Science
AAAS

Supplementary Materials for

Ancient chicken remains reveal the origins of virulence in Marek's disease virus

Authors:

Steven R Fiddaman[1†*], Evangelos A Dimopoulos[2,3†], Ophélie Lebrasseur[4,5], Louis du Plessis[6,7], Bram Vrancken[8,9], Sophy Charlton[2,10], Ashleigh F Haruda[2], Kristina Tabbada[2], Patrik G Flammer[1], Stefan Dascalu[1], Nemanja Marković[11], Hannah Li[12], Gabrielle Franklin[13], Robert Symmons[14], Henriette Baron[15], László Daróczi-Szabó[16], Dilyara N Shaymuratova[17], Igor V Askeyev[17], Olivier Putelat[18], Maria Sana[19], Hossein Davoudi[20], Homa Fathi[20], Amir Saed Mucheshi[21], Ali Akbar Vahdati[22], Liangren Zhang[23], Alison Foster[24], Naomi Sykes[25], Gabrielle Cass Baumberg[2], Jelena Bulatović[26], Arthur O Askeyev[17], Oleg V Askeyev[17], Marjan Mashkour[20,27], Oliver G Pybus[1,28], Venugopal Nair[1,29], Greger Larson[2‡], Adrian L Smith[1*‡], Laurent AF Frantz[30,31*‡]

Affiliations:

[1]Department of Biology, University of Oxford, Oxford, UK

[2]The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK

[3]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

[4]Centre d' Anthropobiologie et de Génomique de Toulouse, Toulouse, France

[5]Instituto Nacional de Antropología y Pensamiento Latinoamericano, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina

[6]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

[7]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[8]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium

[9]Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Brussels, Belgium

[10]BioArCh, Department of Archaeology, University of York, York, UK

[11]Institute of Archaeology, Belgrade, Serbia

[12]Institute of Immunity and Transplantation, University College London, London, UK

[13]Silkie Club of Great Britain, Charing, UK

[14]Fishbourne Roman Palace, Fishbourne, UK

[15]Leibniz-Zentrum für Archäologie, Mainz, Germany

[16]Medieval Department, Budapest History Museum, Budapest, Hungary

[17]Laboratory of Biomonitoring, The Institute of Problems in Ecology and Mineral Wealth, Tatarstan Academy of Sciences, Kazan, Russia

[18]Archéologie Alsace - PAIR, Bas-Rhin, France

[19]Departament de Prehistòria, Universitat Autònoma de Barcelona, Barcelona, Spain

[20]Bioarchaeology Laboratory, Central Laboratory, University of Tehran, Tehran, Iran

[21]Department of Art and Architecture, Payame Noor University (PNU), Tehran, Iran

[22]Provincial Office of the Iranian Center for Cultural Heritage, Handicrafts and Tourism Organisation, Bojnord, Iran

[23]Department of Archaeology, School of History, Nanjing University, China

[24]Headland Archaeology, Edinburgh, UK

[25]Department of Archaeology, University of Exeter, Exeter, UK

[26]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden

[27]CNRS, National Museum Natural History Paris, Paris, France

[28]Department of Pathobiology and Population Sciences, Royal Veterinary College, London, UK

[29]Viral Oncogenesis Group, Pirbright Institute, Woking, UK

[30]Department of Veterinary Sciences, Ludwig Maximilian University of Munich, Munich, Germany

[31]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK

[†]joint-first author

[‡] co-senior authors

[*]corresponding authors. Emails: steven.fiddaman@biology.ox.ac.uk; adrian.smith@biology.ox.ac.uk; laurent.frantz@lmu.de

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S11

Tables S4, S9 and S10

Captions for Data S1

References (26-74)

Other Supplementary Materials for this manuscript include the following:

Data S1, which comprises:

- Table S1: Sample metadata
- Table S2: Screening and capture sequencing results
- Table S3: Modern genome metadata
- Table S5: Integrity of miRNA sequences in ancient MDV
- Table S6: Fixed differences between ancient and modern MDV strains
- Table S7: PAML results
- Table S8: Meq sequence metadata
- Table S11: Metagenomic screening summary data
- Table S12: SNP summary table
- Table S13: Tip dates for BEAST analysis

Materials and Methods:

Archeological site descriptions

1. **Buda Castle**, Teleki Palace, Budapest, Hungary (13th-18th century; OL1385, OL1389; contact: László Daróczi-Szabó).

The former Teleki Palace is located in Buda Castle (Castle Hill, Budapest, Hungary), the medieval royal capital of Hungary, in close vicinity of the Royal Palace, in a general height of 158 meters above Baltic Sea level. Excavations of the remains of the palace were directed in 1999-2000 by Dorottya B. Nyékhelyi. Despite the relatively small area, a large number of archeological finds came to the surface, with a large majority dating between the 13th and 18th Century. Among other types of finds at least 100,000 animal remains were excavated (the analysis of the assemblage is ongoing). In general the assemblage shows a typical picture of animal use of the area: cattle dominating with around 45% of the finds, followed by sheep and/or goat and pig (20%), and others. The domination of domestic animals is almost total, although written sources prove that game was an important part of everyday diet: the most likely explanation is that large wild animals were deboned at the hunting site, and only the meat got to the markets. The relatively small number of bird and fish remains is because of the excavation methods: most of the finds were hand collected, and only a small part went through water sieving – here the number of these small finds drastically increased. Discrepancy from this generalities could be observed when finds connected to non-christian population were examined: pig is almost totally non-existent in layers supposedly connected to the local, 14th Century Jewish population of "Well 8", and objects from the Ottoman Period (mid 16th-late 17th Century) (e.g. "Well 7"). In the latter next to the lack of pig remains a drastic increase of domestic small ruminants was noticed, as well as a decrease in cattle remains, the former becoming dominant.

2. **Castillo de Montsoriu**, Spain (16th century; OL1984, OL1986, OL1999; contact: Maria Sana).

Montsoriu Castle is located in the north-east of the Iberian Peninsula, on a hilltop at 650 m.a.s.l. (4114605800N, 213203000E). Archaeological excavations on-going since 1993 have been able to document the successive changes and re-structuring it underwent from the tenth century onwards, in consonance with the social and political changes in that long span of time. During the 2007 season, an abandoned cistern was excavated. This corresponds to the last stable occupation phase at the castle. It yielded an extremely well-preserved assemblage (UE 10955) consisting of different categories of organic and inorganic remains from the castle's larder. This sample results from a specific action carried out in a very short time. The bone sample recovered from the castle's cistern totals 10,922 remains, being representative of the food stored in the castle's larder. The deposition of this assemblage in the cistern would have been linked with the final abandonment of the castle in the last third of the sixteenth century. The inhabitants of the castle based animal resources exploitation strategies on domestic species (97% of NISP), including sheep and goats (29%), pigs (39%), cattle (32%) and poultry. Hunting was of lesser economic importance, and mainly involved wild rabbits, hares, red deer, boar, roe deer and fox. A large number of birds and fish are also consumed. A total of 863 chicken bones were retrieved from UE10955,

representing a total of 74 specimens. Most of the chicken remains correspond to females (61%). A total of 65 chicken remains show cut marks resulting from food processing and cooking for consumption.

3. **Strasbourg Rue de Lucerne** - Rue du Jeu de Paume, France (16th- early 17th century; OL1936; contact: Olivier Putelat).

An archeological excavation was carried out « Rue de Lucerne – Rue du Jeu de Paume », in 2012-2013, in a suburb of the city of Strasbourg (France, Bas-Rhin). This excavation was carried out by the PAIR/Archéologie Alsace, on an area of 1119m², under the direction of M. Werlé. An agro-pastoral building was uncovered, as well as latrines. The site yielded 3,124 terrestrial animal bones, and about 1,300 fish remains. The filling of latrines 1067 (US 1059), from which the OL1936 sample originated, is dated to the end of the 16th Century. US 1059 yielded 499 remains of terrestrial animals (NISP: 397), including 45 chicken bones (11% of the NISP) and 32 remains of other birds, ducks and geese mainly(*26*).

4. **Rudnik**, Serbia (14th-15th century; OL1008; contacts: Nemanja Marković and Jelena Bulatović).

The medieval settlement of Rudnik comprises four archaeological sites: Stacionar, Imanje Nikić, Imanje Marković, and Kojovača. The settlement is situated on the Rudnik Mountain, near the present-day town of Rudnik in central Serbia (approx. 95 km south of Belgrade (44°08'06.0"N 20°29'42.0"E)). Based on the current archaeological and written records, this area became particularly important in the second half of the 13th century, when German Saxon miners inhabited the country on the invitation of the Serbian king Stefan Uroš I Nemanjić (1241/1242–1276). New technologies in the exploitation and processing of ores provided a significant increase in the economy of medieval Serbia. The established square with an urban settlement developed later into an important mining and trade center, reaching its peak during the 14th and the first part of the 15th century. The main mining activity was the silver extraction, while large amounts of lead and copper were also exploited. Most likely from the very beginning of the mining activities, the mint under the direct management of the Serbian ruler started working at medieval Rudnik. Historical records indicate that a colony of Dubrovnik traders was also established here. So far, archaeological excavations revealed three sacred buildings (two Orthodox and one Catholic churches), numerous profane buildings (some of which are monumental in scale), and numerous and varied movable archaeological finds. All these data indicate that during the last decades of the 13th to the middle of the 15th century, Rudnik was an important developing settlement composed of native and newcomer populations of different social and religious characters(*27*).

Animal remains analyzed so far come from two Rudnik sites of Stacionar and Imanje Nikić. They were collected from cultural layers dated to the period between the 14th and the first half of the 15th century based on the findings of coins and pottery fragments. The Rudnik faunal assemblage comprises the remains of mammals, birds, fish, and molluscs, of which 449 specimens were identified to a species or at least a genus level. Most of the animal remains belonged to the domestic animals.

Caprines (sheep and goats taken together) are the most abundant taxa, followed by domestic cattle and domestic pigs. Other mammal remains were of horse, dog, cat, roe deer, red deer, hare, and squirrel. A small number of fish remains (common carp and sturgeon) and Jacob's scallop shell fragment were also present. Domestic chicken was the only bird species identified and is represented with 20 (4% of the total assemblage) bone fragments(*28*). The specimen OL1008 (a coracoid bone) was found within the cultural layer inside of an economic/residential building at the site of Imanje Nikić.

5. **Fishbourne** (Roman BC50-AD280; OL1128; contact: Robert Symmons).

Discovered in 1960, Fishbourne Roman Palace is situated approximately 2.5km west of the city of Chichester, near the south coast of Britain. The palace itself is the largest domestic Roman building yet found north of the Alps, and today is a visitor center and museum displaying artifacts from the site and 16 *in situ* mosaics. The site seemingly was occupied from the late Iron Age. Following the Roman invasion of 43 CE, the area was used as a military supply base. This later evolved into a "proto palace", which included a large and luxurious bath house. Around 75 CE the proto palace was enlarged into the "Flavian palace": a massive high-status complex comprising 4 wings arranged in a square around a formal garden with an informal garden to the south, leading down to nearby Chichester Harbour. The footprint of the building is approximately 21,000 square meters, although little is known about any ancillary structures that would undoubtedly have serviced the main complex. From the second century the building experienced a gradual period of decline and contraction before it was destroyed by fire in 280 CE.

Sample OL1128 was recovered in 2002, during excavations approximately 30 m east of the northeast corner of the Flavian palace (site code FBE02, context 1040, catalog number CHCFB : FBE02/CB88). The context was interpreted as a midden and dated to the mid-late first century.

6. **Vienna Avar Cemetery**, Austria (7^{th} / 8^{th} century; OL1231; contact: Henriette Baron).

The Avar Cemetery at Vienna Csokorgasse was excavated under the direction of Ludwig Streinz and on behalf of the Historical Museum Vienna (today Wien Museum) in the years 1976 and 1976 within the scope of a building project. In the course of the excavation, 705 burials were unearthed and documented and the cemetery was excavated completely. According to Falko Daim and Ludwig Streinz, the cemetery was continuously used: the first burials in the northeast date to the Early Avar Period II (2^{nd} quarter 7^{th} century CE) and the last graves in the west and south of the burial area stem from the Late Avar Period II to III (2^{nd} and 3^{rd} third 8th century CE). The burial type – inhumation burials orientated west-east, displaying a variety of partially gender-specific burial goods – conforms to the usual customs of the Middle and Late Avar Periods. For the most recent discussion of the site and its zooarchaeological findings, see(*29*).

In 491 (70 %) of the 705 burials animal bones were found. In most cases these were remains of chickens (319 graves, 45 %) and of sheep or goats (313, 44 %). Bones of cattle (240 burials, 34 %) and of pigs (84, 12 %) also occurred in many burials. The chicken was interred in different degrees of

skeletal completeness. Of the mentioned domestic mammals, primarily the thigh portion containing the femur was selected as a burial good. The faunal material also comprises some birds (domestic or greylag geese, Western jackdaw, Northern Goshawk, Eurasian skylark, a pigeon, white-tailed eagle, smew, gray partridge, Eurasian woodcock, as well as some unidentified bird remains), and some fish (pike, a wels catfish, and different cyprinids). Four outstanding rich Late Avar burials in the south of the cemetery area (burials 650, 690, 692, and 693) contained harnessed horses in an age fit for riding usage. In addition, three of these comprised complete skeletons of fully grown large male dogs. From the fourth equestrian burial (692) only a single dog tibia was recovered. Furthermore, a partial skeleton of a young puppy was found in the digging shaft of burial 650. Another one was recovered from a Middle Avar period child burial (burial 462). The equestrian graves contained grown up men with belt fittings and in one case (burial 690) two children and an adolescent, the latter presumably also with belt fittings.

7. **Kazan City**, Russia (16th-17th century; OL1584; contact: Dilyara Shaymuratova).

Kazan City is located in the Republic of Tatarstan in the east of the European part of Russia (55°47'26" N 49°07'19" E), excavations were carried out on the territory of the modern backyard of the Kazan Federal University in 2002. Sample OL1584 was located in layers from the “Early Russian period” (16th-17th centuries). The study of the remains of bird bones was carried out by Igor Askeyev and Dilyara Shaymuratova.

8. **Cheboksary City**, Russia (16th–18th century; OL1585; contact: Dilyara Shaymuratova).

Cheboksary City is situated in the Republic of Chuvashia in the east of the European part of Russia (56°09′07″ N 47°14′48″ E), research and excavations of this archaeological site were carried out in 2004-2005 near the territory of the Vvedensky Cathedral. Sample OL1585 was located in layers of the 16th–18th centuries. Archaeozoological material was studied by Igor Askeyev and Dilyara Shaymuratova.

9. **Elabuga hillfort**, Russia (17th-18th century; OL1599; contact: Dilyara Shaymuratova).

Elabuga hillfort is located in the Republic of Tatarstan in the east of the European part of Russia (55°44′48″ N 52°01′57″ E), excavations were carried out on the territory of the location called "Chertovo gorodishche" in 2003. Sample OL1599 was located in layers from the “Russian period” (17th-18th centuries). Research of bone remains of birds was carried out by Igor Askeyev and Dilyara Shaymuratova.

10. **Andlau**, France (10th-12th century; OL1934; contact: Olivier Putelat).

The village of Andlau (France, Bas-Rhin) is located in Alsace, about forty kilometers southwest of Strasbourg. An archaeological excavation was carried out in 2008 by the PAIR/Archéologie Alsace, at the "12 Cour de l'Abbaye", on an area of 900 m², under the direction of A. Koziol. It concerned the gardens, located about thirty meters south of the Roman abbey church, in the center of the current village. This operation constitutes the first archaeological approach of the medieval abbey, which until then was known only from written sources. The latter indicates that this Benedictine monastery, reserved for women from the aristocracy, was founded in 879/880, by Richarde, the wife of the emperor *Charles le Gros*. The excavation yielded more than 7300 bone remains, 4304 of which are dated to phase A2 of the site (10th-12th century CE), from which the OL 1936 sample originated. Of these 4304, 2108 were determined, the 85 chicken bones accounting for 4% of the NISP. It should be noted that the examination of some of the sieve refusals yielded several thousand fragments of eggshells, mainly attributable to the hen, as well as the discovery of a dwarf hen(*30*).

11. **Studenica Monastery**, Serbia (14th-15th century; OL1214; contacts: Nemanja Marković and Jelena Bulatović).

The Studenica Monastery is located on a flat plateau in the valley of the Studenica River, 49 km southwest of Kraljevo in southwestern Serbia (approx. 220 km from Belgrade; 43°29'11.4"N 20°31'54.9"E). It represents one of the largest and richest medieval Orthodox monasteries in the country which has been on the UNESCO list of the world cultural heritage since 1986. It was founded at the end of the 12th century as an endowment and burial place of the progenitor of the Serbian medieval ruling dynasty of Nemanjić, the Grand Prince (*Veliki Župan*) Stefan Nemanja (1166–1196, †1199). Within its unique circular walls, the monastery complex contains: the Church of the Virgin Mary (the central sacred building), the King's Church, as well as the churches of St. John and St. Nicholas, several smaller sacred buildings, residential and economic facilities. Archaeological excavations (with minor and major interruptions) were conducted between 1949–2014. The study of everyday life in the medieval Studenica was part of the systematic archaeological research carried out in two phases between 1989–1998 and 2010–2014. Several waste areas were discovered inside the monastery complex and along the outer side of the southeastern rampart in the immediate vicinity of the eastern gate(*31*). These waste areas contained a large amount of animal remains whose analysis provided the first insights into the strategies of animal exploitation in one Orthodox medieval monastery(*32*).

Faunal remains originated from midden areas both outside and inside the monastery walls. Based on the stratigraphic data, coins and ceramic fragments findings, these middens with faunal remains were dated to the period from the beginning of the 14th till the middle of the 15th century. Out of the total number of identified specimens (NISP=1949), 1527 belonged to mammals, 282 to birds, and 140 to fish. The majority of mammal remains (92%) were from economically important domesticates: sheep, goat, pig, and in a smaller number cattle. Hare is the best represented game species, whereas only one red deer tine fragment was found(*32*). The diversity of fish species and written sources (f.e.,

Studenica Typikon) point to the fact that the exploitation of fresh fish had an important place in the medieval economy of the monastery. Besides the remains of the fish available more or less locally (such as Wels catfish, carp, and pike), remains of migratory sturgeons (such as beluga, Russian sturgeon, and stellate sturgeon) which were probably transported from the Danube area about 200 km far from the monastery, were found too(*33*).

Bird remains were found in three of the four analyzed midden areas. Out of 282 bird remains, 243 were determined to a species level. Remains of the five species were identified: chicken, duck, goose, pigeon, and eagle. Remains of chicken are the most frequent and comprise 227 specimens (i.e., 93.5% of the total bird remains identified to the species level). Chicken age and sex profiles shows that older hens were the most numerous in all midden areas implying that they were exploited primarily for the egg production(*34*). The sample OL1214 (a coracoid) was recovered in the midden area formed on the Buildings V and VII ruins inside the monastery walls and dated to the last decade of the 14th century and the first decades of the 15th century.

12. **Naderi Tepe**, Iran (19th century; OL2267, OL2268, OL2272; contact: Marjan Mashkour). Three samples:

a. Tepe Naderi 2018 - Tr1, S2, F5, A12. DNA ID: OL2267
b. Tepe Naderi 2016 - Tr1, S2, F5, 34. Depth 290cm. DNA ID: OL2268
c. Tepe Naderi 2016 - Tr1, S2, F5, A41. Depth 330cm. DNA ID: OL2272

Tepe Naderi is a 20m high artificial mound with a long occupational sequence that lasted from the late Chalcolithic/early Bronze Age through the early Iron Age (5128-158BP to 2752-27BP; Achaemenid to Sassanid period), a short medieval age settlement from 10th-11th Century CE, and eventually an occupational phase from Qajar period (18th-19th Century), when a fortress was built at the top of the mound overlooking a lower town which was a fortified enclosure.

F5 is a bag-shaped pit dug into the medieval layers at the foot of the Tepe in the lower town. The pit is 1.16m at the opening and 2.96m at the bottom (2.48m deep) in the south of Trench 1, from which dozens of fragments of blue-white stonepaste wares and many bone fragments were discovered. The pit was found and partly emptied during the first campaign in 2016 and partly in 2018. Therefore, it is plausible that samples from 2016 and 2018 are from the same individual. All samples are from the same context.

At first, since the pit cuts through the medieval layers in the lower town at the foot of the tepe, it was attributed to the medieval period. However, this was later revised based on a finer study on the material and some radiometric dating. A calibrated radiocarbon date of a piece of animal bone places the pit within the range of 146-145 BP, which falls in the Qajar period. Two thermoluminescence dates of two pottery samples are 180-250 and 150-210 BP, or between 1700 and 1800 CE, which are consistent with the radiocarbon-dated animal bone.

Further radiocarbon dating conducted in the present study corroborates the previous dating. The calibrated age range and associated probabilities for sample OL2272 is: 1698-1723 (23.6%); 1814-1835 (20.8%); 1885-1910 (23.8%). See section below for full description of radiocarbon dates obtained in the present study.

13. **Bard-e Mar** (Bem), Kurdistan (18th-early 21st century; OL2178; contact: Amir Saed Mucheshi and Marjan Mashkour).

Bard-e Mar is situated in the Zargos region (35°9'39.63" N, 46°22'14.35" E, 790 m above sea level), on an old terrace of the Sirwan River in Sarvabad County, (a part of the Hawraman region), in the Kurdistan Province, Western Iran. Bard-e Mar is a small site of approximately two hectares with a steep slope towards the river. Bard-e Mar was recorded during the Darian Dam Archaeological Salvage Project (DDASP) directed by Fereidoun Biglari in 2014 and excavated by Amir Saed Mucheshi in 2015. Excavations at Bard-e Mar were conducted in two trenches, named Trench I and Trench II, both reaching the virgin soil. The cultural finds included potsherds, ground stones, clay pipes, rich faunal remains and rectangular and circular stone architectural structures. The Hawraman region has had a traditional and special masonry architectural style due to steep slopes of the mountains, and its continuity to the modern day is observable in evidence from Bard-e Mar. It is worthy to note that the diagnostic glazed pottery vessels were not discovered at site. As a result of relative and absolute chronological comparisons, two distinct periods were identified in Bard-e Mar, including the Middle Islamic (13th to 14th century A.D.) and the Late Islamic, the reign of Qajar Dynasty (18th to early-20th century A.D.). The finds from the Middle Islamic layers are less than its upper part, due to the reduction of the dimension of the trenches. The data from the Middle Islamic period are comparable to the western regions of Iran. The Qajar Dynasty/Late Islamic finds, especially clay pipes, are comparable with the Late Ottoman period in the Iraqi Kurdistan and Kurdish potteries reported from the Zagros region in the Late Islamic period(*35*).

Faunal remains were studied by H. Davoudi, R. Khazaeli and M. Mashkour in the National Museum of Iran and Bioarchaeology Laboratory, Central Laboratory of the University of Tehran. A total of 2807 pieces of animal bones have been discovered from excavations in Bard-e Mar, of which 1718 samples belong to Trench I and 1089 pieces belong to Trench II. The bones are generally placed in the group mammals and birds, and only one piece of fish bone was identified in the collection. Cattle, sheep and goats are dominant, comprising 50% of the collection. A small amount of the collection derives from wild sheep and goats, deer, equines, foxes, forest otters and rabbits. A significant part of animal remains from the Late Islamic period belong to birds, of which some are domestic chickens and their relatives. Examining the animal remains of Bard-e Mar shows that its subsistence economy is based on animal husbandry and the people living in this area used goats, sheep and cattle as the main source of livelihood, a method that is still common in this region(*36*).

In this paper, Sample OL2178 was analyzed, which belonged to the right femur of *Gallus gallus*, derived from Trench I, Locus 14, Code a#56, the Qajar Dynasty/Late Islamic period (18th to early 20th century A.D.).

Description of bird used as positive control

To validate the shotgun sequencing and baiting procedure, we also processed DNA from a modern bird displaying symptoms of Marek's Disease from a naturally acquired infection. A feather from the bird, a Silkie chicken, was provided by Gabrielle Franklin of the Silkie Club of Great Britain. At the time of infection, the bird was unvaccinated and 3-4 years of age. Symptoms included a sudden loss of weight and dropped wing/paralysis on the left side. Following the development of these more severe symptoms, the bird was euthanized to prevent suffering.

Radiocarbon dating

In total, four samples were radiocarbon dated (Table S10) using either the Oxford Research Laboratory for Archaeology and the History of Art or Beta Analytic. Uncalibrated dates were calibrated using IntCal20: Northern Hemisphere(*37*). In all cases, there was a degree of ambiguity in the measured sample ages, so raw plots are given below (Fig. S8).

Ancient DNA extraction

DNA extractions were performed in dedicated ancient DNA (aDNA) facilities at the PalaeoBARN (University of Oxford). DNA was extracted from chicken bone samples (Table S1) in a dedicated ancient DNA laboratory using the appropriate sterile techniques and equipment. Prior to grinding, ~0.5mm of the exterior surface of the bone was removed using a Dremel 3000 electric hand-drill. Extraction was carried out using the Dabney extraction protocol(*38*).

Library preparation and preliminary sequencing

Illumina libraries were built following either the protocol in(*39*) or(*40*), but with the addition of a six base-pair barcode added to the IS1_adapter.P5 and IS3_adapter. P5+P7 adapter pair. The libraries were then amplified on an Applied Biosystems StepOnePlus Real-Time PCR system to check that library building was successful, and to determine the minimum number of cycles to use during the indexing amplification PCR reaction. A six base-pair barcode was used during the indexing amplification reaction resulting in each library being double-barcoded with an "internal adapter" directly adjacent to the ancient DNA strand and which would be the first bases sequenced, and a traditional external barcode that would be sequenced during Illumina barcode sequencing. An additional sequencing library was built for sample OL1385, which had two external 6 bp indices appended to the P5 and P7 Illumina sequencing primers, and had no internal index tag. Up to 200 libraries with unique barcode combinations were pooled at equimolar levels (as determined by an Agilent Technologies 2200 TapeStation) and an 80 - 150 bp run was carried out on an Illumina HiSeq 2500/4000/X sequencer at Novogene or Macrogen.

Preliminary analyses of sequencing data using HAYSTAC

We screened 995 ancient chicken (*Gallus gallus domesticus*) DNA libraries, with HAYSTAC(*6*), with default settings, for the Gallid Alphaherpesvirus 2 (MDV) genome, using a custom database of complete herpesvirus genomes obtained from in the NCBI RefSeq database (Table S11). This analysis identified 18 samples with at least one MDV read (range: 1–785; Table S2). To ensure the validity of the positive identifications in samples with a higher number of confidently assigned MDV reads (>50), we ensured that the respective evenness of coverage ratio (genome coverage / fraction of genome covered) was less than 10, and that appropriate chemical damage patterns characteristic of ancient DNA were identified ((*41*); Fig. S2). These 18 samples, as well as a modern sample with MD symptoms (OL1099; positive control) and a sample for which we did not identify any MDV reads (OL1214; negative control) were subsequently genome captured (see below).

Design of probes for in-solution capture

Custom DNA probes for a tiled baiting approach were designed and synthesized by Arbor BioScience, based on the *Gallid alphaherpesvirus 2* genome (strain RB-1B; accession EF523390.1, NCBI). Baits were only constructed for one copy of each of the two terminal repeats (coordinates 1-14004 and 165217-178246 in EF523390.1 were excluded), and regions of low complexity in the genome were masked . Oligonucleotide baits were designed approximately every 28 nucleotides, yielding 8403 50-mer probe sequences for the whole MDV target genome. The probes allowed for the hybridisation of both modern and ancient strains of MDV, while limiting carryover of endogenous DNA.

In-solution capture and sequencing of captured libraries

Additional libraries were either amplified or re-built from extracts of the samples that were selected for targeted genome enrichment (Table S2), following the metagenomic screening for MDV DNA reads. These selected libraries were amplified again following the same indexing and PCR amplification strategies as previously described, and were condensed in a final volume of 25 µl. Libraries from samples OL1008, OL1128, OL1214, OL1231, OL1385, OL1585, OL1987, OL1999 were built once and were also amplified and captured once, whereas for samples OL1389, OL1584, OL1599, OL1934, OL1936, OL1984, OL1986, OL2178, OL2267, OL2268, OL2272 three different libraries were built from the same extract and each library was subsequently amplified and captured (Table S2). Any additional libraries built for the samples that were captured more than once, were indexed and amplified for sequencing, using conventional full-length P7 and P5 Illumina primers that were both indexed with external 6 bp indices, instead of having one internal and one external index(*40*).

MDV genome capture and paired-end sequencing on Illumina NovaSeq lanes were performed by Arbor BioScience. Specifically ~200 million 150 bp paired-end reads were generated per capture sequencing pool. As a positive control, one modern chicken feather sample (OL1099) which displayed symptoms characteristic of MD was also processed using the above procedure and yielded a

4× genome. Sample OL1214 yielded no MDV-specific reads following screening and was included as a negative control in the genome capture protocol.

MDV-positive samples

In total, we performed genome captures for 20 samples, comprising a modern positive control (OL1099), an ancient negative control (OL1214) and 18 ancient samples with varying degrees of MDV positivity (Table S2). Of these, 16 samples (including the positive control) were unambiguously positive for MDV DNA. The negative control was enriched to 451 post-capture reads, but a majority of these were short (<25bp) and therefore likely nonspecific (Table S2). Three further samples (OL1008, OL1231 and OL1128) yielded very few (<60) post-capture reads, and these were overwhelmingly short (<25bp), so were considered negative for MDV (Table S2).

Capture efficiency

For MDV-positive samples, enrichment for MDV sequence ranged from 1.65 to 273.12-fold increase in unique DNA templates, allowing us to increase the depth of coverage of the ancient MDV isolates (range: 0.13 - 41.92), reduce the percentage of missing sequence data and reconstruct consensus genomes (Table S2). The genome enrichment captured virtually all the unique DNA templates that were available in the analyzed libraries, as shown by the high percentage of duplicated reads post capture (range: 56.6 - 99.98%; Fig. S1, Table S2).

Short read alignment of ancient data

The captured data were aligned with the bwa aln algorithm against the RB-1B (EF523390.1) MDV reference genome with default parameters apart from disabling the seed option ("-l 1024") (*42*). Because the MDV genome has terminal long (TRL) and short (TRS) repeat regions (14kbp and 13kbp, respectively), one repeat from each of these duplicated regions was masked (coordinates 1-14004 and 165217-178246 in EF523390.1) in order to avoid low mapping quality stemming from a DNA read aligning to multiple positions, and to retain the maximum number of reads.

Testing read sharing between chicken and MDV

To detect sequence homology between chicken and MDV genomes that could lead to false positive identification of ancient MDV-positive samples, we simulated short read sequencing data based on the RB-1B (EF523390.1) genome and mapped these regions against the chicken genome (GRCg7b). To simulate reads, ART was used (art_illumina -ss HS10 -i [input_fasta] -l 100 -f 50 -p -m 350 -s 50 -o [output_fastq]). Reads were then mapped against the chicken genome using the bwa aln algorithm.

Genotyping of ancient data

Aligned reads had 5 base pairs trimmed from both their 5' and 3' ends with bamUtil v 1.0.14(*43*), and read group identifiers were added to the resulting bam files by using picard tools v 2.16.0(*44*).

The GATK package(*45*) was used for variant calling and sample genotyping. HaplotypeCaller v 4.1.2.0 was employed for SNP (single nucleotide polymorphism) calling with the following parameters: the EF523390.1 MDV genome (masked for TRL and TRS; see above) was used as reference, a value of 30 for the minimum base quality (Q) was required, all sites were outputted in GVCF files, ploidy was set to 1, and physical phasing was not allowed (gatk HaplotypeCaller --reference EF523390.1_mask.fasta --min-base-quality-score 30 --output-mode EMIT_ALL_SITES --sample-ploidy 1 -ERC GVCF --do-not-run-physical-phasing true).

The GVCF files, from each individual sample, were aggregated into one with the CombineGVCFs v 4.1.2.0 tool. The aggregated data were input into GenotypeGVCFs for joint genotyping of the different samples. The raw variants were further filtered using VariantFiltration, filtering in variants with a minimum base quality (Q) value of 30 and a minimum sequence depth of 5 reads (gatk VariantFiltration --filter-expression "QUAL >= 30.0 && DP >= 5"). VariantsToTable was subsequently employed to convert the resulting VCF files containing the filtered SNPs to tables (Table S12).

Generating fasta files

Consensus fasta files were also obtained from aligned reads in bam format using the htsbox pileup tool (https://github.com/lh3/htsbox). The following parameters were used: a minimum read length of 25 base pairs, 5 base pairs were trimmed from each end of an aligned read, a value of 30 for the minimum base quality (Q) and for the minimum read alignment quality (q) was set, for polymorphic sites the majority frequency allele was called, and a minimum sequence depth of 5 reads was also required (htsbox pileup -l 25 -T 5 -q 30 -Q 30 -M -s 5).

Publicly available (modern) data

Raw Illumina sequencing reads for samples KU173115.1, KU173116.1, KU173117.1, KU173118.1, KU173119.1 were downloaded from the SRA using fasterq-dump v 2.11.0 (https://github.com/ncbi/sra-tools), and raw reads for samples MF431493.1, MF431494.1, MF431495.1, MF431496.1 were provided to us by Jakob Trimpert(*7*). Modern samples AF243438.1, AY510475.1, DQ530348.1, EF523390.1, EU499381.1, FJ436096.1, FJ436097.1, JF742597.1, JQ314003.1, JQ806361.1, JX844666.1, KU744555.1, KU744556.1, KU744557.1, KU744558.1, KU744559.1, KU744560.1, KU744561.1, KX290013.1, KX290014.1, MG432697.1, did not have any associated raw sequencing reads deposited on the SRA, so full genome sequences were downloaded from NCBI's Nucleotide database using a custom biopython (v 1.79) script (*25*). Metadata associated with modern MDV strains can be found in Table S3.

For publically available sequenced MDV strains that have undergone serial passage under experimental conditions, only the lowest passage number was retained for time-calibrated phylogenetic analyses. Therefore the following accessions were excluded from our BEAST analysis

(but were included in maximum likelihood phylogenies): JQ806362.1 (passage 31), JQ809691.1 (passage 41), JQ809692.1 (passage 61), JQ820250.1 (passage 81), JQ836662.1 (passage 101), KT833852.1 (passage 70), KX290015.1 (passage 75), KX290016.1 (passage 110). We also excluded strains that were duplicated in the dataset (KT833851.1 and NC_002229.3, both Md5). We further excluded accession AF147806.2, since it was isolated in 1964 and subsequently experimentally passaged in cell culture at least 45 times before being sequenced(*4*). On a maximum-likelihood phylogeny AF147806.2 falls on a long terminal branch and a root-to-tip divergence analysis further showed that it is enriched for substitutions compared to contemporary strains (see Fig. S9A and C and "Temporal signal" section below). Repeated passaging likely introduced mutations in this strain that obfuscates the temporal signal and makes it unsuitable for time-calibrated phylogenetic analyses. Accession MG518371.1 was also excluded from our BEAST dataset as it lay in a clade with a vaccine strain and thus likely represented contamination.

Removing BAC sequences

MDV reference genomes with accession numbers KT833851.1, KT833852.1, and FJ436097.1 were contaminated with sequences from BAC (bacterial artificial chromosome) cloning vectors. BAC sequences were removed from these genomes with a custom biopython script, and their non-contaminated respective genomes were outputted in fasta format.

Masking of the modern genomes

Since the consensus MDV sequences in the ancient viral dataset had one of each of the terminal repeats (TRL and TRS) masked, we also masked the same terminal repeats in our modern viral sequence dataset.

The modern MDV genomes, which were downloaded from the NCBI nucleotide database, (cleaned from BAC contamination) were split into 500 bp fragments, with a 5 bp overlap between each k-mer, using the pyfasta split v 0.5.2 tool (https://github.com/brentp/pyfasta). Each fragmented genome was subsequently aligned against the masked EF523390.1 MDV genome using the bowtie2 algorithm v 2.4.4(*46*) with default settings. The resulting bam files for each modern MDV genome were sorted and passed to htsbox pileup with default settings (htsbox pileup -l 25 -M), in order to build new consensus fasta sequences that had the TRL and TRS repeats masked, as for the ancient MDV sequence dataset. Two sets of consensus fasta sequences were built for the ancient samples, one set with ≥1x and and one with ≥5x coverage depth per site (htsbox pileup -l 25 -M -s 1/5).

For the modern samples for which sequencing reads were available, Bowtie2 v 2.4.4(*46*) with default settings was used to align these short reads against the masked EF523390.1 MDV reference genome, and consensus fasta sequences were built using htsbox (htsbox pileup -l 25 -T 3 -q 30 -Q 30 -M -s 5).

Sequencing depth for sample OL1385

To calculate per-base sequencing depth for the highest coverage sample (OL1385), samtools depth (v. 1.10) was used. To plot the sequencing depth (Fig. S10), a custom R script was used to plot the mean depth across non-overlapping bins of 300bp (*25*).

Phylogenetic clustering analysis

The R package fastbaps v 1.0.4(*47*) was used to identify phylogenetic clusters within our MDV dataset, using the baps algorithm to perform Bayesian hierarchical clustering(*48*), unconstrained by any phylogenetic tree. The 35 ancient and modern genomes used in the BEAST analysis were used as input for the lineage clustering analysis. The resulting lineages were compared and plotted side-by-side with the Maximum Credibility Tree, produced by TreeAnnotator v 1.10.4 from the BEAST software suite(*49*), using the R library ggtree v 3.2.0(*50*); Fig. S11).

Positive selection analysis

The same alignments used for the BEAST analysis were also analyzed for positive selection using codeml implemented in PAML (v4.9;(*13*)). Open reading frames were sliced out of the genomic alignment with reference to coordinates of the annotated RB-1B genome (accession EF523390.1). Exons from multi-exonic genes were concatenated and ORFs on the reverse strand were reverse complemented. For ORFs in the repeated regions, only the internal repeat was included in the analysis, meaning ORFs MDV000.5–MDV006.6 and MDV097.3–MDV103 were excluded. The branch-site test in PAML was used (model=2; NSsites=2; fix_omega=0/1; omega = 1), specifying the modern MDV sequences as the 'foreground' lineage and the ancient MDV sequences as the 'background' lineage. Pairs of log-likelihood values for each locus were used to perform a likelihood ratio test (2ΔlnL) with d.f. = 1, then the final p-value calculated by dividing the original p-value by 2 (as described in the PAML manual). P-values were transformed using the Benjamini-Hochberg procedure in R (v4.2.1;(*14*)). Of the 154 non-redundant loci included in the analysis, 49 genes showed significant evidence (corrected $p<0.05$) for positive selection in modern MDV with respect to ancient MDV. Full PAML results are presented in Table S7. Positive selection results were plotted on a circular representation of the RB1B genome in Fig. 2 using Circos(*51*).

Maximum Likelihood trees and Multisequence Alignments

Maximum likelihood (ML) phylogenetic trees (Fig. S4) were built using RAxML v 8.2.9(*52*), and Neighbour Joining (NJ) trees (Fig. S3) were constructed using seaview v4. The percentage of unknown bases (Ns) was calculated for each genome with the seqtk comp tool v 1.2.95, and all samples with at least 1% of the genome covered (1782 sites) were kept, in an effort to retain and analyze the maximum number of ancient MDV genomes. In total 42 modern MDV genomes from previous studies (Table S3) and 11 captured MDV genomes from the present study were included. We used consensus fasta sequences built with htsbox for this analysis (see above - for the captured

samples one set was built with coverage depth of 1x and one set with 5x). Of the 11 captured genomes that were included, 1 genome was from a modern sample (OL1099; sampled 2014) while the remaining 10 were obtained from archeological samples from the 14th-20th century (Table S1).

RAxML was used to build an unrooted tree with the following parameters: a random seed value for rapid bootstrapping, a parsimony-based starting tree, a GTRGAMMA substitution model, and 100 bootstrap runs (raxmlHPC-PTHREADS -f a -T 10 -x $RANDOM -k -# 100 -p $RANDOM -m GTRGAMMA). Seaview was used to build an NJ tree using the Jukes-Cantor model, with 100 bootstrap runs.

Maximum likelihood tree based on transversions

A midpoint-rooted maximum likelihood tree based only on transversions (Fig. S5), was built using 4 ancient MDV genomes (each with at least 80% coverage at 5x), 1 modern positive control from the current study, and 30 modern MDV genomes from public sources. All transition polymorphisms were set to "N" (unknown base) using a custom biopython script (*25*). All monomorphic positions were also removed using the snp-sites tool v 2.5.1(*53*). A phylogenetic tree was then constructed using the Lewis correction method for ascertainment bias(*54*) along with the ASC_GTRGAMMA substitution model (raxmlHPC-PTHREADS -f a -T 10 -x $RANDOM -k -# 100 -p $RANDOM -m ASC_GTRGAMMA --asc-corr lewis)

Maximum likelihood tree rooted with Herpesvirus of Turkeys

To confirm the placement of the root in the tree based only on transversions we used genomic data from a *Meleagrid herpesvirus 1* (HVT; accession: NC_002641.1) as an outgroup to build a rooted maximum-likelihood tree (Fig. S6a). Mafft v 7.123(*55*) was used to align the sequences of the outgroup to the MDV genomes (mafft --maxiterate 1000 --thread 5 --nwildcard). For the phylogenetic analysis we filtered out positions that were missing in the HVT genome (either "N" or gaps) using a custom biopython(*56*) script. We removed all the positions that were either monomorphic or unknown throughout all the samples in the alignment by using the snp-sites tool v 2.5.1(*53*). A phylogenetic tree was constructed using the Lewis correction method for ascertainment bias(*54*) along with the GTRGAMMA substitution model (raxmlHPC-PTHREADS -f a -T 10 -x $RANDOM -k -# 100 -p $RANDOM -m GTRGAMMA), .

Filtering overlapping open reading frames for BEAST analyses

To obtain divergence time with BEAST v 1.10.4 we removed overlapping open reading frames (ORF) based on a bed file containing ORF coordinates for EF523390.1 using a custom biopython script (*25*). Non-coding regions of EF523390.1 were also included, which resulted in a bed file that contained non-overlapping ORF and non-coding regions. These regions were extracted from the htsbox consensus fasta files for each of the ancient and modern genomes using the seqkit subseq tool

of the seqkit package v 0.12.1(*57*). ORFs that were found on the minus (-) strand were reverse complemented with the seqtk seq tool v. 1.2.95(*58*).

Filtering for coverage for BEAST analyses

In order to maximize the information content of the sequence alignments provided to BEAST and reduce the noise introduced by missing data in our ancient samples, we used a custom biopython script to calculate the percentage of missing data within each genomic region included in the BEAST analysis (see above). Only ORFs and intergenic regions with at least 10% of sites covered in all individual sequences in the alignment (i.e. 100% coverage: no missing data) were retained. We then concatenated the ORF and intergenic regions separately using the seqkit concat tool v. 0.12.1.

Temporal signal

To assess whether our data possess a temporal signal we first examined the correlation between root-to-tip divergence (in substitutions/site) and sampling date, using maximum likelihood trees constructed in RAxML v. 8.2.9(*52*). Sampling dates of ancient sequences were fixed to the means of the C14 calibrated distributions and the maximum-likelihood trees were rooted based on the outgroup analysis (Fig. S6a). A total of 36 MDV genomes were included in this analysis: 30 modern, 1 modern extensively passaged strain (AF147806.2), 1 modern positive control from the present study (OL1099), and 4 ancient captured genomes from the present study (OL1385, OL1389, OL1986, OL2272). Four variations were used: (i) all strains (Fig. S9A), (ii) all strains except AF147806.2 (Fig. S9B), (iii) all modern strains (Fig. S9C) and (iv) all modern strains except AF147806.2 (Fig. S9D). This analysis confirmed that AF147806.2 (cell passaged strain) did not undergo clock-like evolution and can be considered as an outlier.

We further used a Bayesian date randomization test(*59–61*) (DRT) to examine the strength of the temporal signal, by permuting sampling dates among genomes and performing 100 replicate analyses. For the analyses the same dataset (n = 35, excluding AF147806.2) and BEAST model as below were used, using an uncorrelated lognormally distributed (UCLD) relaxed clock and fixing the sampling dates of the ancient sequences to the means of the C14 calibrated distributions. Chains were run for 50 million steps and the parameter sampling frequency set at every 5,000 steps. This resulted in ESS>200 for all parameters in all replicate analyses. The HPD intervals of the mean clock rates compared to an analysis with unpermuted sampling dates is shown in Fig. S12.

Regressing the root-to-tip divergence against sampling date showed a strong positive correlation (Fig. S9) and the Bayesian DRT further showed strong evidence for a temporal signal in our dataset (Fig. S12).

Finally, we performed a Bayesian Evaluation of Temporal Signal (BETS (*62*)) to assess the effect of age uncertainty of the ancient MDV samples on the temporal signal. Here we compared the BEAST model described below (using the probability densities of the radiocarbon dates of ancient samples as

priors for their ages) to a model using only the mean radiocarbon ages and a model without sample ages. We used generalized stepping stone sampling (GSS (*63*)), implemented in BEAST, to obtain log marginal likelihood estimates (MLEs) of the different models. The results show that a model including sample ages and date uncertainty (log MLE = -123231.16) is preferred over a model using only mean radiocarbon ages (log MLE = 136774.82) and a model without sample ages (log MLE = -136777.73).

BEAST analysis

The BEAST v 1.10.4(*49*) package was used for the divergence dating and molecular clock rate estimation of the MDV phylogeny. A total of 35 MDV genomes were included in this analysis: 30 modern, 1 modern positive control from the present study (OL1099), and 4 ancient captured genomes from the present study (OL1385, OL1389, OL1986, OL2272). Publically available sequenced MDV strains that have undergone serial passage under experimental conditions or contaminated with vaccine strains were excluded from this analysis (see "Publicly available (modern) data" section for accessions). Tip dates were provided to BEAST for both the modern and archaeological samples (Table S13). We used the probability density function of the radiocarbon date as prior for the tip dates of the 4 ancient samples. To do so we used the empirical calibrated radiocarbon sampler (ECRS;(*64*) as implemented in BEAST. We used the same prior for OL1385 (which was directly dated) and OL1389 (not directly dated) as the bones from which these sequences were derived were from the same archaeological context.

The concatenated ORF alignment was partitioned further into codon positions with a custom python script (*25*). Each partitioned alignment along with the concatenated intergenic region alignment, was used as input for BEAST. The tree topology was fixed to the topology of the outgroup-rooted maximum likelihood tree (without the outgroup) constructed only from transversion polymorphisms. An independent GTR+$\Gamma_{4(65, 66)}$ substitution model was used for each alignment position and a constant size coalescent model was specified as a tree prior.

We used an uncorrelated lognormally distributed (UCLD) relaxed clock model(*67*). We ran 8 MCMC chains of 300 million steps each, with the parameter and tree sampling frequencies set at every 10,000 steps. The BEAST parameter log files were inspected with Tracer to ensure that all chains converged to the same posterior distribution. After discarding 10% as burn-in we combined the chains which resulted in ESS>200 for all parameters. The 95% CI of the coefficient of variation estimated under the UCLD model excluded 0 (median=0.524, 95% HPD 0.34-0.76) indicating that the MDV sequences in this tree did not evolve under a strict clock. To compare the placement of the root and the topology of the ancient samples, we also repeated the analysis, but without imposing any tree topology constraints.

Pairwise divergence

Average pairwise divergence between ancient (OL1385) and modern sequences was computed using the R (v. 4.2.1) Ape package(*68*) and a window size of 100 bp and step size 25 bp. This analysis used the same genomic alignment used for other analyses in the study. Pairwise divergence was plotted on Fig. 2 using Circos as described elsewhere.

Functional validation

In order to functionally test the transactivation ability of the Meq oncogene in ancient vs. modern MDV strains, an *in vitro* study system was constructed. The full length coding sequence of Meq from the highest coverage ancient sample (OL1385; Buda Castle, Hungary), along with Meq from modern strains (RB1B and Md5), were codon optimized and purchased from Integrated DNA Technologies (IDT; IA, USA) with a 3× FLAG sequence appended to the N-terminus (Table S9). Since Meq preferentially forms a heterodimer with chicken c-Jun, the chicken c-Jun coding sequence was also synthesized with the addition of a 3× FLAG sequence on the N-terminus (UniProtKB accession: P18870.2). All Meq and c-Jun sequences were cloned into the mammalian expression vector pTarget (accession: AY540613) using the NotI and XmaI sites. A reporter construct comprising a 773 bp section of the Meq promoter (which contains the putative AP-1 binding site AGTCATGCATGACGT bound by Meq itself) upstream of the firefly luciferase gene on the pGL3-Basic vector backbone was a kind gift from Venugopal Nair (Pirbright Institute, UK). A further reporter construct to normalize transfection efficiency – pRL-TK – comprises the *Renilla* firefly luciferase downstream of the relatively weak constitutive HSV-thymidine kinase promoter and was purchased from Promega (UK). All vectors underwent whole-plasmid sequencing (Plasmidsaurus) prior to their use experimentally to validate sequence integrity.

The chicken embryonic fibroblast cell line DF-1 was maintained in high glucose Dulbecco's Modified Eagle Medium with GlutaMAX (ThermoFisher, UK) and 10% fetal bovine serum (ThermoFisher, UK) in a humidified incubator at 37 °C and 5 % CO_2. The day before transfection, cells were passaged and seeded into 96-well tissue culture plates at a density of 2×10^4 cells/well. Wells were transfected with a Meq construct (240 ng), c-Jun (240 ng), Meq-pGL3 Basic reporter (200 ng) and pRL-TK reporter (4 ng) using 1.92 µl TransIT-2020 transfection reagent (Mirus; WI, USA). Twenty-four hours post transfection, cell supernatants were discarded and cells were washed in cold phosphate-buffered saline. Additional cell lysates were also prepared for immunoblotting of FLAG-tagged protein. Luciferase measurements were conducted using the Dual-Luciferase Reporter Assay System (Promega). Briefly, cells were incubated with 20 µl passive lysis buffer at room temperature for 15 minutes. In opaque white 96-well plates, the Dual-Luciferase assay was completed in a GloMAX plate reader (Promega), programmed to inject 100 µl Luciferase Assay Reagent II, then quench and read *Renilla* luciferase with 100 µl Stop & Glo Reagent, both with a 2-second delay and 10-second integration time.

Over several experiments, it became apparent that the *Renilla* reporter did not serve as an adequate background control because *Renilla* luciferase varied considerably depending on whether Meq was

present. *Renilla* luciferase measurements were substantially higher in wells containing Meq, suggesting that the TK promoter was being driven by the transactivation ability of Meq. The same phenomenon has been reported with pRL-SV40 constructs previously(*69*, *70*). As a consequence, firefly luciferase values normalized to *Renilla* luciferase were artificially diminished in Meq-containing wells. Instead of normalizing to *Renilla* luciferase, we present the raw firefly luciferase measurements. We confirmed via immunoblotting that the expression of ancient Meq was comparable to that of other Meq constructs.

Supplementary Text

Sequence analyses of Meq gene: extended description

Given that Meq is known to be a major determinant of MDV virulence, we elected to consider this gene in more detail. Meq exerts transcriptional control on downstream targets via its C-terminal transactivation domain. This domain is characterized by PPPP (tetraproline) repeats, and the number of tetraproline repeats is inversely proportional to the virulence of the MDV strain(*22*). Standard Meq is 339 aa in length, but length variants of Meq exist: long(l)-Meq is 398 aa due to a tandem duplication in the transactivation domain; short(s)-Meq is 298 aa; very short(vs)-Meq is 247 aa(*71*). We conducted an analysis of 413 standard-length Meq sequences (comprising four ancient sequences and one modern sequence derived from the present study, and 408 modern sequences derived from public databases), and found that all modern Meq sequences have between two and five tetraproline repeats. However, all ancient sequences have an extra tetraproline repeat (totaling six) that has been disrupted in the modern lineage (Table S8).

We reconstructed the phylogeny of all Meq sequences using the RAxML(*52*) tree builder in Geneious (v. 2019.2.3) with 100 bootstrap replicates. For each of the sequences, we also determined which of the tetraproline repeats had been disrupted and the causative mutation for any disruptions. This information was plotted onto the Meq phylogeny using iTOL (v. 6.5.8;(*72*)) along with the internal branches where each disruption is likely to have taken place (Fig. S7).

Close inspection of the tetraproline motifs revealed that each motif has been lost at multiple points throughout MDV phylogeny, confirmed by the presence of independent disruptive mutations. Moreover, there is evidence that the loss of tetraproline motifs is ordered. Following the loss of the 6th tetraproline motif (which occurred in the common ancestor to all modern strains), the 4th tetraproline motif is lost independently in two major lineages of European and Asian/N. American strains. Next, the 2nd tetraproline motif is most commonly lost, followed by either the 5th or the 1st tetraproline motif, usually in terminal branches of the tree. The selective pressure for the ordered loss of the 4th then 2nd tetraproline appears to be particularly strong – in the major Asian/N. American lineage, the 2nd tetraproline was lost 6 independent times. Intense selection pressure within the 2nd and 4th tetraproline motifs was also confirmed using a small number of Meq sequences derived from whole genomes in the positive selection analysis described above (where codons 176 and 217 were under

742 selection in the 2nd and 4th tetraproline, respectively). Occasionally, the 3rd tetraproline is lost after the
743 6th tetraproline, but this typically occurs at terminal branches (Fig. S7).

744 The loss of tetraproline motifs appears to act as a ratchet, whereby each subsequent loss results in an
745 increase in virulence, and once lost, motifs cannot be regained. This results in a stepwise scaling of
746 the fitness landscape wherein the order of losses is important. Although there are some observations
747 of virus lineages exhibiting an alternative loss order, such lineages are not widespread, suggesting that
748 they may become stuck in local fitness peaks and are outcompeted by lineages following the order
749 described above.

750 Functional validation results: extended description
751 Having identified the crucial oncogene Meq as being positively selected between the ancient and
752 modern strains, we sought to test whether the polymorphisms translate into a change in Meq function.
753 Meq is a regulator of transcriptional activity, and the strength of the transcriptional activation of target
754 genes is strongly linked to virulence(*73*). In order to compare transcriptional activation between
755 ancient and modern MDV strains, the Meq gene was cloned based on the sequence from the highest
756 coverage ancient sample (OL1385; Buda Castle, Hungary) alongside Meq from three modern MDV
757 strains (RB1B and Md5, both very virulent pathotypes). Meq forms a functional heterodimer with a
758 chicken protein – c-Jun – which is required for transcriptional activation, and so chicken c-Jun was
759 cloned and expressed alongside each Meq construct in DF-1 cells. As a reporter of transcriptional
760 activity, the Meq promoter (which contains an AP-1 binding site) in an expression vector upstream of
761 the firefly luciferase gene was co-transfected with the Meq/c-Jun pair.
762
763 As anticipated, Meq from the modern very virulent strains of MDV (RB-1B and Md5) showed the
764 greatest transactivation ability, with RB-1B Meq exceeding a ten-fold enhancement of luciferase
765 signal compared to empty vector (**Fig. 3c**). As previously reported, we found that c-Jun is important
766 for Meq function, exemplified by RB1B Meq transactivation being severely abrogated without c-Jun
767 co-expression. There was, however, a small transactivation effect of RB1B Meq in isolation, possibly
768 as a result of endogenous c-Jun in DF-1 cells forming a small amount of functional heterodimer. Most
769 significantly, we found that the ancient Meq sequence derived from the ancient Hungarian strain of
770 MDV (OL1385), was a very weak transactivator. Transactivation from this construct was still
771 significantly elevated relative to baseline, but was considerably less than Meq from modern strains.
772 These findings support the hypothesis that ancient Meq, and thus ancient MDV, was considerably less
773 virulent than modern strains.
774
775 Choice of tree root: extended description
776 When running the BEAST analysis with a phylogenetically unconstrained tree, we found that BEAST
777 placed the root immediately prior to sample with the earliest tip date (Fig. S6b). This is expected as
778 BEAST has no specific mechanism to determine the root without inclusion of an outgroup. As a
779 result, a tree topology determined by BEAST showed the ancient samples in a paraphyletic group. In
780 contrast, when the tree topology was determined (using maximum likelihood) while including an

outgroup (herpesvirus of turkeys; HVT), the ancient samples fall into a well-supported monophyletic clade (bootstrap: 94; Fig. S6a). Although a closely related virus, HVT is too divergent to have been included in the BEAST analysis as an outgroup along with MDV samples. To ensure accurate root placement, our BEAST analysis therefore used a tree that was constrained to the topology as determined by the maximum likelihood analysis with HVT as an outgroup, but with the HVT sequence removed.

BEAST results: extended description

The mean root age of the tree was estimated around 1602 CE, with the mean times of the most recent common ancestors (tMRCA) for the Eurasian and North American modern lineages tracing back to 1903 and 1909 CE respectively. The mean UCLD clock rate was estimated to be 7.69E-6 substitutions/site/year, an estimate that is in the same order of magnitude as other dsDNA viruses(*74*). The posterior age distributions estimated for the ancient MDV samples OL1385 (1803), OL1389 (1802), OL1986 (1594) and OL2272 (1821) resembled the corresponding posterior density functions of the radiocarbon dates (Fig. S8), which were used as empirical priors on the sample ages, showing that the MCMC chain sufficiently integrated across the uncertainty in the ages of the ancient MDV samples. One exception is OL1986; while there is some support from the radiocarbon dating for the sample dating to before 1627 CE, there is little support from the BEAST results. Similarly, the BEAST results show less support for OL1385 and OL1389 to be more recent than 1858 and 1871 CE, respectively.

Phylogenetic clustering analysis: extended description

A total of 4 phylogenetic clusters were identified by fastbaps (Fig. S11). The first cluster included the 4 ancient samples with the highest genome coverage, followed by a Eurasian and a North American cluster, and a cluster of modern samples of both Eurasian and North American origin. As the clustering is performed without being constrained by the BEAST MCC tree, the identified phylogenetic clusters further corroborate the presence of a monophyletic Eurasian and North American lineages (as it can be observed in the topology estimated by BEAST) as well as the presence of a “transitional” lineage between the two modern MDV lineages.

Read sharing between chicken and MDV: extended description

To test whether the chicken and MDV genomes share significant homology that could result in read sharing between the host and virus, we simulated short-read data from the MDV genome and mapped these against the chicken genome. We identified one region which shares significant homology between the chicken and MDV and could be a source of read sharing: MDV001/MDV080 is a RNA telomerase subunit (vTR), 435bp in length, that is presumed to be a recent gene capture from the chicken genome. There was no evidence of significant read sharing between the chicken and MDV genomes – depth of coverage in our highest coverage ancient sample (OL1385) over the homologous region (coordinates: 141336-141771; depth: 58.8 ± 24.8) was in line with the wider region (coordinates:140000-142000; depth 77.0 ± 47.9).

Supplementary Figures

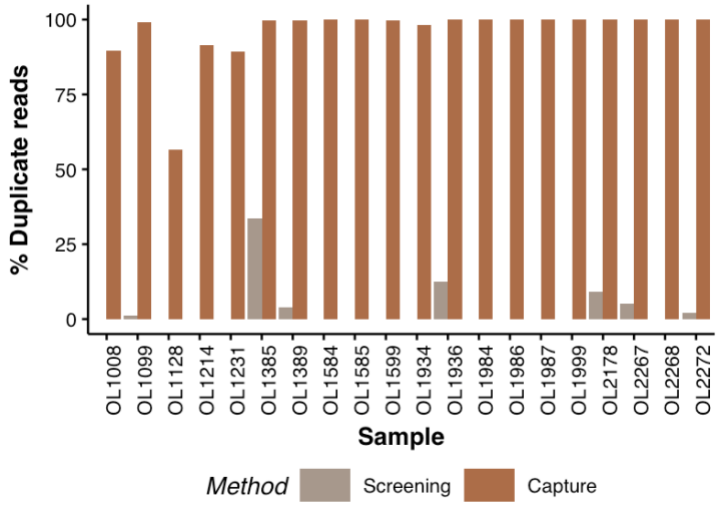


Fig. S1.

Percentage of duplicated reads after sequencing of screening and bait-captured libraries. In most cases, baiting of the sample resulted in near-saturation of duplicated reads, meaning almost all unique molecules in the sample were sequenced.

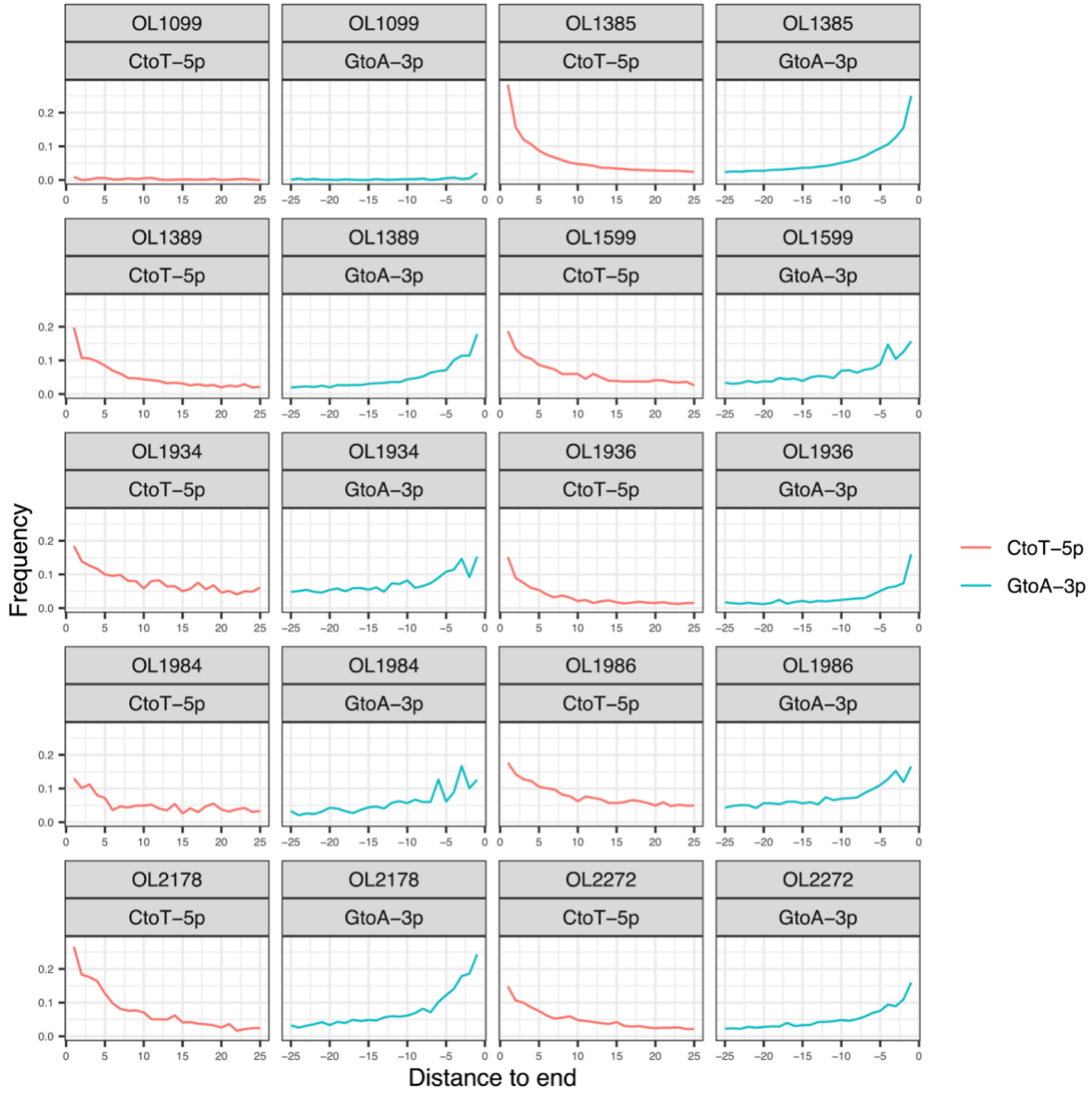


Fig. S2.

DNA damage profiles of DNA derived from archaeological samples. DNA reads from all samples with genomic coverage >1x were analyzed using the MapDamage program(*41*) which assesses the rate of C-to-T and G-to-A transitions as a result of the spontaneous deamination of cytosine. As a control, reads from a modern sample (OL1099; 2014 CE) were also included.

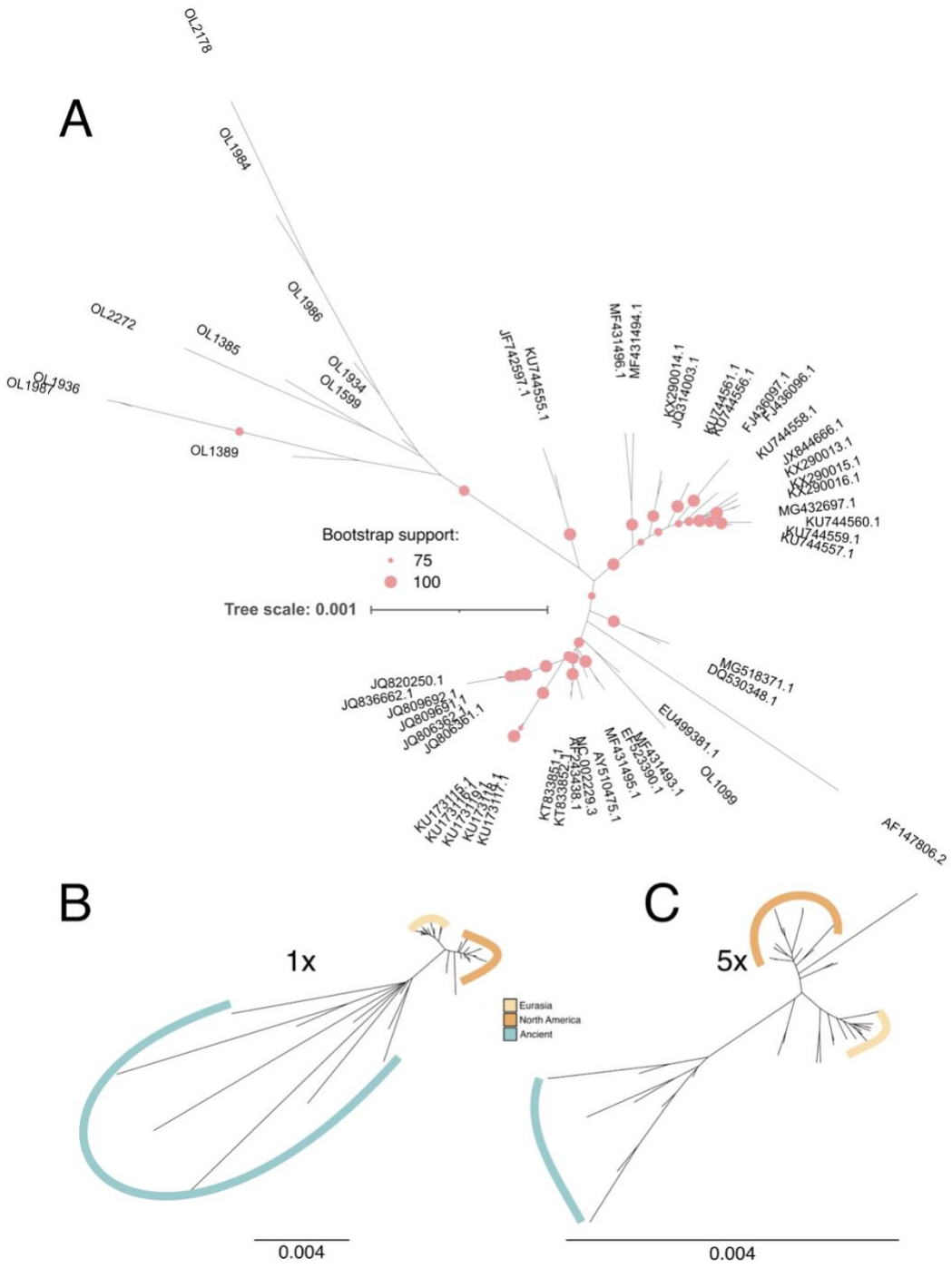


Fig. S3.

Neighbor-joining phylogenies of modern and ancient MDV genomes. (A) Bootstraps of ≥75 are displayed as filled circles. Captured samples (with 'OL' prefix) from the present study were included if the genome had ≥1% coverage at 5x. Differences in branch lengths for ancient samples are highlighted when using a depth of coverage threshold of 1x (B) and 5x (C).

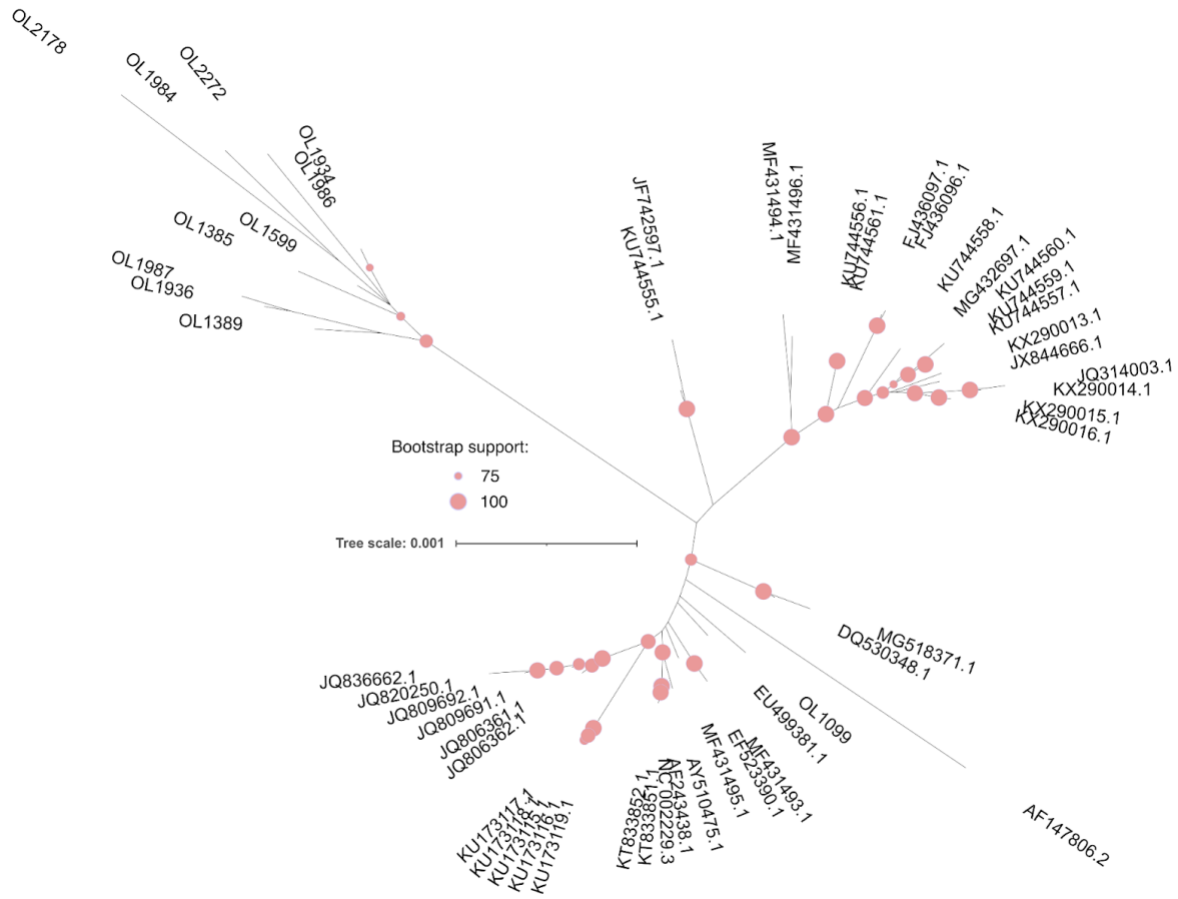


Fig. S4.

Maximum likelihood phylogeny of modern and ancient MDV genomes. Bootstraps of ≥75 are displayed as filled circles. Captured samples (with 'OL' prefix) from the present study were included if the genome had ≥1% coverage.

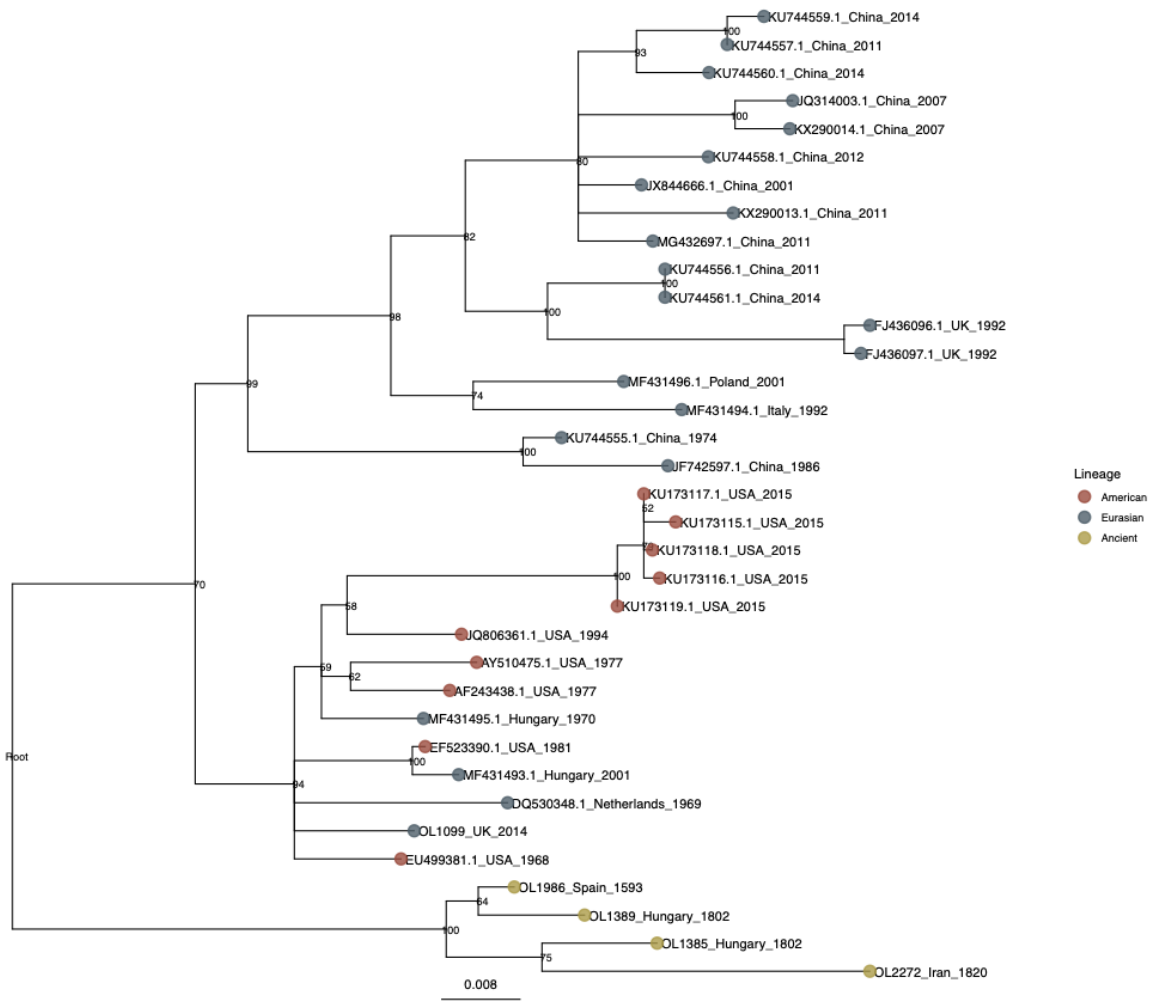


Fig. S5.

Midpoint-rooted maximum likelihood tree using transversion SNPs only. Tree was built using RAxML(*52*), with ascertainment correction for SNPs, and included all ancient samples with a genomic coverage of ≥20%. The placement of the root was confirmed using an outgroup (Fig. S6a).

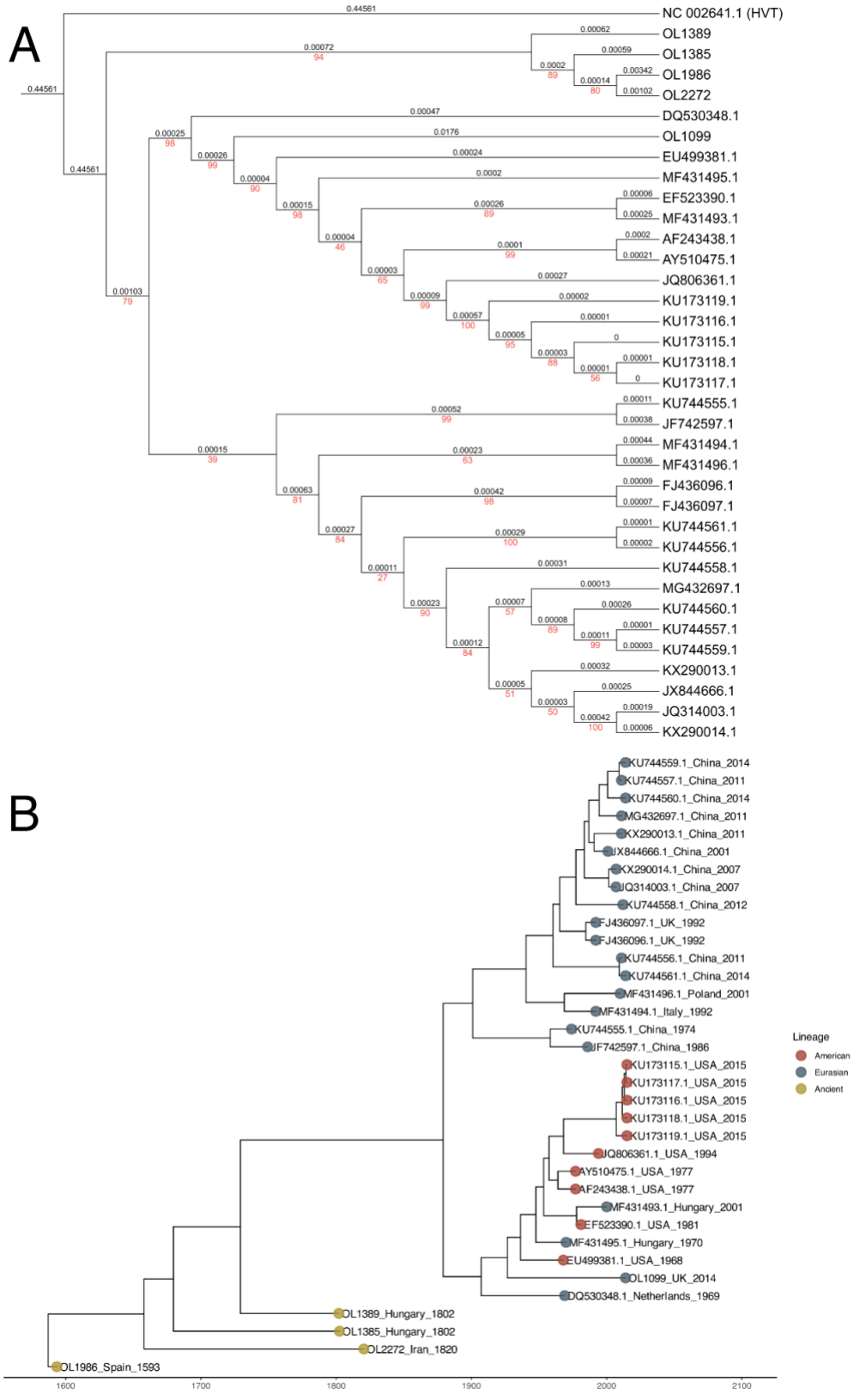


Fig. S6. Rooting of the MDV phylogeny. (Legend on next page).

Fig. S6. Rooting of the MDV phylogeny. (A) Tree was built using RAxML(*52*) and included all ancient samples from the present study with a genomic coverage of ≥20% along with 30 modern samples from public sources, one modern sample from the present study (OL1099) and the Meleagrid

herpesvirus 1 (Herpesvirus of turkeys, accession: NC_002641.1) as an outgroup. Because of the large distance between Meleagrid herpesvirus 1 and MDV strains, the tree is displayed as a cladogram with branch lengths in black and bootstrap values in red. (B) As a comparison, the unconstrained phylogeny was run in BEAST to allow the program to infer the root. BEAST determined the root to be immediately prior to the earliest sample included in the run (OL1986, Spain, 1593CE).

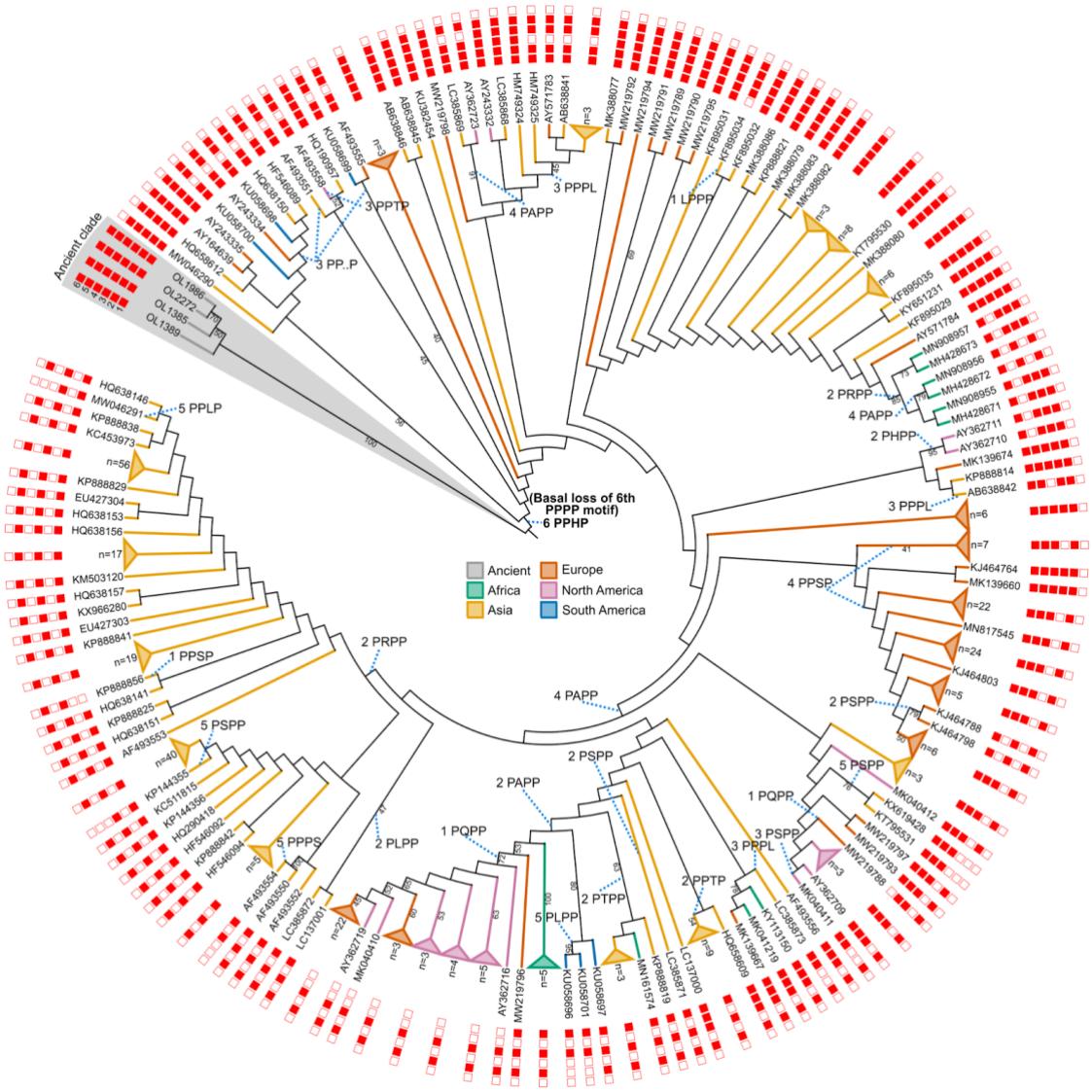
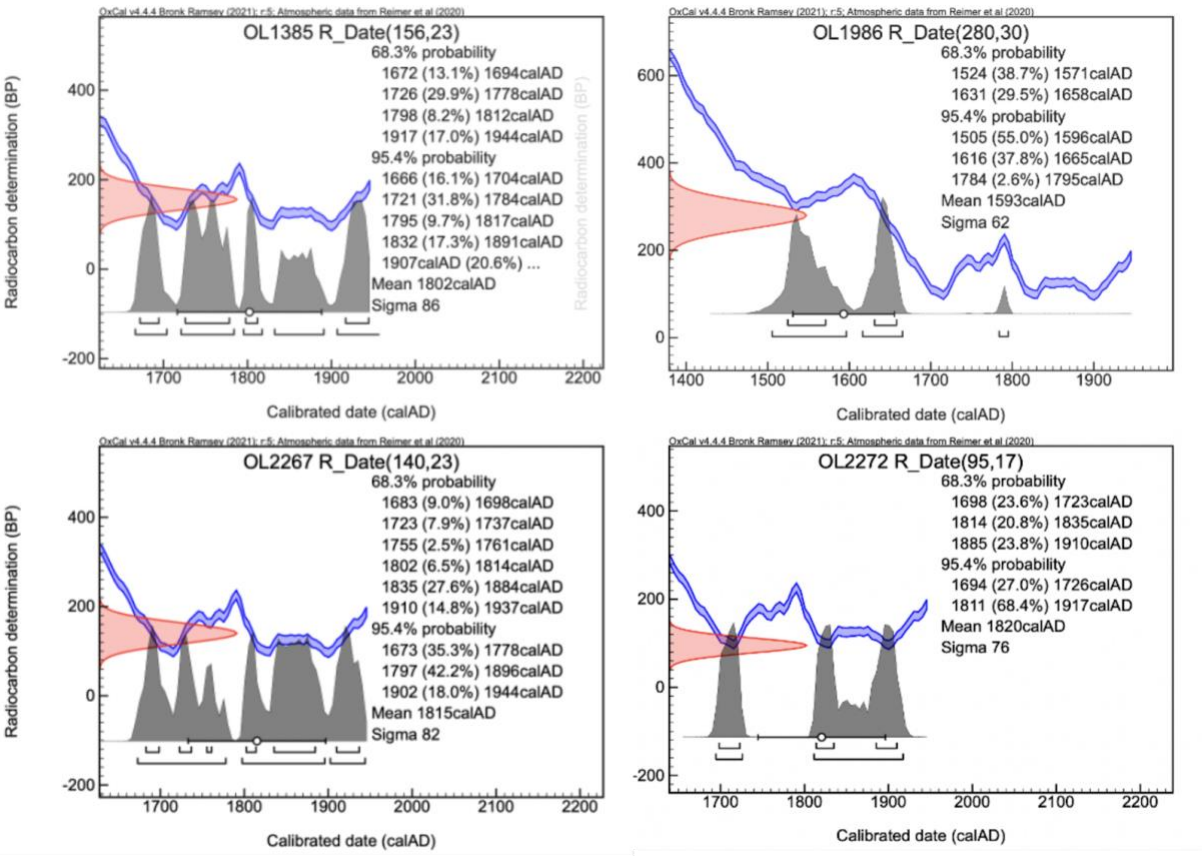


Fig. S7.

Meq phylogeny showing ordered loss of tetraproline motifs. The tree comprises 413 Meq sequences of the standard length (339aa) built using RAxML with 100 bootstrap replicates and visualized in iTOL(*72*). Around the edge, the integrity of each tetraproline motif is depicted (filled squares representing an intact tetraproline and open squares for a disrupted tetraproline). Meq sequences from ancient samples described herein are highlighted in gray, and the basal loss of the 6th tetraproline motif (common to all modern strains) is shown in bold. The label connected by a blue dotted line indicates the polymorphism that is found instead of tetraproline and the position of the tetraproline motif. For instance, '4 PAPP' indicates that the 4th tetraproline motif is disrupted by a proline-to-alanine substitution in the second proline position. '3 PP..P' denotes a deletion of the 3rd proline in the 3rd tetraproline motif.



873

874 **Fig. S8.**

875 **Probability density plots for radiocarbon dating of ancient samples**. Calibration of dates was done

876 in OxCal (https://c14.arch.ox.ac.uk/oxcal.html) using the IntCal20: Northern Hemisphere(*37*) method.

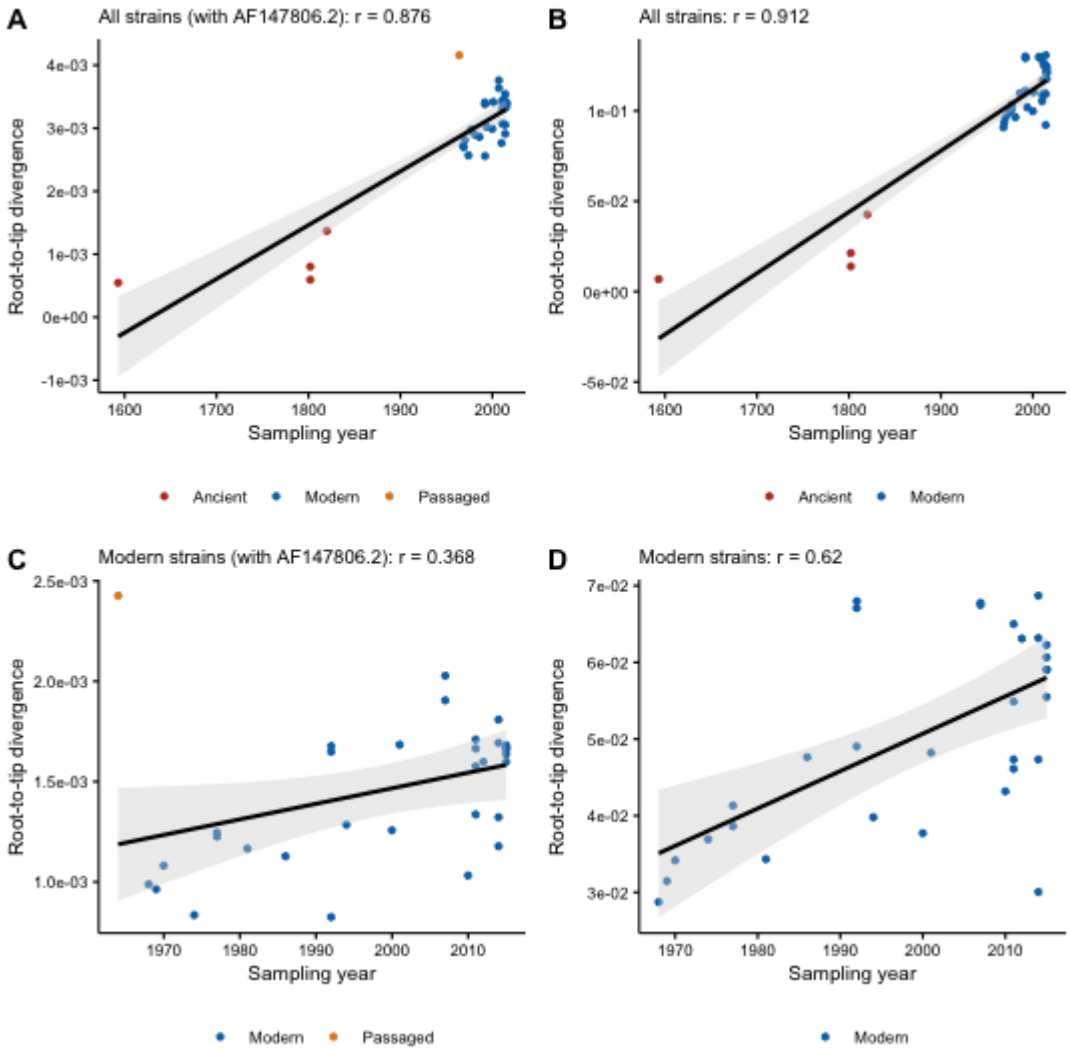


Fig. S9.

Root-to-tip divergence against sampling date of Marek's Disease Virus strains and respective Pearson correlation coefficients. (A) Ancient and modern strains (including AF147806.2, which was extensively passaged). **(B)** Ancient and modern strains (excluding AF147806.2). **(C)** Modern strains only (including AF147806.2) **(D)** Modern strains only (excluding AF147806.2).

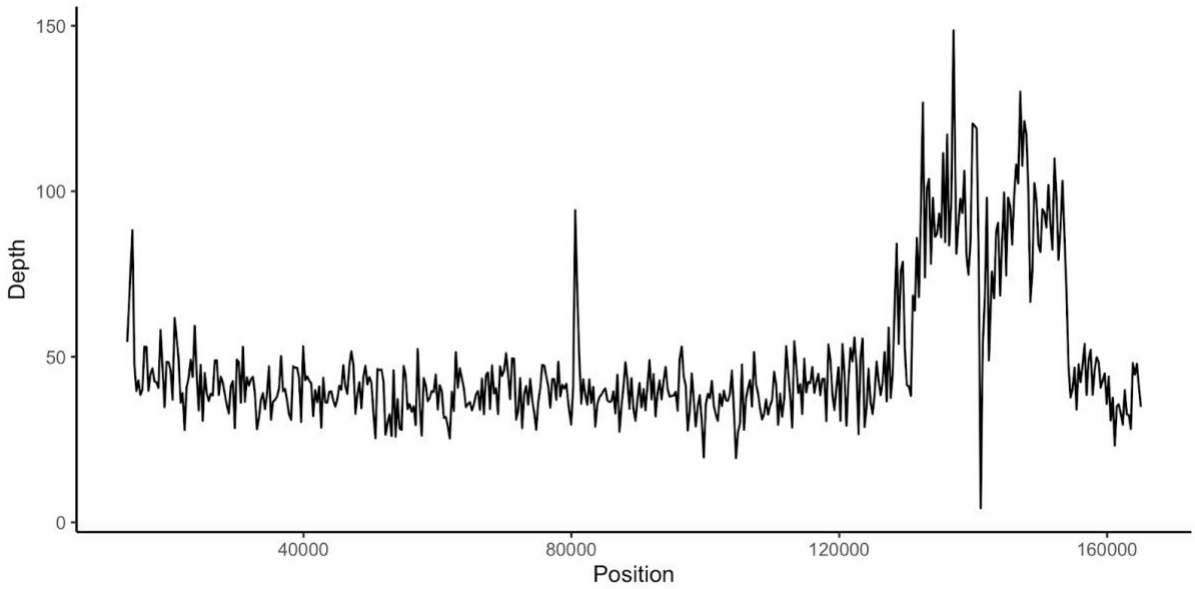


Fig. S10.

Sequencing depth of the highest coverage ancient sample (OL1385). Per-base sequencing depth averaged over 300bp windows illustrating no significant drop-off in sequence depth across the genome. The region of elevated coverage from ~130 kb is a duplicated region that encompasses the internal repeat long (IRL) and the internal repeat short (IRS). The sharp decline in coverage at ~142 kb is a short repetitive region that was not baited.

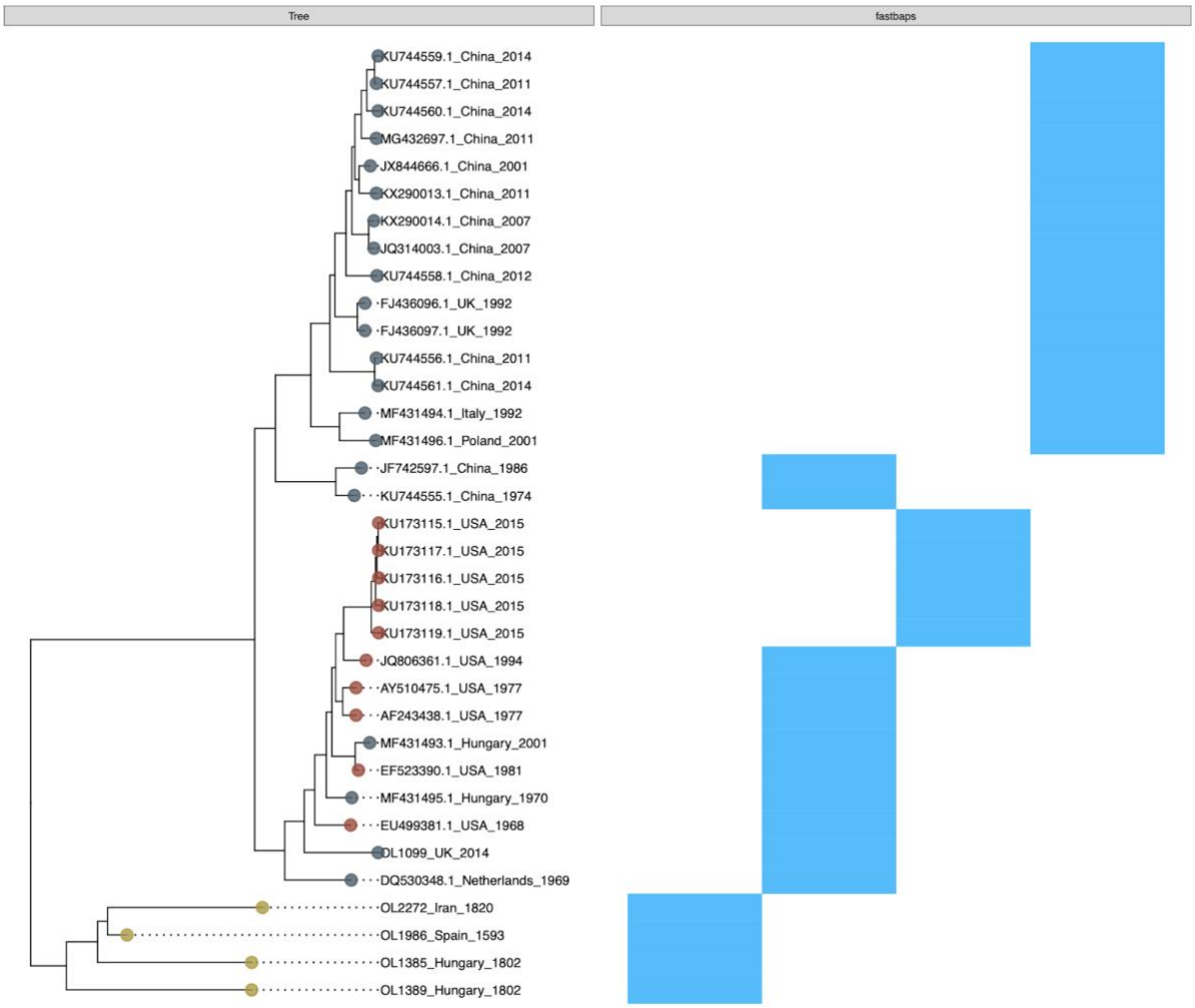


Fig. S11.

Phylogenetic clustering analysis using fastbaps. Phylogenetic clusters (blue bars) were identified in the dataset using the fastbaps algorithm, under the baps prior model and unconstrained by a phylogenetic tree.

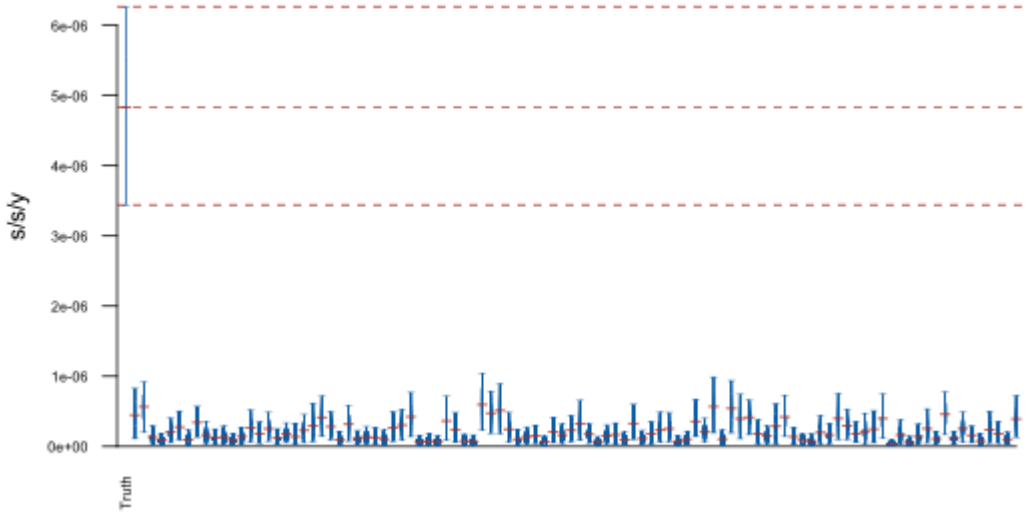


Fig S12.

The Bayesian date randomization test (DRT) performed with 100 replicates, under a UCLD clock model in BEAST v1.10.4. The plot shows the posterior distributions for the mean clock rate, using the true, unpermuted sampling dates (far left) and 100 replicates with sampling dates permuted among tips. Distributions are truncated at the upper and lower limits of the 95% HPD interval and horizontal red lines indicate the median estimates. The red dashed lines indicate the median and upper and lower limits of the 95% HPD interval of the clock rate inferred under the true sampling dates.

Supplementary Tables

Table S4. Summary of estimated parameters from the ME and UCLD clock model BEAST runs.

Mean ages tip dating runs	UCLD (95% HPD Interval)
Root age	1602 (1486, 1768)
tMRCA(Modern)	1881 (1822, 1929)
tMRCA(American superclade)	1909 (1868, 1944)
tMRCA(Eurasian)	1904 (1854, 1945)
age OL1389	1737 (1669, 1858)
age OL1385	1752 (1672, 1871)
age OL1986	1675 (1627, 1801)
age OL2272	1799 (1701, 1910)
Mean clock rate (substitutions/site/year)	9.90E-6 (5.42E-6, 1.50E-5)

Table S9.

Strain	Year (CE)	Pathotype	Meq coding length	Sequence source
OL1385 (ancient)	1802 ± 86	Unknown	1020 bp	This study
RB1B	1981	Very virulent	1020 bp	EF523390.1
Md5	1977	Very virulent	1020 bp	NC_002229.3

Meq sequences cloned for *in vitro* functional analysis. Meq was cloned from the highest coverage ancient sample (OL1385) as well as from two modern strains.

Table S10.

Sample	Country	Uncalibrated date (YBP)	Error +/-	Calibrated date	Error +/-	Laboratory	Lab Code
OL1385	Hungary	156	23	1802	86	Oxford RLAHA	OxA-40466
OL2267	Iran	140	23	1815	82	Oxford RLAHA	OxA-40467
OL2272	Iran	95	17	1820	76	Oxford RLAHA	OxA-40491
OL1986	Spain	280	30	1593	62	Beta Analytic	Beta-638008

Radiocarbon dating of ancient chicken samples. Raw dates were calibrated using IntCal20: Northern Hemisphere(*37*). See also Figure S8 for probability density plots.

Data S1.

- **Table S1**: **Sample metadata**. Metadata for all ancient samples sequenced in the present study.
- **Table S2**: **Screening and capture sequencing results**. Sequencing statistics for all MDV-positive screened samples that were then submitted for in-solution capture.
- **Table S3**: **Modern genome metadata**. Metadata for all modern MDV sequences included in phylogenetic analyses.
- **Table S6**: **Fixed differences between ancient and modern MDV strains**. Table of single nucleotide polymorphisms found to be fixed between ancient strains (with at least 2 representatives) and modern strains.
- **Table S7**: **PAML results**. Results from the branch-site analysis of positive selection.
- **Table S8**: **Meq metadata**. Metadata of all Meq sequences included in the Meq analysis, plus additional information about the tetraproline content for each sequence.
- **Table S11**: **Pre-capture metagenomic IDs**. Summary of the pre-capture HAYSTAC results.
- **Table S12**: **SNP summary table**. Summary table of all SNPs and depth of coverage in all ancient samples.
- **Table S13**: **BEAST tip dates**. Tip dates used as priors in the BEAST analysis.