lithuanian child mummy [36] and from two specimens from the Czech National Museum in Prague dated to the nineteenth and early twentieth century AD [37]. Molecular clock analyses based on these genomes suggested somewhat contrasting dates for a common ancestor of all twentieth century AD circulating VARV strains, covering a time period between 1350 and 1645 AD [36–39]. Despite a lack of clear consensus on the exact dates, there is an agreement that the time of divergence between historic and twentieth century AD circulating strains predates the start of widespread vaccination in 1796 AD [1]. Furthermore, the divergence between P-I and P-II, the two clades which include all twentieth century AD VARV strains [11], also appears to predate modern vaccination, and the diversity within the modern clades seems to be as recent as the late nineteenth or early twentieth century AD. This suggests that VARV strains circulating at the time underwent a severe bottleneck as global smallpox vaccination programmes started at the turn of the twentieth century AD [1], resulting in the extinction of several older lineages [36].

During the eighteenth century AD, smallpox was endemic in Europe [3], with recorded increases in both frequencies of epidemics and mortality [40–42]. High mortality rates were recorded particularly for children, increasing from one in ten of all children burials in the second half of the seventeenth century AD to nearly one in three in the eighteenth century AD among London Quakers [43]. In order to capture more of the VARV diversity at this time and provide an additional calibration point for the dating of the common ancestor of modern and historic VARV strains, we computationally reconstructed an eighteenth century AD VARV genome from a museum specimen originally prepared by surgeon and anatomist John Hunter (1728–1793 AD) between 1760 and 1793 AD [44]. Phylogenetic analysis of our newly reconstructed historic VARV genome, together with all available historic and modern genomes, supports a recent common ancestor dated to the seventeenth century AD and suggests greater VARV diversity in the past than previously assumed. Our study also serves as an example of how increasing the number

of available historic or ancient genomes can provide better resolution in phylogenetic inferences.

2. Material and methods

(a) Sampling, DNA extraction, library preparation and sequencing

The sample used in this study was collected at the Hunterian Museum at the Royal College of Surgeons of England from specimen RCSHC/P 328, an ethanol-fixed infant leg embedded in liquid paraffin (figure 1a). The specimen is described as follows in the museum's online catalogue SurgiCat [44]: object name: 'Leg, smallpox, Cases of Small Pox, Mounted wet tissue'; description: 'Part of the leg from the same case as P327 showing smallpox lesions covering the surface of the limb. The epidermis has been reflected from the thigh to show the elevation of the lesions. This infant is likely to have contracted smallpox while *in utero*'. A thin cross section was sampled under clean conditions (i.e. tools and surfaces were cleaned with a sodium hypochlorite solution and protective equipment was worn to reduce contamination of the specimen by the workers). Only the cross section was transported to the University of Zurich. The tissue specimen was used in its entirety for DNA extraction and only DNA libraries are retained. The original specimen remained in the custody of the Royal College of Surgeons of England, from which it was not removed.

The DNA extraction was performed according to an adapted protocol by Devault and colleagues [46]. One hundred milligrams of tissue were homogenized using a sterile scalpel blade and incubated at 95°C and 1000 rpm for 5 min in 800 µl of extraction buffer (25 mM Tris-HCl, 5 mM CaCl₂, 25 mM sodium citrate, 2.5 mM ethylenediaminetetraacetic acid, 1% sodium dodecyl sulfate, 50 mM 1,4-dithiothreitol, 10 mM N-phenacylthiazolium bromide). After adding 80 µl of proteinase K (20 mg ml⁻¹), digestion was performed at 50°C and shaken for 24 h. Following centrifugation, the supernatant was extracted twice with a 25:24:1 phenol, chloroform and isoamyl alcohol mixture followed by a final chloroform step. DNA was isolated using QIAquick spin columns (QIAGEN), with two elutions in 30 µl elution buffer and reduced centrifugation speed (6–10 krpm) to prevent the loss of short DNA fragments. The DNA extraction was performed in an aDNA clean laboratory [47] following standard anti-contamination protocols [48–50] with parallel non-template control extractions.

Ten microliters of extract or extraction blank were used to generate Illumina sequencing libraries following a protocol optimized for aDNA [51,52] with modified P5 primers for compatibility with Illumina v4 sequencing chemistry. Two sample-specific indexes were added to each library via PCR amplification [52]. Blunt-end repair, adapter ligation and setup of indexing PCRs were performed in an aDNA clean laboratory. Six libraries were generated for sample P328 and non-template library blanks were generated in parallel. Indexed libraries went through 9–15 cycles of re-amplification in one or four 100 µl reactions containing one unit AccuPrime™ Pfx DNA polymerase (Thermo Fisher Scientific) or Herculase II Fusion DNA polymerase (Agilent), 1× AccuPrime™ Pfx reaction mix or Herculase II reaction buffer, 0.3–0.4 µM primers IS5 and IS6 [51] and 4–5 µl library template. Libraries were purified with MinElute spin columns (QIAGEN) following the manufacturer's instructions. Quantitative PCR [51] and analysis on an Agilent 2200 TapeStation were used to assess library quality and sequencing was performed on the HiSeq2500 and HiSeq4000 Illumina platforms with 2 × 125 + 7 + 7 or 2 × 75 + 8 + 8 cycles, respectively, by the Functional Genomics Center Zurich (Switzerland).

(b) Risk assessment

All laboratory work was carried out under conditions stipulated by Swiss federal regulations (Swiss Federal Act on Research involving

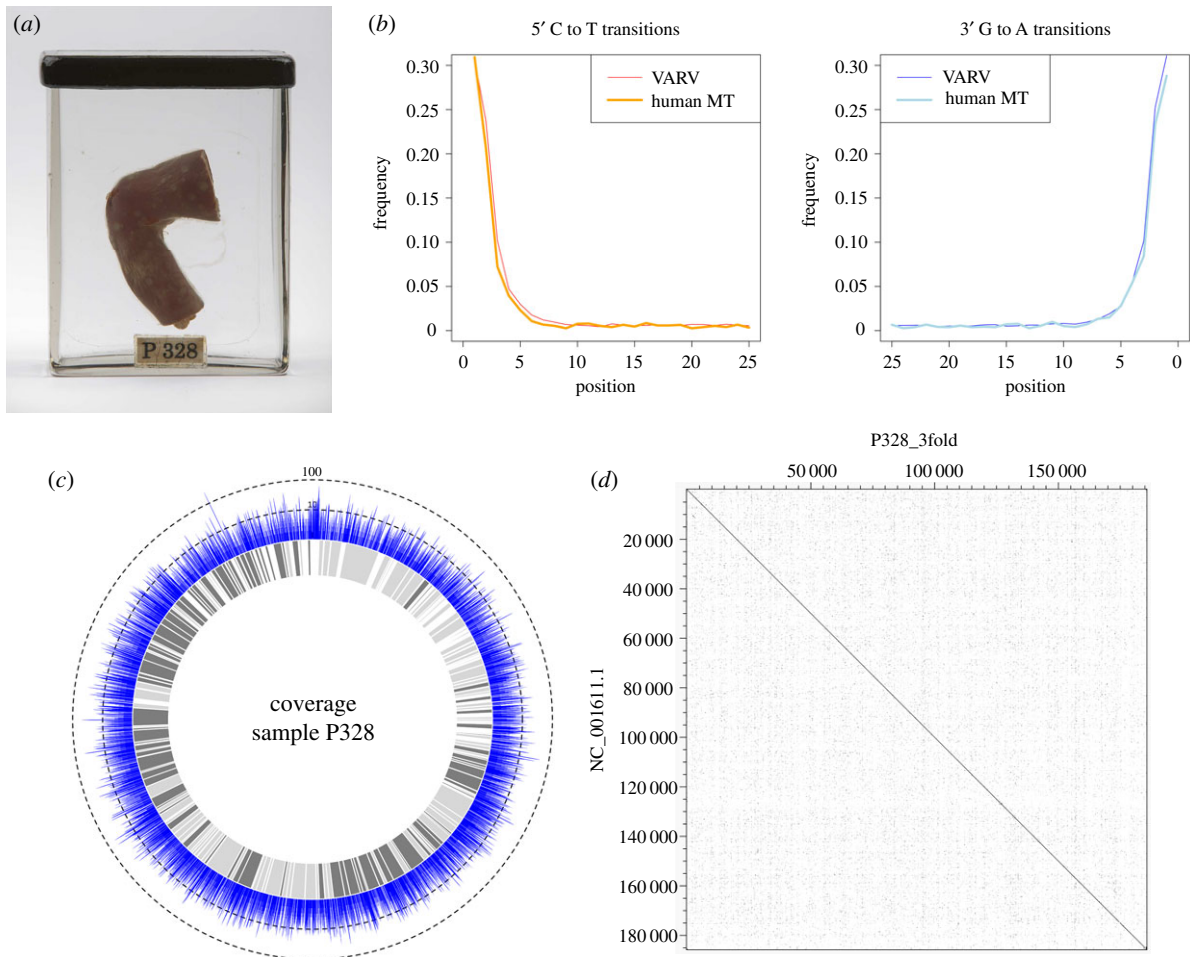


Figure 1. (a) Specimen RCSHC/P 328 showing smallpox lesions. Copyright © Museums at the Royal College of Surgeons. (b) Damage profile of reads mapping to the human mitochondrial genome (thick line) and the VARV reference genome (thin line). (c) Coverage diagram of the variola virus genome from sample P328. Blue indicates the coverage at a particular position (outer ring), coding areas of the genome are in grey (inner ring). The dotted circles indicate the coverage. (d) Conservation of genomic sequence between P328 and the variola virus reference genome NC_001611.1. The plot was generated with DOTTER [45]. (Online version in colour.)

human beings. Human Research Act, HRA, RS 810.30). Owing to the age of the specimen and conditions of preservation, any identifiable DNA was expected to be damaged, highly fragmented and non-infectious [53,54] and no additional precautions were therefore taken. Furthermore, the DNA libraries were retained in a non-infectious form that cannot be used for reassembly owing to extended ligands on the ends of fragments of genetic material. In detail, the preparation process involved the addition of synthetic adapter molecules [51,52], typically much longer than the DNA fragments of VARV themselves, which would prevent any possible *in vitro* reassembly of the genome.

We followed the recommendations of the World Health Organization (WHO) for handling DNA from VARV samples [55] with the following exceptions: (i) owing to the highly fragmented and damaged nature of the aDNA [53,54], no autoclaving step of the sample and its byproducts was performed; (ii) temporary retention of greater than 20% of the VARV genome in a non-infectious and fragmented form with full extraction of sample and no remaining original material; and (iii) handling of the ancient VARV DNA in our dedicated laboratory at the University of Zurich. The WHO has viewed our handling procedure (as it had been carried out at the time the work was done) through a detailed risk assessment and supported this publication.

(C) Data processing and analysis

(i) Metagenomic screening

The metagenomic screening of all six libraries of sample P328, as well as six non-template controls, was carried out with MALT

v. 0.4.1 [56] using all complete bacterial, viral and archaeal genomes available in GenBank [57] as a reference (version May 2018). MALT was executed with the following mapping parameters: only reads with a minimum 85% identity (`-minPercentIdentity`) were considered as a possible match to the reference. Moreover, the minimum support parameter (`-minSupport`) was set to 5, i.e. only nodes with minimum support of five reads were kept. BlastN mode and SemiGlobal alignment were applied and a top per cent value (`-topPercent`) of 1 was set. All other parameters were set to default. For more details, see the electronic supplementary material, file S1. All contaminant genera previously detected are excluded from the resulting metagenomic composition [58]. MALT results were analysed and visualized using MEGAN6 v. 6.12.6 [59]. The ancient origin of the reads mapping to prevalent genomes was assessed by calculating damage profiles consisting of cytosine to thymine and guanine to adenine base misincorporation at the 5' and 3' ends of the fragments, respectively, typical for aDNA [53], using DAMAGEPROFILER v. 0.3.12 [60].

(ii) Read processing, mapping and variant calling

All sequenced libraries were processed using EAGER v. 1.92.55 [61]. To summarize, the sequencing quality was inspected with FASTQC v. 0.11.5 [62], the reads were adapter trimmed and read-pairs merged into higher quality consensus sequences based on a minimum overlap of 10 bases with ADAPTERREMOVAL v. 2.2.1a [63] and subsequently aligned to the VARV reference genome (NC_001611.1) using BWA v. 0.7.17 [64] with a minimum quality score of 20 and a maximum edit distance of $n=0.01$. Duplicates were removed with MARKDUPLICATES v. 2.15.0 [65], and

DAMAGEPROFILER v. 0.3.12 [60] was used to investigate the damage patterns. The Genome Analysis Toolkit (GATK) v.3.8.0 [66,67] was used to generate a mapping assembly and single nucleotide polymorphism (SNP) calling. The reference base was called if the position was covered at least three times and the quality score was at least 30. The base was called as an SNP if the quality score was at least 30 and 90% of the mapped reads contained this variant.

Furthermore, EAGER v. 1.92.55 [61] was also applied to map all reads against the human reference genome (GRCh37.p13) using the same parameters for quality assessment, adapter trimming and mapping as described above. Modern human DNA contamination was calculated using schmutzi [68] and the haplogroup was determined with HAPLOGREP2 v. 2.1.19 [69].

(iii) Phylogeny

A whole-genome alignment was calculated based on the newly sequenced and computationally reconstructed genome of sample P328, 44 modern and three historic publicly available VARV genomes [37,39,70–72], as well as eight additional *Orthopoxvirus* genomes (camelpox virus, taterapox virus, cowpox virus, horsepox virus, monkeypox virus, raccoonpox virus, skunkpox virus and volepox virus) [72–77] and a horsepox virus genome reconstructed from a vaccine manufactured in 1902 AD [78]. All accession IDs are listed in the electronic supplementary material, table S1. The sequences were aligned with MAFFT v. 7.407 [79] using the FFT-NS-2 algorithm. Based on the resulting alignment and using all sites, a maximum-likelihood tree was calculated using RAxML v. 8 [80] with 100 bootstraps.

(iv) Beast analysis

As there are discrepancies about the dating of samples V563 and V1588 [37,38], we used the Bayesian framework BEAST 2.5.2 [81] to estimate the tip dates for the historic genomes based on the time intervals given in the electronic supplementary material, table S2. These intervals cover the years from the original publication [37] and the corresponding comment [38]. We used uniform distributions for P328 (1760–1793 AD) and VD21 (1643–1665 AD) and normal distributions for V563 (mean: 1925, s.d.: 20) and V1588 (mean: 1929, s.d.: 60). Next, BEAST 2.5.2 was used to estimate divergence times and substitution rate based on 48 published VARV genomes [37,39,70–72], including strain P328. We excluded all non-human strains. The analysis was performed using a strict molecular clock, the K81 substitution model (bMODELTEST [82]) and assuming constant population size [83] as tested best previously [36]. For modern samples, the isolation dates were used as tip dates. The Markov chain Monte Carlo was run with 100 000 000 iterations rejecting the first 10 000 000 as burn-in. For more detailed parameters, see the electronic supplementary material, files S2 and S3.

3. Results

(a) Sample information and processing

To computationally reconstruct the genome of an eighteenth century AD VARV strain, we collected a sample from a museum specimen with a known diagnosis of smallpox, specimen RCSHC/P 328 (later referred to as P328, figure 1a) from the Hunterian collection at the Royal College of Surgeons of England. The infant leg, whose surface is covered in smallpox lesions, was originally prepared by surgeon and anatomist John Hunter between 1760 and 1793 AD [44,84], and it is believed that the infant contracted the disease *in utero*. In 1780 AD, John Hunter published an account of a similar but unrelated case [85], where he describes the transmission of smallpox from a mother, who had recently recovered from

smallpox, to her stillborn child. Following DNA extraction, we generated six P328 DNA libraries for shotgun sequencing, obtaining a total of approximately 150 million raw reads.

(b) General metagenomic assessment

All P328 and non-template control libraries were processed the same way. First, all reads mapping to the human genome were excluded. The remaining reads were mapped against all available complete bacterial, viral and archaeal genomes in GenBank [57] using MALT [56] to monitor the metagenomic content in our dataset. The percentage of reads that could be assigned to the database per library varied from 0.49% to 31.11% (5.81% on average). The high number of unassigned reads may come from thus far unsequenced environmental bacteria. In general, P328 libraries are dominated by viruses (53.77%), followed by bacteria (45.47% on average) and archaea (0.76% on average) (electronic supplementary material, table S3). Moreover, we detected a high amount of reads mapping to Poxviridae (6.85%–24.38%), especially to the VARV genome (NC_001611.1) in all libraries except the non-template controls (electronic supplementary material, figure S1). To verify the ancient origin of the reads, the damage profiles of the reads mapping to VARV were determined (electronic supplementary material, figure S2). However, no specific metagenomic background could be assigned, which may be owing to the conservation of sample P328 in ethanol and paraffin. The top three species are plant-infecting viruses as well as bacteria specialized in the degradation of organic compounds such as *Methylothera versatilis* without reliable damage profiles. Lastly, we also detected bacteria that were identified as contaminants in sequenced blanks [58].

We then merged all P328 libraries into a single file and separately mapped it against the VARV genome (NC_001611.1) and the human reference genome (GRCh37.p13) (electronic supplementary material, table S4). Based on the reads mapping to the human genome (62 250 656 reads), we were able to reconstruct 95.42% of the human mitochondrial genome with a mean coverage of 19.62 X and a damage profile of 28.07% to 30.68% (figure 1b), verifying the ancient origin of the sample. Next, contamination was calculated using schmutzi [68], resulting in 1% of modern-day human contamination. The haplogroup H3b1b1, which is common in Europe, was determined using HAPLOGREP2 v. 2.1.19 [69]. Furthermore, the analysis of reads mapping to the VARV reference resulted in a damage profile of 29.76% to 31.32% (figure 1b), which is consistent with the damage profile of the reads mapping to the human genome, further confirming the ancient origin of the reads.

(c) Computational genome reconstruction and phylogeny

Using the merged shotgun sequencing data of all P328 libraries, 56 549 unique reads mapped to the VARV reference genome. Based on these reads, 85.18% of the VARV genome with a mean coverage of 14 X could be reconstructed with uniform coverage (figure 1c) using EAGER [61] (electronic supplementary material, table S4). Moreover, to investigate the synteny of strain P328, we compared it to the VARV reference genome, resulting in no major rearrangements and strong conservation of gene arrangements (figure 1d) [45].

To access the phylogenetic placement of strain P328, a maximum-likelihood tree was calculated based on the

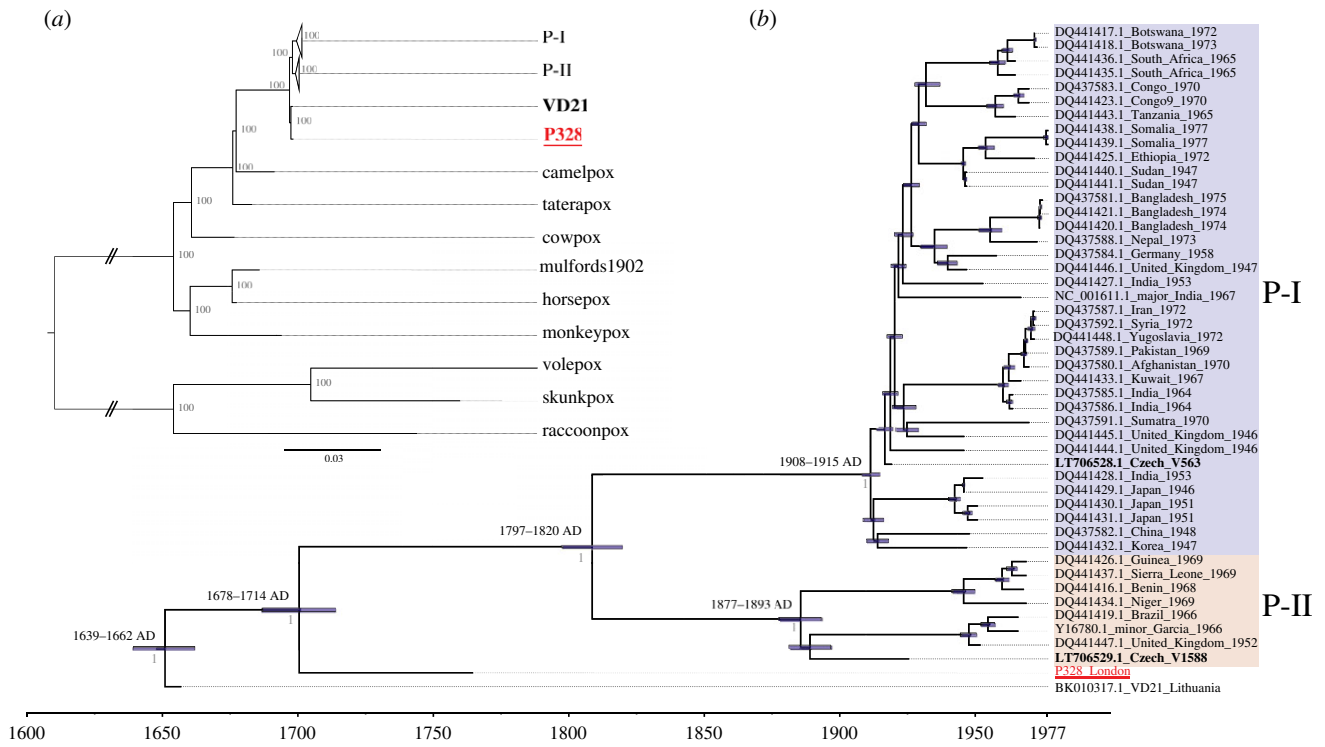


Figure 2. (a) Maximum-likelihood tree based on 57 *Orthopoxvirus* genomes. The historic genomes are bolded, the newly sequenced genome is in red and underlined. The bootstrap values are given as node labels in grey (100 BS). (b) Dated Bayesian maximum clade credibility tree reconstructed with BEAST 2.5.2 [81] (using a strict clock and constant population size). The nodes are labelled with the 95% HPD interval. Historic genomes are in bold, the newly added genome in red and underlined. Posterior values are given as node labels in grey. (Online version in colour.)

whole-genome alignment of the newly sequenced genome of sample P328, 44 modern and three historic VARV genomes that were publicly available [37,39,70–72], as well as eight additional *Orthopoxvirus* genomes (camelpox virus, taterapox virus, cowpox virus, horsepox virus, monkeypox virus, raccoonpox virus, skunkpox virus and volepox virus) [72–77] and a horsepox virus genome reconstructed from a vaccine manufactured in 1902 AD [78] (figure 2a, the complete tree is shown in the electronic supplementary material, figure S3). The newly reconstructed genome falls, together with historic strain VD21 [36,39], basal to all modern VARV strains, which is consistent with previous studies [36,37,39]. Additionally, the root splits the pox strains into New and Old World *Orthopoxviruses* [74].

Owing to discrepancies in the dating of two historical samples from the Czech National Museum (V1588 and V563) [37,38], we used BEAST 2.5.2 to estimate the tip dates based on the intervals given by the two studies. We used a normal distribution around 1925 AD for V563 and around 1929 AD for V1588 as suggested by Porter *et al.* [38], but allowed the priors to cover the entire age range suggested by the two studies. These analyses resulted in the year 1925 AD for V1588 and 1920 AD for V563. The divergence times of the viral genomes were estimated based on the whole-genome alignment of 44 modern and three historic publicly available VARV genomes [37,39,70–72], together with the newly sequenced P328 genome using BEAST 2.5.2 [81] (figure 2b and table 1). We used a strict molecular clock with a constant population size and substitution model K81. The mean evolutionary rate of VARV is estimated to be 10.67×10^{-6} nucleotide substitutions per site per year. P328, dated to 1766 AD, forms a sister group to all modern human strains [70–72], as well as to the historic Lithuanian strain VD21 [36,39] dated to 1656. The time to the most recent common ancestor of VD21, P328 and all twentieth century AD VARV

strains is estimated to 1651 AD (1639–1662, 95% highest posterior density (HPD)). The median divergence time of P328 and the modern strains is dated to 1701 AD (1687–1714 AD, 95% HPD). V563 and V1588 fall within the modern P-I and P-II clades, respectively, consistent with previous studies [37]. To test the robustness of the divergence time estimation, we reran the BEAST analysis excluding the strains V563 and V1588, resulting in almost identical divergence times (electronic supplementary material, figure S4).

4. Discussion

Here, we successfully computationally reconstructed an eighteenth century AD VARV genome at a mean coverage of 14X using shotgun metagenomic data. Because pathogen DNA typically represents only a very small fraction of an ancient or historic samples' metagenomic profile, hybridization capture is needed in most cases to reconstruct genomes (e.g. [14,36,46,86]). While this technique has the advantage of efficiently reconstructing pathogen genomes at a higher coverage and optimized sequencing costs, it can introduce bias, because evolutionary events such as genomic rearrangements, i.e. insertions, deletions, duplications, inversions and translocations, in addition to horizontal gene transfer, are likely to be missed when using extant genomes as a reference for hybridization probe design. Examples of genome reconstruction without enrichment for pathogen DNA exist [17,21,28,29,87–90] but represent a minority of aDNA studies. The P328 VARV genome we obtained without hybridization capture is highly consistent with the VD21 VARV genome obtained by Duggan *et al.* [36] using a capture approach, suggesting that this bias may be less of an issue, given the presence of strong genomic conservation such as for VARV [11,91].

Table 1. Comparison of the time to the most recent common ancestor (tMRCA) for the dated VARV phylogeny and individual branches for this and previously published studies [36,37,39]. (Dates are given in calendar years (AD). HPD, highest posterior density.)

branch splits, (AD)	this study		Duggan <i>et al.</i> [36]		Pajer <i>et al.</i> [37]	Smithson <i>et al.</i> [39]	
	mean tMRC	95% HPD	mean tMRCA	95% HPD	mean tMRCA	mean tMRCA	95% HPD
split VD21/P328/ modern VARV	1651	1639–1662	1617	1588–1645	1350	1517	1470–1563
split P328/modern VARV	1701	1687–1714	na	na	na	na	na
split P-I/P-II	1809	1797–1820	1764	1734–1793	1695	1623	1579–1667
split P-I internal	1911	1908–1915	1910	1902–1917	1887	1881	1861–1897
split P-II internal	1886	1877–1893	1870	1855–1885	1808	1794	1754–1828

This has also shown to be true for other ancient pathogen genomes, e.g. *Mycobacterium leprae* [17].

Besides investigating the species of interest, shotgun sequencing also provides the possibility to analyse the metagenomic composition of a sample by comparing it against a reference database. The top 10 identified species are dominated by plant-infecting viruses, such as the Dasheen mosaic virus, but these reads do not show a reliable damage profile. For the P328 libraries, the metagenomic analysis yielded a non-specific metagenomic background. This may be owing to a high content of unknown or unsequenced species, but also to the sample's conservation treatment in ethanol and paraffin and the lack of studies investigating the metagenomic content of this preservation method. Moreover, a relaxed identity parameter was applied when mapping to accommodate age-related DNA damage [53]. However, we clearly identified a high amount of reads mapping to the VARV genome, all showing a consistent damage profile, confirming the authenticity of the reads (electronic supplementary material, figure S2).

Phylogenetic analysis of the newly computationally reconstructed P328 VARV genome together with all available modern and historic VARV genomes [37,39,70–72], as well as eight modern and one historic *Orthopoxvirus* genomes [72–78], places P328 in a sister group to all twentieth century AD VARV strains together with the VD21 strain (figure 2a). This is consistent with the phylogenetic analysis reported by Duggan *et al.* [36]. Bayesian analysis assuming a strict molecular clock and constant population size suggests a common ancestor between 1639 and 1662 AD for all VARV strains included in this study [37,39,70–72]. This is consistent with the common ancestor suggested between modern strains and VD21, dated to between 1588 and 1645 AD [36], and clearly younger than the common ancestor dated to 1350 AD suggested by Pajer *et al.* [37].

Similarly, our phylogenetic analysis suggests a divergence between the P-I and P-II clades between 1797 and 1820 AD with P-I dating to between 1908 and 1915 AD and P-II to between 1877 and 1893 AD (figure 2b). While these dates are marginally younger than those reported by Duggan *et al.* [36], they are consistent with the hypothesis that the P-I and P-II clades diverged prior or contemporary to the development of smallpox vaccination in 1796 AD [1] (table 1) with evidence of a severe bottleneck following the rise in vaccination rates during the late nineteenth and early twentieth century AD.

Furthermore, this is consistent with the observation that currently available twentieth century AD VARV genomes probably represent only a fraction of VARV genetic diversity in the past, because several strains are believed to have disappeared or were no longer detected owing to a loss in virulence [3]. Of the four historic strains included in our study, the younger V563 and V1588 [37] fall within the diversity of modern VARV strains, clustering in the P-I and P-II clades, respectively. By contrast, the seventeenth century VD21 [36,39] and the eighteenth century P328 both form their own sister groups to the twentieth century strains. This observation is consistent with modern VARV strains not being representative of past viral diversity and agrees with the scenario that selective pressure from increasing levels of vaccination caused several VARV lineages to disappear [3,36]. However, acquiring additional ancient and historic VARV genomes is crucial to capture the true diversity of VARV prior to the development of smallpox vaccination.

Finally, we addressed the discrepancies in dating of strains V563 and V1588, which were obtained from two specimens from the Czech National Museum in Prague [37]. These specimens were dated by Pajer *et al.* [37] to 60 and 160 years ago, respectively, but were contested to be of similar age and dating to the 1920s AD [38]. We used a tip calibration with normal distribution around the proposed younger dates, while allowing for the possibility of the specimens to be the age originally proposed. We obtained a mean date of 1920 AD for V563 and 1925 AD for V1588, consistent with the dates proposed by Porter *et al.* [38]. Furthermore, our proposed seventeenth century AD dating of the common ancestor between modern and historic strains is much closer to the sixteenth- to seventeenth century AD dating proposed by Duggan *et al.* [36] and Porter *et al.* [38] than to the fourteenth century AD dating obtained by Pajer *et al.* [37] when assuming an older date for strain V1588. Furthermore, excluding V563 and V1588 from the analysis still yielded similar dates (electronic supplementary material, figure S4).

An additional study by Smithson *et al.* [39] proposed an improved assembly for VD21, which when used in phylogenetic analysis, pushed back the date of the common ancestor between VD21 and modern VARV strains to the late fifteenth or early sixteenth century AD and the divergence between P-I and P-II to the late sixteenth or early seventeenth century AD (table 1). Here, we used the newly improved assembly for VD21 and

obtained dates closer to those suggested by Duggan *et al.* [36] and older than those suggested by Smithson *et al.* [39].

The differing, and at times contested, dates assigned to VARV phylogeny presented here and in previous studies (e.g. [36,37,39]) show the importance of improving existing genomic assemblies, in addition to increasing the resolution power of such analyses by exploring and reconstructing additional historic genomes from an expanded spatio-temporal range. This is not only crucial to better understand pathogen evolution over a larger time transect and geographical distribution, but also to provide additional calibration points for phylogenetic analyses. To this end, metagenomic shotgun approaches, which do not use hybridization, such as the one presented here, can serve as a valuable tool to study a wide range of hosts and their associated viruses.

Ethics. The sample used in this study was obtained with the approval of the Board of Trustees of the Hunterian Collection and the Museums Department of the Royal College of Surgeons of England. The specimen is fully encoded and older than 70 years (post-mortem). Therefore, it does not require additional approval under Swiss law (Swiss Federal Act on Research involving human beings. Human Research Act, HRA art.1 and art.36; RS 810.30). The responsible ethics committee (Kantonale Ethikkommission Zürich, Switzerland) approved the use of this sample (application number KEK-ZH-Nr. 2014-0316).

Data accessibility. The raw sequencing data have been deposited at the European Nucleotide Archive (ENA) under project PRJEB35140 with accession IDs ERS3935826 - ERS3935837. The MALT parameters and the BEAST xml files used for our analyses have been uploaded as part of the electronic supplementary material.

Authors' contributions. G.F., J.N., C.P., F.R., A.B. and V.J.S. have been involved in conceptualization of the study; G.F., J.N. and A.B. developed the methodology; G.F., J.N., H.T.B. and M.R. carried out formal analysis; G.F. and A.M.B. have been involved in the investigation; C.P. provided resources; J.N. carried out the data curation; G.F. and J.N. wrote the original draft; G.F., J.N., H.T.B., A.M.B. and V.J.S. took care of the writing-review and editing; G.F. and J.N. carried out the visualization; F.R., A.B. and V.J.S. supervised the work; G.F., J.N., F.R., A.B. and V.J.S. took care of the project administration; F.R. and V.J.S. provided funding. All authors approved the final version of the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This study was funded by the University of Zurich's University Research Priority Program 'Evolution in Action: From Genomes to Ecosystems' (G.F., J.N., grants awarded to A.B., V.J.S., F.R.) as well as the Mäxi Foundation Zurich (A.M.B., A.B., awarded to F.R.). G.F. was also supported by the Research Council of Norway (project 262777).

Acknowledgements. We are grateful to Sam Alberti and the Board of Trustees of the Hunterian Collection at the Royal College of Surgeons of England for donating the sample used in this study and to Martyn Cooke for carrying out the sampling. We thank Sirisha Aluri, Jelena Kühn-Georgijevic, Catharine Aquino and Lennart Opitz at the Functional Genomics Center Zurich for their assistance with sequencing. We are also thankful to Claudia Viganò for assisting with laboratory work, Heidi Lischer for help with preliminary data analysis and Bastiaan Star for comments on the manuscript. The analyses were performed on the high-performance computing (HPC) cluster at the University of Zurich, operated by the Central Informatics at the University of Zurich and on the Abel Cluster at the University of Oslo and the Norwegian metacentre for High Performance Computing (NOTUR), operated by the Department for Research Computing at the University of Oslo IT-department (USIT).

References

- Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID. 1988 *Smallpox and its eradication*. Geneva, Switzerland: World Health Organization
- Hopkins DR. 2002 *The greatest killer: smallpox in history*. Chicago, IL: University of Chicago Press.
- Thèves C, Biagini P, Crubézy E. 2014 The rediscovery of smallpox. *Clin. Microbiol. Infect.* **20**, 210–218. (doi:10.1111/1469-0691.12536)
- Moss B. 2007 Poxviridae: the viruses and their replication. In *Fields virology* (eds DM Knipe, PM Howley), pp. 2905–2946. Philadelphia, PA: Lippincott, Williams & Wilkins.
- Moore ZS, Seward JF, Lane JM. 2006 Smallpox. *Lancet* **367**, 425–435. (doi:10.1016/S0140-6736(06)68143-9)
- Babkin IV, Babkina IN. 2015 The origin of the variola virus. *Viruses* **7**, 1100–1112. (doi:10.3390/v7031100)
- Hughes AL, Irausquin S, Friedman R. 2010 The evolutionary biology of poxviruses. *Infect. Genet. Evol.* **10**, 50–59. (doi:10.1016/j.meegid.2009.10.001)
- Shchelkunov SN. 2009 How long ago did smallpox virus emerge? *Arch. Virol.* **154**, 1865–1871. (doi:10.1007/s00705-009-0536-0)
- Babkin IV, Babkina IN. 2012 A retrospective study of the orthopoxvirus molecular evolution. *Infect. Genet. Evol.* **12**, 1597–1604. (doi:10.1016/j.meegid.2012.07.011)
- Dixon CW. 1962 *Smallpox*. London, UK: Churchill.
- Li Y, Carroll DS, Gardner SN, Walsh MC, Vitalis EA, Damon IK. 2007 On the origin of smallpox: correlating variola phylogenics with historical smallpox records. *Proc. Natl Acad. Sci. USA* **104**, 15 787–15 792. (doi:10.1073/pnas.0609268104)
- Marciniak S, Poinar HN. 2019 Ancient pathogens through human history: a paleogenomic perspective. In *Paleogenomics: genome-scale analysis of ancient DNA* (eds C Lindqvist, OP Rajora), pp. 115–138. Cham, Switzerland: Springer International Publishing.
- Bos KI *et al.* 2019 Paleomicrobiology: diagnosis and evolution of ancient pathogens. *Annu. Rev. Microbiol.* **73**, 639–666. (doi:10.1146/annurev-micro-090817-062436)
- Bos KI *et al.* 2011 A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510. (doi:10.1038/nature10549)
- Spyrou MA *et al.* 2018 Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234. (doi:10.1038/s41467-018-04550-9)
- Rascovan N, Sjögren K-G, Kristiansen K, Nielsen R, Willerslev E, Desnues C, Rasmussen S. 2019 Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic decline. *Cell* **176**, 295–305. (doi:10.1016/j.cell.2018.11.005)
- Schuenemann VJ *et al.* 2013 Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183. (doi:10.1126/science.1238286)
- Bos KI *et al.* 2014 Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497. (doi:10.1038/nature13591)
- Schuenemann VJ *et al.* 2018 Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. *PLoS Pathog.* **14**, e1006997. (doi:10.1371/journal.ppat.1006997)
- Sabin S, Herbig A, Vagene AJ, Ahlström T, Bozovic G, Arcini C, Kühnert D, Bos KI. 2019 A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. *BioRxiv*. (doi:10.1101/588277)
- Kay GL *et al.* 2015 Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat. Commun.* **6**, 6717. (doi:10.1038/ncomms7717)
- Krafft A, Lichy JH, Lipscomb TP, Klaunberg BA, Kennedy S, Taubenberger JK. 1995 Postmortem diagnosis of morbillivirus infection in bottlenose dolphins (*Tursiops truncatus*) in the Atlantic and Gulf of Mexico epizootics by polymerase chain reaction-based assay. *J. Wildl. Dis.* **31**, 410–415. (doi:10.7589/0090-3558-31.3.410)
- Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. 1997 Initial genetic characterization of the 1918 'Spanish' influenza virus. *Science* **275**, 1793–1796. (doi:10.1126/science.275.5307.1793)

24. Reid AH, Fanning TG, Hultin JV, Taubenberger JK. 1999 Origin and evolution of the 1918 'Spanish' influenza virus hemagglutinin gene. *Proc. Natl Acad. Sci. USA* **96**, 1651–1656. (doi:10.1073/pnas.96.4.1651)
25. Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. 2007 The emergence of HIV/AIDS in the Americas and beyond. *Proc. Natl Acad. Sci. USA* **104**, 18 566–18 570. (doi:10.1073/pnas.0705329104)
26. Worobey M *et al.* 2008 Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664. (doi:10.1038/nature07390)
27. Taubenberger JK, Hultin JV, Morens DM. 2007 Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antivir. Ther.* **12**, 581–591.
28. Xiao Y-L, Kash JC, Beres SB, Sheng Z-M, Musser JM, Taubenberger JK. 2013 High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J. Pathol.* **229**, 535–545. (doi:10.1002/path.4145)
29. Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. 2014 A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological barley stripe mosaic virus. *Sci. Rep.* **4**, 4003. (doi:10.1038/srep04003)
30. Kahila Bar-Gal G *et al.* 2012 Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* **56**, 1671–1680. (doi:10.1002/hep.25852)
31. Patterson Ross Z *et al.* 2018 The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathog.* **14**, e1006750. (doi:10.1371/journal.ppat.1006750)
32. Mühlemann B *et al.* 2018 Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423. (doi:10.1038/s41586-018-0097-z)
33. Krause-Kyora B *et al.* 2018 Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* **7**, e36666. (doi:10.7554/eLife.36666)
34. McCollum AM, Li Y, Wilkins K, Karem KL, Davidson WB, Paddock CD, Reynolds MG, Damon IK. 2014 Poxvirus viability and signatures in historical relics. *Emerg. Infect. Dis.* **20**, 177–184. (doi:10.3201/eid2002.131098)
35. Biagini P *et al.* 2012 Variola virus in a 300-year-old Siberian mummy. *N. Engl. J. Med.* **367**, 2057–2059. (doi:10.1056/NEJMc1208124)
36. Duggan AT *et al.* 2016 17th century variola virus reveals the recent history of smallpox. *Curr. Biol.* **26**, 3407–3412. (doi:10.1016/j.cub.2016.10.061)
37. Pajer P *et al.* 2017 Characterization of two historic smallpox specimens from a Czech Museum. *Viruses* **9**, 200. (doi:10.3390/v9080200)
38. Porter AF, Duggan AT, Poinar HN, Holmes EC. 2017 Comment: characterization of two historic smallpox specimens from a Czech Museum. *Viruses* **9**, 276. (doi:10.3390/v9100276)
39. Smithson C, Imbery J, Upton C. 2017 Re-assembly and analysis of an ancient variola virus genome. *Viruses* **9**, 253. (doi:10.3390/v9090253)
40. Davenport RJ, Boulton J, Schwarz L. 2016 Urban inoculation and the decline of smallpox mortality in eighteenth-century cities: a reply to Razzell. *Econo. Hist. Rev.* **69**, 188–214. (doi:10.1111/ehr.12112)
41. Duncan CJ, Duncan SR, Scott S. 1996 Oscillatory dynamics of smallpox and the impact of vaccination. *J. Theor. Biol.* **183**, 447–454. (doi:10.1006/jtbi.1996.0234)
42. Davenport RJ, Satchell M, Shaw-Taylor LMW. 2018 The geography of smallpox in England before vaccination: a conundrum resolved. *Soc. Sci. Med.* **206**, 75–85. (doi:10.1016/j.socscimed.2018.04.019)
43. Landers J. 1993 *Death and the metropolis: studies in the demographic history of London, 1670–1830*. Cambridge, UK: Cambridge University Press.
44. SurgiCat. 2019 Home page. See <http://surgicat.rcseng.ac.uk/> (accessed on 1 November 2019).
45. Sonhammer EL, Durbin R. 1995 A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**, GC1–10. (doi:10.1016/0378-1119(95)00714-8)
46. Devault AM *et al.* 2014 Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N. Engl. J. Med.* **370**, 334–340. (doi:10.1056/NEJMoa1308663)
47. Krüttli A, Bouwman A, Akgül G, Della Casa P, Rühli F, Warinner C. 2014 Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central Europe. *PLoS ONE* **9**, e86251. (doi:10.1371/journal.pone.0086251)
48. Cooper A, Poinar HN. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139. (doi:10.1126/science.289.5482.1139b)
49. Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. 2005 Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544. (doi:10.1016/j.tree.2005.07.005)
50. Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. 2017 From the field to the laboratory: controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Sci. Tech. Archaeol. Res.* **3**, 1–14. (doi:10.1080/20548923.2016.1258824)
51. Meyer M, Kircher M. 2010 Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448. (doi:10.1101/pdb.prot5448)
52. Kircher M, Sawyer S, Meyer M. 2012 Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3. (doi:10.1093/nar/gkr771)
53. Briggs AW *et al.* 2007 Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14 616–14 621. (doi:10.1073/pnas.0704665104)
54. Dabney J, Meyer M, Pääbo S. 2013 Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567. (doi:10.1101/cshperspect.a012567)
55. World Health Organization. 2016 WHO recommendations concerning the distribution, handling and synthesis of variola virus DNA. See <https://www.who.int/csr/disease/smallpox/variola-virus-dna/en/> (accessed on 22 June 2020).
56. Vågene ÅJ *et al.* 2018 *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528. (doi:10.1038/s41559-017-0446-6)
57. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013 GenBank. *Nucleic Acids Res.* **41**, D36–D42. (doi:10.1093/nar/gks1195)
58. Salter SJ *et al.* 2014 Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87. (doi:10.1186/s12915-014-0087-z)
59. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. 2016 MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957. (doi:10.1371/journal.pcbi.1004957)
60. Neukamm J, Peltzer A. 2018 Integrative-Transcriptomics/DamageProfiler v0.3.12. See <http://github.com/Integrative-Transcriptomics/DamageProfiler>. (doi:10.5281/zenodo.1288880)
61. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016 EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60. (doi:10.1186/s13059-016-0918-z)
62. Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. See <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
63. Schubert M, Lindgreen S, Orlando L. 2016 AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88. (doi:10.1186/s13104-016-1900-2)
64. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
65. Picard Toolkit. 2019 Broad Institute, Github Repository. See <http://broadinstitute.github.io/picard/>.
66. DePristo MA *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. (doi:10.1038/ng.806)
67. Auwera GA *et al.* 2013 From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11. (doi:10.1002/0471250953.bi1110s43)
68. Renaud G, Slon V, Duggan AT, Kelso J. 2015 Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224. (doi:10.1186/s13059-015-0776-0)
69. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, Schönherr S. 2016 HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63. (doi:10.1093/nar/gkw233)

70. Shchelkunov SN *et al.* 2000 Alastrim smallpox variola minor virus genome DNA sequences. *Virology* **266**, 361–386. (doi:10.1006/viro.1999.0086)
71. Shchelkunov SN, Totmenin AV, Sandakhchiev LS. 1996 Analysis of the nucleotide sequence of 23.8 kbp from the left terminus of the genome of variola major virus strain India-1967. *Virus Res.* **40**, 169–183. (doi:10.1016/0168-1702(95)01269-9)
72. Esposito JJ *et al.* 2006 Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science* **313**, 807–812. (doi:10.1126/science.1125134)
73. Afonso CL, Tulman ER, Lu Z, Zsak L, Sandybaev NT, Kerembekova UZ, Zaitsev VL, Kutish GF, Rock DL. 2002 The genome of camelpox virus. *Virology* **295**, 1–9. (doi:10.1006/viro.2001.1343)
74. Smithson C, Tang N, Sammons S, Frace M, Batra D, Li Y, Emerson GL, Carroll DS, Upton C. 2017 The genomes of three North American orthopoxviruses. *Virus Genes* **53**, 21–34. (doi:10.1007/s11262-016-1388-9)
75. Shchelkunov SN *et al.* 2001 Human monkeypox and smallpox viruses: genomic comparison. *FEBS Lett.* **509**, 66–70. (doi:10.1016/S0014-5793(01)03144-1)
76. Tulman ER *et al.* 2006 Genome of horsepox virus. *J. Virol.* **80**, 9244–9258. (doi:10.1128/JVI.00945-06)
77. Dabrowski PW, Radonić A, Kurth A, Nitsche A. 2013 Genome-wide comparison of cowpox viruses reveals a new clade related to variola virus. *PLoS ONE* **8**, e79953. (doi:10.1371/journal.pone.0079953)
78. Schrick L, Tausch SH, Dabrowski PW, Damaso CR, Esparza J, Nitsche A. 2017 An early American smallpox vaccine based on horsepox. *N. Engl. J. Med.* **377**, 1491–1492. (doi:10.1056/NEJMc1707600)
79. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
80. Stamatakis A. 2014 RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
81. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537. (doi:10.1371/journal.pcbi.1003537)
82. Bouckaert RR, Drummond AJ. 2017 bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42. (doi:10.1186/s12862-017-0890-6)
83. Kingman JFC. 1982 The coalescent. *Stochastic Process. Appl.* **13**, 235–248. (doi:10.1016/0304-4149(82)90011-4)
84. Hunter J. 1835 *The works of John Hunter: with notes*. London, UK: Longman.
85. Hunter J. 1780 Account of a woman who had the small pox during pregnancy, and who seemed to have communicated the same disease to the foetus. *Phil. Trans. R. Soc.* **70**, 128–142. (doi:10.1098/rstl.1780.0008)
86. Maixner F *et al.* 2016 The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **351**, 162–165. (doi:10.1126/science.aad2545)
87. Warinner C *et al.* 2014 Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344. (doi:10.1038/ng.2906)
88. Martin MD, Vieira FG, Ho SYW, Wales N, Schubert M, Seguin-Orlando A, Ristaino JB, Gilbert MTP. 2016 Genomic characterization of a South American *Phytophthora* hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. *Mol. Biol. Evol.* **33**, 478–491. (doi:10.1093/molbev/msv241)
89. Yoshida K *et al.* 2013 The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**, e00731. (doi:10.7554/eLife.00731)
90. Rasmussen S *et al.* 2015 Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582. (doi:10.1016/j.cell.2015.10.009)
91. Moss B. 2001 Poxviridae: the viruses and their replication. In *Fields virology* (eds DM Knipe, PM Howley), pp. 2849–2883. Philadelphia, PA: Lippincott Williams & Wilkins.