DNA from dust: the first field-isolated genomes of MDV-1, from virions in poultry dust and chicken feather follicles

Utsav Pandey[1], Andrew S. Bell[2], Daniel Renner[1], David Kennedy[2], Jacob Shreve[1], Chris Cairns[2], Matthew Jones[2], Patricia Dunn[3], Andrew Read[2], Moriah L. Szpara[1*]

[1]Department of Biochemistry and Molecular Biology, and the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, 16802, USA

[2] Center for Infectious Disease Dynamics, Department of Biology and Entomology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

[3] Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, Pennsylvania, 16802, USA

***Corresponding Author**
Moriah L. Szpara
Dept. of Biochemistry & Molecular Biology
The Huck Institutes of the Life Sciences
W-208 Millennium Science Complex (MSC)
Pennsylvania State University
University Park, PA 16802 USA
Phone: 814-867-0008
Email: moriah@psu.edu

Author emails:
Utsav Pandey: uup104@psu.edu
Andrew S. Bell: asb15@psu.edu
Daniel W. Renner: dwr19@psu.edu
David Kennedy: dak30@psu.edu
Jacob T. Shreve: jtshreve@iu.edu
Chris Cairns: clc5303@psu.edu
Matthew Jones: mjj16@psu.edu
Patricia Dunn: pad7@psu.edu
Andrew Read: a.read@psu.edu
Moriah L. Szpara: moriah@psu.edu

Abstract (382 words at present)

Marek's disease (MD) is a lymphoproliferative disease of chickens caused by airborne gallid herpesvirus type 2 (GaHV-2, aka MDV-1). Mature virions are formed in the feather follicle epithelium cells of infected chickens from which the virus is shed as fine particles of skin and feather debris, or poultry dust. Poultry dust is the major source of virus transmission between birds in agricultural settings. Despite both clinical and laboratory data that show increased virulence in field isolates of MDV-1 over the last 40 years, we do not yet understand the genetic basis of MDV-1 pathogenicity. Our present knowledge on genome-wide variation in the MDV-1 genome comes exclusively from laboratory-grown isolates. MDV-1 isolates tend to lose virulence with increasing passage number *in vitro,* raising concerns about their ability to accurately reflect virus in the field. The ability to rapidly and directly sequence field isolates of MDV-1 is critical to understanding the genetic basis of rising virulence in circulating wild strains. Here we present the first complete genomes of uncultured, field-isolated MDV-1. These five consensus genomes were derived directly from poultry dust or single chicken feather follicles without passage in cell culture. These sources represent the shed material that is transmitted to new hosts, vs. the virus produced by a point source in one animal. We developed a new procedure to extract and enrich viral DNA, while reducing host and environmental contamination. DNA was sequenced using Illumina MiSeq high-throughput approaches and processed through a recently described bioinformatics workflow for *de novo* assembly and curation of herpesvirus genomes. We comprehensively compared these genomes to one another and also to previously described MDV-1 genomes. The field-isolated genomes had remarkably high DNA identity when compared to one another, with few variant proteins between them. In an analysis of genetic distance, the five new field genomes grouped separately from all previously

58 described genomes. Each consensus genome was also assessed to determine the level of
59 polymorphisms within each sample, which revealed that MDV-1 exists in the wild as a
60 polymorphic population. By tracking a new polymorphic locus in ICP4 over time, we found that
61 MDV-1 genomes can evolve in short period of time. Together these approaches advance our
62 ability to assess MDV-1 variation within and between hosts, over time, and during adaptation to
63 changing conditions.

64 Introduction

65 Marek's disease (MD) is a lymphoproliferative disease of chickens caused by airborne gallid
66 herpesvirus type 2 (GaHV-2), also known as Marek's disease virus 1 (MDV-1). MDV-1 is an
67 alphaherpesvirus, in the family *Herpesviridae*, genus *Mardivirus*. In addition to MDV-1, the
68 genus *Mardivirus* also contains the non-oncogenic MDV 2 (MDV-2, or gallid herpesvirus 3),
69 infectious laryngotracheitis virus (ILTV, or gallid herpesvirus 1) and turkey herpesvirus 1 (HVT-
70 1, or meleagrid herpesvirus type 1) (1–4). These Mardivirus family members belong to distinct
71 phylogenetic sub-groups of the Mardivirus family tree with low DNA sequence identity (e.g.
72 only 61% DNA identity between MDV-1 and MDV-2) (4).
73
74 Since the late 1960s, MD infections have been controlled via mass agricultural vaccinations of
75 poultry. All three serotypes, MDV-1, MDV-2 and HVT-1, have been used to make modified live
76 vaccines, which are employed either singly or in combination (3, 5). The vaccines currently
77 being used to control Marek's disease are live-attenuated vaccines that are administered either to
78 18-day old embryos or immediately after hatching. These vaccines prevent birds from
79 developing disease symptoms, but do not prevent these hosts from becoming infected, or block

80 transmission of the virus (8–10). The virulence of MDV-1 has risen in the last 40 years, which
81 has been attributed to changes in farming practices, as well as widespread vaccination (1, 11, 9,
82 10). Despite both clinical and laboratory data that demonstrate increased virulence in recent field
83 isolates of MDV-1, the genetics underlying MDV-1 evolution into more virulent forms is not
84 well understood. Understanding MDV-1 evolution in the field would provide new insight into
85 future evolution of this pathogen and facilitate precautionary measures to prevent outbreaks. A
86 few genes such as *Meq, UL36,* and *ICP4* have been associated with MDV-1 pathogenicity,
87 however, very little is known about their interaction partners during infection, or if these are the
88 only genes involved in the evolution of increased virulence (12).
89
90 Remarkably, our understanding of MDV-1 genomics and genetic variation comes exclusively
91 from the study of 10 different laboratory-grown strains (13–20). Most herpesviruses share this
92 limitation, where the large genome size and need for high-titer samples has led to a
93 preponderance of genome-studies being done with cultured virus, instead of clinical or field
94 samples (21–27). MDV-1 isolates tend to lose virulence with increasing passage number *in vitro*
95 (28, 29), as has been found for other herpesviruses such as herpes simplex virus and human
96 cytomegalovirus (24, 30). This raises concerns about the ability of cultured strains to accurately
97 reflect the genetic basis of virulence in wild circulating strains. Recent applications of next-
98 generation sequencing (NGS) have demonstrated that herpesvirus genomes can be captured from
99 human clinical samples using genome amplification techniques such as oligonucleotide
100 enrichment, and PCR-based approaches (31–36). The rising virulence of MDV in the field,
101 combined with these new approaches in next-generation sequencing of direct field samples, will
102 enable new insights into the genetic basis of rising virulence in circulating wild strains.

103

104 NGS approaches have the potential to reveal how genetic variation impacts the natural ecology
105 of MDV-1 in the field by identifying new genetic markers that correlate with increased
106 virulence, detecting loci that show allelic variation over time, and evaluating the extent of
107 standing variation in the field. The deep sequencing aspect of NGS makes it possible to sample
108 many members of a viral population, rather than just the dominant member. Large outputs
109 generated through NGS allow the analysis of genome-wide variations with higher confidence
110 than was possible with traditional approaches such as, restriction fragment length polymorphism
111 (RFLP) analysis and Sanger sequencing of individual genes. There is growing evidence for
112 sequence polymorphisms in large DNA virus genomes. Recent studies on human
113 cytomegalovirus (HCMV), a betaherpesvirus, have shown that populations of HCMV in human
114 patients can have levels of diversity that are comparable to RNA viruses such as dengue virus
115 (DENV) and HIV (33, 34, 37, 38). The potential for MDV to harbor limited heterogeneity has
116 been previously suggested in a serial passage study of cultured MDV-1 (28). Polymorphic
117 populations allow viruses to adapt to diverse environments and withstand changing selective
118 pressures, such as evading the host immune system, adapting to different tissue compartments,
119 and facilitating transmission between hosts (25, 32–34, 39–41). Drug resistance and vaccine
120 failure have also been attributed to the variation present in viral populations (32, 40, 41). The
121 ability to capture and sequence viral genomes directly from host infections and sites of
122 transmission will be crucial to decipher when and where variations arise, and which one(s)
123 spread into future host generations.

124

125 Here we present a method for the enrichment and isolation of viral genomes from poultry dust
126 and feather tips. Poultry dust is the source of transmitted infections, while feather follicle
127 epithelial cells are the only sites of fully productive virus replication. Deep sequencing of viral
128 DNA from these two materials enabled us to observe, for the first time, the complete genome of
129 MDV-1 directly from field samples. A subsequent comparison of these genomes provided our
130 first view of variation between source farms and individual birds. A further analysis of the rare
131 polymorphisms within each sample provided new insights into the potential for within-sample
132 variation and competition in the field. The MDV-1 isolation technique developed here is highly
133 effective in terms of purity, speed, and cost. When combined with suitable computational tools,
134 this method can be used to obtain full-length genomes of wild-type MDV-1 from diverse field
135 samples. This presents an unprecedented level of detail into the viral genomes that are shed from
136 infected hosts. As has been done with varicella zoster virus (VZV) vesicles in human patients
137 (32, 41), we can now study whether virus found in different body sites or individual feathers
138 harbors unique genetic variations. These approaches will reveal the genetic variability of MDV-1
139 that is present in poultry farms, allow us to map viral distribution and spread, and aid in the
140 identification of markers associated with virulence and vaccine resistance. The extension of these
141 methods to other field and clinical samples may enable similar new directions for human
142 herpesvirus infections.
143

144 # Methods

145 ## Collection of poultry dust and feathers

146 Two commercial-scale farms in central Pennsylvania (11 miles apart) were chosen for this study
147 due to a high incidence of MDV-1 **(Figure 1A)**. These operations raise broiler birds, housing
148 approximately 25,000-30,000 individuals per poultry house. The poultry on these farms were
149 vaccinated with bivalent vaccine composed of MDV-2 (strain SB-1) + HVT (strain FC126),
150 which allowed us to distinguish wild MDV-1 from concomitant shedding of vaccine strains.
151 These farms do not use the Rispens vaccine, which is derived from MDV-1. Poultry dust
152 samples were collected into 1.5 ml tubes from fan louvers. This location contains less moisture
153 and contaminants than floor-collected samples, and represents a mixture of air-borne virus
154 particles and feather dander. Sequential samples from Farm A (Table 1) were collected 11
155 months apart, from adjacent buildings on the same farm (**Figure 1A**).
156
157 Feathers were collected at animal maturity and just prior to processing for sale, to maximize the
158 potential for infection and high viral titer. At the time of collection the animals were ~12-weeks
159 old. Ten birds were chosen randomly throughout the entirety of each house for feather collection.
160 Two feathers from each bird were collected from the axillary track (breast feathers). One feather
161 from each bird was tested for the presence and quantity and MDV-1 present (see below for
162 quantitative PCR details). The remaining feather from the two birds with highest apparent MDV-
163 1 titer were used for a more thorough DNA extraction (see below for details) and next-generation
164 sequencing. Animal procedures were approved by the Institutional Animal Care and Use
165 Committee of the Pennsylvania State University (IACUC Protocol#: 46599).
166

167 **Viral DNA isolation from poultry dust**

168 MDV nucleocapsids were isolated from poultry dust as indicated in **Figure 2A**. Dust collected
169 from poultry houses was stored in 50 ml polypropylene Falcon® tubes (Corning) at 4°C until
170 required. 500 mg of poultry dust was suspended in 6.5 ml of 1X phosphate buffered saline
171 (PBS). The mixture was vortexed vigorously until homogenous and centrifuged at 2000 × g for
172 10 minutes. This supernatant was agitated on ice for 30 sec. using a Sonica Ultrasonic Processor
173 Q125 (probe sonicator with 1/8th inch microtip) set to 20% amplitude. It was then vortexed
174 before being centrifuged for a further 10 minutes at 2000 × g. This supernatant (approximately 5
175 ml in volume) was introduced to a 5 ml syringe and pushed through a 0.8 μM Corning®
176 surfactant-free cellulose acetate (SFCA) filter that had been soaked overnight in fetal bovine
177 serum (FBS). The flow-through was then passed through a Millipore Express® PLUS Membrane
178 (0.22 μM) vacuum filter and the membrane subsequently washed twice with 2.5 ml of PBS. The
179 resulting flow-through (approximately 10 ml in volume) was treated with DNase (Sigma) at a
180 concentration of 0.1 mg/ml for 30 minutes at room temperature. The MDV nucleocapsids present
181 in the DNase-treated solution were captured on a 0.1 μM polyethersulfone (PES) membrane
182 (VWR). An increased MDV purity, but ultimately reduced total nanograms of DNA yield, may
183 be achieved by washing this membrane once with 2.5 ml PBS (see Supplemental Tables S1-S3).
184 The membrane was then carefully excised using a sterile needle and forceps, and laid – exit side
185 downwards – in a sterile 5 cm diameter plastic petri-dish where it was folded twice lengthwise.
186 The "rolled" membrane was then placed into a 2 ml micro-tube containing 1.8 ml of lysis
187 solution (ATL buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen).
188 Digestion was allowed to proceed at 56°C for 1 hour on an incubating microplate shaker (VWR)
189 set to 1100 rpm. The membrane was then removed, held vertically over a tilted sterile 5 cm

190 diameter plastic petri-dish and washed with a small volume of the lysis solution (from the 2 ml
191 micro-tube). This wash was subsequently returned to the 2 ml micro-tube and the tube replaced
192 on the heated shaker where it was allowed to incubate overnight. The following day, the DNA
193 was isolated as per manufacturer's instructions using the DNeasy® Blood and Tissue kit
194 (Qiagen). DNA was eluted in 200 µl DNase-free water. Ten to fourteen aliquots of 500 mg each
195 were used to obtain sufficient DNA for each dust sample (see Supplemental Tables S1-S3).
196

197 **Isolation of viral DNA from feather follicles**

198 Individual feathers were plucked from the breast region and the distal 0.5-1.0 cm proximal shaft
199 (feather tip, which contains the feather pulp) was snipped into a sterile 1.5 ml micro-tube
200 containing a single sterile 5 mm steel bead (Qiagen). On return to the laboratory, tubes were
201 stored at -80°C until processing. Each tube containing a single feather tip was allowed to thaw,
202 and then 200 µl of PBS was added and the sample bead-beaten for 30 seconds at 30 Hz using a
203 Tissuelyser (Qiagen) (**Figure 2B**). A further 720 µl of PBS and 80 µl of 2.5 mg/ml trypsin
204 (Sigma) were then added (final trypsin concentration: 0.8 mg/ml), and the solution was
205 transferred to a new sterile 2 ml micro-tube. Digestion was allowed to proceed for 2 hours at
206 37°C on an incubating microplate shaker (VWR) set to 700 rpm. The suspension was then
207 sonicated on ice for 30 seconds using a Sonica Ultrasonic Processor Q125 (probe sonicator with
208 1/8th inch microtip) set to 50% amplitude. DNase I was added to a final concentration of 0.1
209 mg/ml and allowed to digest for 1 hour at room temperature. An equal volume of lysis solution
210 (ATL buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen) was added and
211 the sample was incubated over night at 56°C on an Incubating Microplate Shaker (VWR) set to

212 1100 rpm. The following day, the DNA was isolated as per manufacturer's instructions using the
213 DNeasy® Blood and Tissue kit (Qiagen).
214

215 **Measurement of total DNA and quantification of viral DNA**

216 The total amount of DNA present in the samples was quantified by fluorescence analysis using a
217 Qubit® fluorescence assay (Invitrogen) following the manufacturer's recommended protocol.
218 MDV genome copy numbers were determined using serotype-specific quantitative PCR (qPCR)
219 primers and probes, targeting either the MDV-1 pp38 (LORF14a) gene or MDV-2 (SB-1 strain)
220 DNA polymerase (DNA-Pol) gene. The MDV-1 assay was designed by Sue Baigent: forward
221 primer (Spp38for) 5'-GAGCTAACCGGAGAGGGAGA-3'; reverse primer (Spp38rev) 5'-
222 CGCATACCGACTTTCGTCAA-3'; probe (MDV-1) 6FAM-CCCACTGTGACAGCC-BHQ1
223 (S. Baigent, *pers. comm.*). The MDV-2 assay is that of Islam et al. (42), but with a shorter MGB
224 probe in place of their BHQ-2 probe (6FAM-GTAATGCACCCGTGAC-MGB). Real-time
225 quantitative PCRs were performed on an ABI Prism 7500 Fast System with an initial
226 denaturation of 95°C for 2 minutes followed by 40 cycles of denaturation at 95°C for 3 seconds
227 and annealing and extension at 60°C for 30 seconds. Both assays included 4 µl of DNA in a total
228 PCR reaction volume of 20 µl with 1X PerfeCTa™ qPCR FastMix™ (Quanta Biosciences),
229 forward and reverse primers at 300 nM and TaqMan ® BHQ (MDV-1) or MGB (MDV-2)
230 probes (Sigma and Life Sciences, respectively) at 100 nM and 200 nM, respectively. In addition
231 each qPCR reaction incorporated 2 µl BSA (Sigma). Absolute quantification of genomes was
232 based on a standard curve of serially diluted plasmids cloned from the respective target genes.
233 The absolute quantification obtained was then converted to concentration. The concentration of
234 total DNA present was calculated using a Qubit® fluorometer. Once the concentration of the

235 total DNA, MDV-1, and MDV-2 DNA present in the sample were known, we calculated the
236 percentage of MDV-1 and MDV-2 genomic DNA in the total DNA pool (see Supplemental
237 Tables S1-S4).
238

239 **Illumina next-generation sequencing**

240 Sequencing libraries for each of the isolates were prepared using the Illumina TruSeq Nano DNA
241 Sample Prep Kit, according to the manufacturer's recommended protocol for sequencing of
242 genomic DNA. Genomic DNA inputs used for each sample are listed in **Table 1**. The DNA
243 fragment size selected for library construction was 550 base pairs (bp). All the samples were
244 sequenced on an in-house Illumina MiSeq using version 3 sequencing chemistry to obtain
245 paired-end sequences of 300 × 300 bp. Base calling and image analysis was performed with the
246 MiSeq Control Software (MCS) version 2.3.0.
247

248 **Consensus genome assembly**

249 As our samples contained many more organisms than just MDV, we developed a computational
250 workflow to preprocess our data prior to assembly. A local BLAST database was created from
251 every *Gallid herpesvirus* genome available in GenBank. All sequence reads for each sample
252 were then compared to this database using BLASTN (43) with a loose e-value less than or equal
253 to 10^{-2} in order to computationally enrich for sequences related to MDV. These "MDV-like"
254 reads were then processed for downstream genome assembly.
255

256 MDV genomes were assembled using the viral genome assembly VirGA (44) workflow which
257 combines quality control preprocessing of reads, *de novo* assembly, genome linearization and
258 annotation, and post-assembly quality assessments. For the reference-guided portion of viral
259 genome assembly in VirGA, the Gallid herpesvirus 2 (MDV-1) strain MD5 was used (GenBank
260 Accession: NC_002229). These new genomes were named according to recent
261 recommendations, as outlined by Kuhn et al (45). We use shortened forms of these names
262 throughout the manuscript (see Table 1 for short names). The full names for all five genomes are
263 as follows: MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 1; MDV-1
264 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 2; MDV-1 *Gallus domesticus*-
265 wt/Pennsylvania, USA/2015/Farm B-dust; MDV-1 *Gallus domesticus*-wt/Pennsylvania,
266 USA/2015/Farm B-feather 1; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm B-
267 feather 2.
268

269 **Between-sample: consensus genome comparisons**

270 Clustalw2 (43) was used to construct pairwise global nucleotide alignments between whole
271 genome sequences, and pairwise global amino acid alignments between open reading frames.
272 These alignments were utilized by downstream custom Python scripts to calculate percent
273 identity, protein differences, and variation between samples.
274

275 The proline-rich region of MDV049, which contains an extended array of tandem repeats of the
276 amino acids PQ, was removed from all five consensus genomes prior to comparison. The amount
277 of polymorphism seen in this region of MDV049 is driven by fluctuations in the length of these
278 tandem repeats, as has been seen in prior studies with other alphaherpesviruses such as HSV,

279 VZV, and pseudorabies virus (PRV) (32,48–50). Since the length of extended arrays of perfect
280 repeats cannot be precisely determined by *de novo* assembly (21, 22, 25, 26), we excluded this
281 region from pairwise comparisons of genome-wide variation.
282

283 **Within-sample: polymorphism detection within each consensus genome**

284 VarScan v2.2.11 (48) was used to detect variants present within each consensus genome. To aid
285 in differentiating true variants from potential sequencing errors (49), two separate variant calling
286 analyses were explored. (41). Our main polymorphism-detection parameters (used in Figures 4-5
287 and Supplemental Tables S5-S6) were as follows: minimum variant allele frequency ≥ 0.02; base
288 call quality ≥ 20; read depth at the position ≥ 10; independent reads supporting minor allele ≥ 2.
289 Directional strand bias ≥ 90% was excluded; a minimum of two reads in opposing directions was
290 required. For comparison and added stringency, we also explored a second set of parameters
291 (used in Supplemental Figure S2): minimum variant allele frequency ≥ 0.05; base call quality ≥
292 20; read depth at the position ≥ 100; independent reads supporting minor allele ≥5. Directional
293 strand bias ≥ 80% was excluded. The variants obtained from VarScan were then mapped back to
294 the genome to understand their distribution and mutational impact using SnpEff and SnpSift (50,
295 51). Polymorphisms in the proline-rich region of MDV049 were excluded, as noted above.
296

297 **Testing for signs of selection acting on polymorphic viral populations**

298 For each of our five consensus genomes, which each represent a viral population, we classified
299 the polymorphisms detected into categories of synonymous, non-synonymous, genic-
300 untranslated, or intergenic, based on where each polymorphism was positioned in the genome.

301 For these analyses (Figure 5), we were only able to include polymorphisms detected in the three
302 dust genomes, since the total number of polymorphisms obtained from feather genomes was too
303 low for chi-square analysis. First, we calculated the total possible number of single nucleotide
304 mutations that could be categorized as synonymous, non-synonymous, genic-untranslated or
305 intergenic. To remove ambiguity when mutations in overlapping genes could be classified as
306 either synonymous or non-synonymous, genes with alternative splice variants or overlapping
307 reading frames were excluded from these analyses. This removed 25 open reading frames
308 (approximately 21% of the genome). This tally of potential mutational effects constituted the
309 expected number of mutations in each category. We preformed chi-squared tests on each dataset
310 to assess whether the observed distribution of polymorphisms matched the expected distribution.
311 We also performed a similar analysis in pairwise fashion (**Supplemental Table S6**), to assess
312 whether the fraction of variants differed from what would be expected by random chance.
313 Pairwise combinations included the following: synonymous vs. non-synonymous, synonymous
314 vs. intergenic, synonymous vs. genic-untranslated, non-synonymous vs. intergenic, non-
315 synonymous vs. genic-untranslated, and intergenic vs. genic-untranslated. Statistically significant
316 outcomes would suggest that recent or historical selection differed between those categories of
317 variants.
318

319 **Sanger sequencing of polymorphic locus in ICP4**

320 A potential locus of active selection within the ICP4 (MDV084) gene was detected during deep-
321 sequencing of Farm B-dust. This locus was examined using Sanger sequencing. An
322 approximately 400 bp region of the ICP4 gene was amplified using a Taq PCR Core Kit
323 (Qiagen) and the following primers at 200 nM: forward primer (ICP4selF)

324 5'AACACCTCTTGCCATGGTTC 3'; reverse primer (ICP4selR)
325 5'GGACCAATCATCCTCTCTGG 3'. Cycling conditions included an initial denaturation of
326 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at
327 55°C for 30 seconds and extension at 72°C for 1 minute, with a terminal extension at 72°C for 10
328 minutes. The total reaction volume of 50 µl included 10 µl of DNA and 4 µl BSA (final
329 concentration 0.8 mg/ml). Amplification products were visualized on a 1.5% agarose gel, the
330 target amplicon excised and then purified using the E.Z.N.A. Gel Extraction Kit (Omega Bio-
331 tek). Sanger sequencing was performed by the Penn State Genomics Core Facility utilizing the
332 same primers as used for DNA amplification. The relative peak height of each base call at the
333 polymorphic position was analyzed using the ab1PeakReporter tool (52).
334

335 **Genetic distance and dendrogram**

336 Multiple sequence alignments of complete MDV-1 (*Gallid herpesvirus 2*) genomes from
337 GenBank (13, 16–20, 53–55) and those assembled by our lab were generated using MAFFT (56).
338 The evolutionary distances were computed using the Jukes-Cantor method (57) and the
339 evolutionary history was inferred using the neighbor-joining method (58) in MEGA6 (59), with
340 1,000 bootstrap replicates (60). Positions containing gaps and missing data were excluded.
341

342 **Taxonomic estimation of non-MDV sequences in poultry dust and feathers**

343 All sequence reads from each sample were submitted to a quality control preprocessing method
344 to remove sequencing primers, artifacts, and areas of low confidence (44). Sequence annotation
345 was performed using a massively iterative all-vs.-all BLASTN (E-value 10^{-2}) approach using the

346 all-nucleotide-database from NCBI. Only a portion of the total sequence read pool could be
347 identified with confidence using this method. We then used *de novo* assembly to extend the
348 length of these unidentified sequences, therefore elongating them into contigs. These were
349 iterated through BLASTN again, which revealed alignment to repetitive regions of the *Gallus*
350 *domesticus* (chicken) genome. Since the viral DNA enrichment procedures include a level of
351 stochasticity in removal of host and environmental contaminants, the proportion of taxa present
352 is not a definitive outline of those present initially. The results of these classifications are shown
353 in **Supplemental Figure S3** and listed in **Supplemental Table S7**.
354

355 Results and Discussion

356 Enrichment of capsids and isolation of viral DNA from poultry dust and feathers

357 Field samples containing MDV were collected from poultry dust and feathers of two large-scale
358 agricultural operations (>25,000 animals; **Figure 1A**). In pilot DNA extractions, we found that
359 poultry dust contained less than 0.04% MDV-1 DNA, and feather follicles contained less than
360 1% (data not shown). We thus developed a protocol that used the proteinaceous C-type capsid of
361 MDV to protect the viral DNA during initial enrichment from the milieu of environmental source
362 material (**Figure 2**) (61). Due to the complexity of contaminating environmental sources of
363 DNA, enrichment of viral DNA from poultry dust involved a more extensive procedure than that
364 required for feathers (**Figure 2A vs. 2B**). Initial steps in the isolation of capsids from the poultry
365 dust were important for the release of cell-associated virus into the supernatant. Vortexing
366 distributed the dust into solution and aided in the release of virions from the cells. Centrifugation
367 allowed large dust particles to sediment out, leaving a portion of virus in solution. Sonication of

368 the resulting supernatant further aided in release of virus. A second round of centrifugation
369 removed additional particulates from the supernatant. To enrich viral capsids away from the
370 remaining contaminants, we then utilized a series of filtration techniques. First we passed the
371 supernatant through a 0.8 µm filter, to filter out dust particles, host cells and bacteria. The flow-
372 through was then passed through a 0.22 µm filter, to remove additional bacteria and fungal
373 spores. Next we treated the flow-through from the 0.22 µm filter with DNase, to remove
374 contaminant DNA. In the absence of DNase treatment we observed a higher yield of viral DNA,
375 but with much lower purity (data not shown). Finally the DNase-treated suspension was passed
376 over a 0.1 µm filter. This filter membrane trapped the viral nucleocapsids, which are between
377 0.1-0.2 µm (62). The filter membrane with trapped virions was excised and digested to open the
378 capsids and release the DNA. Quantitative PCR was used to assess the copy number of viral
379 genomes in the resulting DNA. Total yield and percent MDV-1 vs. MDV-2 DNA are listed in
380 **Supplemental Tables S1-S3**. When the captured capsids were washed with PBS before excising
381 the membrane, the percent MDV DNA in the total DNA pool increased, while the total DNA
382 yield decreased. In the future, samples with a higher percentage of MDV DNA could be obtained
383 by applying these wash steps to all components of the sample pool.
384
385 A different protocol was developed for extraction of MDV DNA from feather follicles, to
386 accommodate both the smaller input material and our expectation of higher purity (**Figure 2B**).
387 Sequential size filters were not used to filter out contaminants from feather follicles as these
388 direct host samples have fewer impurities than the environmental samples of poultry dust.
389 However, the feather follicle cells were encased inside the keratinaceous shell of the feather tip,
390 which required disruption to release the cells. Vigorous bead-beating achieved the desired

391 destruction of the follicle tip. Next, the feather follicle cells were digested with trypsin to remove
392 the extracellular matrix and to dissociate cells. This suspension was then sonicated to release
393 cell-associated virus, and it was treated with DNase to remove non-encapsidated DNA. Finally,
394 the remaining viral capsids were digested to release the DNA. While the overall amount of DNA
395 obtained from feather follicles was lower than that obtained from pooled dust samples
396 (**Supplemental Table S4)**, it was of higher purity and was sufficient to generate libraries for
397 sequencing (**Table 1**, lines 1-3).
398

399 **Sequencing and assembly of wild MDV-1 genomes**

400 A total of five uncultured wild-type samples of MDV were sequenced on an in-house Illumina
401 MiSeq sequencer (**Table 1**, lines 4-6; see **Methods** for details). Following the nomenclature
402 conventions for viral field isolates, we have named these MDV-1 genomes according to location
403 and sample type (see **Methods** for full isolate name). Here we use shortened names for clarity,
404 e.g. Farm A-dust 1 (**Table 1**). The farms from which we sourced these samples use a bivalent
405 vaccine composed of HVT and MDV-2 strain SB-1. These farms do not use Rispens, which is an
406 alternative vaccine derived from attenuation of MDV-1. The use of bivalent vaccine made it
407 possible for us to readily distinguish sequence reads that resulted from the shedding of virulent
408 MDV-1 vs. vaccine virus (MDV-2 or HVT) strains. The overall DNA identity of MDV-1 and
409 MDV-2 is just 61% (4). In a comparison of strains MDV-1 Md5 (NC_002229) and MDV-2 SB-1
410 (HQ840738), we found no spans of identical DNA greater than 50 bp (data not shown). This
411 allowed us to accurately distinguish these 300 x 300 bp MiSeq sequence reads as being derived
412 from either MDV-1 or MDV-2. After deep sequencing, a BLAST-based filter step was used to
413 computationally enrich MDV-like sequence reads, separating them from the many contaminants

414 in the poultry dust samples (**Supplemental Figure 1**). These sequences were then further
415 separated into MDV-1-like sequences or MDV-2 like sequences.
416
417 The sequence read data derived from poultry dust contained approximately 2-5% MDV-1 DNA,
418 while the feather samples ranged from ~27%-48% MDV-1 (**Table 1**, line 6). Although the two
419 feathers had different proportions of MDV-1 before sequencing (**Table 1**, line 4), approximately
420 half of the sequencing output was MDV-1 specific for each sample (**Table 1**, line 6). Enrichment
421 of MDV-1 specific sequences may be a result of steps involved in the preparation of genomic
422 DNA for sequencing. The shearing of genomic DNA was optimized using MDV DNA, which
423 may have favored the removal of sub-optimally sheared contaminant DNA during subsequent
424 size selection steps. Based on these results, we concluded that feathers were a better source of
425 MDV-1 DNA than poultry dust. However, since poultry dust represents the infectious material
426 that transmits MDV from host to host, and across generations of animals that pass through a farm
427 or field site, we pursued analysis of wild MDV genomes from both types of source material.
428

429 **Annotation of new MDV-1 consensus genomes from the field**

430 Consensus genomes were created for each of the five samples in **Table 1**, using a recently
431 described combination of *de novo* assembly and reference-guided alignment of large sequence
432 blocks, or contigs (44). Like other alphaherpesviruses, MDV genomes have a class E
433 organization (**Figure 3A**) that consists of unique long (UL) and unique short (US) regions (63,
434 64). Each unique region is flanked by inverted repeats, which are named according to their
435 position (terminal, T, vs. internal, I), and whether they flank the unique long (L) or short (S)
436 region (TRL, IRL, TRS, IRS; see **Figure 3A** for illustration). Nearly complete genomes were

437 obtained for all five samples (**Table 1**). The coverage depth for each genome was directly
438 proportional to the number of MDV-1 specific reads obtained from each sequencing library
439 (**Table 1**, line 5,7). The poultry dust sample from Farm B had the highest coverage depth, at an
440 average of almost 600X across the viral genome. Feather 1 from Farm B had the lowest coverage
441 depth, averaging 44X genome-wide, which still exceeds that of most bacterial or eukaryotic
442 genome assemblies. The genome length for all 5 samples was approximately 180 kilobases
443 (**Table 1**), which is comparable to all previously sequenced MDV-1 isolates (13, 15–20, 54).
444
445 For each field sample collected and analyzed here, we assembled a consensus viral genome. We
446 anticipated that the viral DNA present in a single feather follicle might be homotypic, based on
447 similar results found for individual vesicular lesions of VZV (32, 41). We further expected that
448 the genomes assembled from a poultry dust sample would represent a mix of viral genomes,
449 summed over time and space. Viral genomes assembled from poultry dust represent the most
450 common genome sequence, or alleles therein, from all of the circulating MDV-1 on a particular
451 farm. Our subsequent analyses included both a comparison between the consensus genomes, and
452 a search for variability within the viral population that was compiled into each consensus
453 genome. The comparison of consensus genomes provided a view into the amount of sequence
454 variation between Farm A and Farm B, or between two birds on the same Farm (**Table 2**). In
455 contrast, examining the polymorphic loci within each consensus genome assembly allowed us to
456 observe the level of variation within the viral population at each point source (**Table 3, Figures**
457 **4-5**).
458

459 **DNA and amino acid variations between five new field genomes of MDV-1**

460 We began our assessment of genetic diversity by determining the extent of DNA and amino acid
461 variations between the five different consensus genomes. We found that the five genomes are
462 highly similar to one another at the DNA level, with the percent homology ranging from 99.4%
463 to 99.9% in pairwise comparisons (**Table 2**). These comparisons used a trimmed genome format
464 (**Figure 3B**) where the terminal repeat regions had been removed, so that these sequences were
465 not over-represented in the analyses. The level of identity between samples is akin to that
466 observed in closely related isolates of HSV-1 (44). Observed nucleotide differences were
467 categorized as genic or intergenic, and further sub-divided based on whether the differences were
468 insertions or deletions (INDELs) or single-nucleotide polymorphism (SNPs) **(Table 2)**. The
469 number of nucleotide differences was higher in intergenic regions than in genic regions for all
470 genomes. For the INDEL differences, we also calculated the minimum number of events that
471 could have led to the observed differences, to provide context on the relative frequency of these
472 loci in each genome. We anticipate that these variations include silent mutations, as well as
473 potentially advantageous or deleterious coding differences.
474
475 To understand the effect(s) of these nucleotide variations on protein coding and function, we
476 next compared the amino acid (AA) sequences of all open reading frames (ORFs) for the five
477 isolates. The consensus protein coding sequences of all five isolates were nearly identical, with
478 just a few differences (**Table 2**). In comparison to the other four samples, Farm B-dust harbored
479 AA substitutions in four proteins. A single non-synonymous mutation was seen in each of the
480 following: the lipase MDV010, the unknown protein MDV012, and the probable membrane
481 protein MDV056. A single synonymous mutation was observed in the DNA helicase-primase

482 protein MDV020. Finally, a 66 AA insertion unique to Farm B-dust was observed in the DNA
483 polymerase processivity subunit protein MDV055. We did not observe any coding differences
484 between temporally separated dust isolates from Farm A or between feather isolates from
485 different hosts in Farm B, although both of these comparisons (Table 2, bottom) revealed
486 hundreds of noncoding differences. The fact that any coding differences were observed in this
487 small sampling of field viruses suggests that the natural ecology of MDV-1 may include drift and
488 adaptation in protein function, in addition to genetic drift.
489

490 **Detection of polymorphic bases within each genome**

491 Comparing viral genomes found in different sites provides a macro-level assessment of viral
492 diversity. We next investigated the presence of polymorphic viral populations within each
493 consensus genome, to reveal how much diversity might exist within a farm (as reflected in
494 poultry dust genomes) or within a single host (as reflected in feather genomes). For each
495 consensus genome, we used polymorphism detection analysis to examine the depth and content
496 of the sequence reads at every nucleotide position in each genome (see **Methods** for details).
497 Rather than detecting differences between isolates, as in **Table 2**, this approach revealed
498 polymorphic sites within the viral population that contributed to each consensus genome. We
499 detected 2-58 polymorphic sites within each consensus genome (**Figure 4**). The feather genomes
500 had a lower number of polymorphisms as compared to the dust genomes, which may be
501 attributable to their relatively low sequence coverage. Since INDELs were not included in this
502 polymorphism analysis, but clearly contributed to between-sample variation (**Table 2**), this may
503 be an underestimate of the overall amount of within-sample variation. Viral polymorphisms were
504 distributed across the entire length of the genome (**Figure 4**), with the majority concentrated in

505 the repeat regions. Application of a more stringent set of parameters (see **Methods** for details)
506 yielded a similar distribution of polymorphisms, albeit with no polymorphisms detected in
507 feather samples due to their lower depth of coverage (**Supplemental Figure S2**). These data
508 reveal that polymorphic alleles are present in field isolates, including in viral genomes collected
509 from single feather tips.
510
511 To address the potential effect(s) of these polymorphisms on MDV biology, we categorized the
512 observed polymorphisms into categories of synonymous, non-synonymous, genic-untranslated,
513 or intergenic (**Supplemental Table S5**). The majority of all polymorphisms were located in
514 intergenic regions (**Figure 4**). We next investigated whether evidence of selection could be
515 detected from the distribution of polymorphisms in our samples. One way to assess this is to
516 determine whether the relative frequencies of synonymous, non-synonymous, genic-untranslated,
517 and intergenic polymorphisms can be explained by random chance. If the observed frequencies
518 differ from those expected from a random distribution, it would suggest genetic selection. After
519 calculating the expected distribution in each sample (as described in **Methods**), we determined
520 that the distribution of variants differed from that expected by chance in each of our dust samples
521 (**Figure 5**, Farm A-Dust 1: χ^2=68.16, d.f.=3, p<0.001; Farm A-Dust 2: χ^2=128.57, d.f.=3,
522 p<0.001; Farm B-Dust 1: χ^2=63.42, d.f.=3, p<0.001). In addition, we found in pairwise tests that
523 the number of observed intergenic polymorphisms was significantly higher than the observed
524 values for other categories (**Table S6**). This suggests that the mutations that occurred in the
525 intergenic regions were better tolerated and more likely to be maintained in the genome.
526

527 **Tracking shifts in polymorphic loci over time**

528 In addition to observing polymorphic SNPs in each sample at a single moment in time, we
529 explored whether any shifts in polymorphic allele frequency were detected in the two sequential
530 dust samples from Farm A. We found one locus in the ICP4 (MDV084) gene (nucleotide
531 position 5,495) that was polymorphic in the Farm A-dust 2 sample, with nearly equal proportions
532 of sequence reads supporting the major allele (C) and the minor allele (A) (**Figure 6A**). In
533 contrast, this locus had been 99% A and only 1% C in Farm A-dust 1 (collected 11 months
534 earlier), such that it was not counted as polymorphic in that sample by our parameters (see
535 **Methods** for details). At this polymorphic locus, the nucleotide C encodes a serine, while
536 nucleotide A encodes a tyrosine. The encoded AA lies in the C-terminal domain of ICP4
537 (position AA1832). ICP4 is an important immediate-early protein in all herpesviruses, where it
538 serves as a major regulator of viral transcription (65). The role of ICP4 in MDV pathogenesis is
539 also considered crucial because of its proximity to the latency associated transcripts (LAT) and
540 recently described miRNAs (12).
541
542 We examined dust samples from Farm A over a period of 9 months, observing variation at this
543 ICP4 locus using Sanger sequencing (**Figure 6B**). We found that this locus was highly
544 polymorphic in time-separated dust samples. The C (serine) allele was present in different
545 proportions at different time points. In four of the dust samples the major allele switched from C
546 (serine) to A (tyrosine). This reversible fluctuation in allele frequencies over a short period of
547 time is unprecedented for alphaherpesviruses. However, recent studies on HCMV have shown
548 that selection can cause viral populations to evolve in short period of times (33, 34). While this is
549 only one example of a polymorphic locus that shifts in frequency over time, similar approaches

550 could be used at any of the hundreds of polymorphic loci detected here (**Supplemental Table**
551 **S5**). This combination of deep-sequencing genomic approaches to detect new polymorphic loci,
552 and fast gene-specific surveillance to track changes in SNP frequency over a larger number of
553 samples, provides a powerful new approach for field ecology.
554

555 **Comparison of field isolates of MDV-1 to previously sequenced isolates**

556 To compare these new field-based MDV genomes to previously sequenced isolates of MDV, we
557 created a multiple sequence alignment of all available MDV-1 genomes (13, 16–20, 53–55). The
558 multiple sequence alignment was used to generate a dendrogram depicting genetic relatedness
559 (see **Methods**). We observed that the five new isolates form a separate group when compared to
560 all previously sequenced isolates (**Figure 7**). This may result from geographic differences as
561 previously seen for HSV-1 and VZV (66–69, 26), or from temporal differences in the time of
562 sample isolations, or from the lack of cell-culture adaptation in these new genomes. In this
563 dendrogram, all of the isolates from USA form a separate group, relative to the European and
564 Chinese MDV-1 genomes. Within the USA cluster, a separate group was formed by the five new
565 isolates. No prior MDV-1 genomes were available from Pennsylvania, so this separation may
566 well result from geographic separation. Further MDV-1 genomes, representing wider geographic
567 diversity and additional uncultured field-derived genomes, will be needed to distinguish which of
568 these has a greater impact on the observed clustering.
569

570 **Assessment of taxonomic diversity in poultry dust and chicken feathers**

571 As noted in **Table 1**, only a fraction of the reads obtained from each sequencing library were
572 specific to MDV-1. We analyzed the remaining sequences to gain insight into the taxonomic
573 diversity found in poultry dust and chicken feathers. Since our enrichment for viral capsids
574 removed most host and environmental contaminants, the taxa observed here represent only a
575 fraction of the material present initially. However it provides useful insight into the overall
576 complexity of each sample type. The results of the classification for Farm B- dust, Farm B-
577 feather 1, and Farm B-feather 2 are shown in **Supplemental Figure 3**. We divided the sequence
578 reads by the different kingdoms they represent. Complete lists of taxonomic diversity for all
579 samples to the family level are listed in **Supplemental Table S7**. As expected, the taxonomic
580 diversity of poultry dust is greater than that of feather samples. The majority of sequences in the
581 dust samples mapped to the chicken genome, and only about 2-5% were MDV specific (see also
582 **Table 1**, line 6). We found that single feathers were a better source of MDV DNA, due to their
583 reduced level of taxonomic diversity and higher percentage of MDV-specific reads (**Table 1**, line
584 6 and **Supplemental Figure 3**).
585

586 ## Conclusions

587 This study presents the first-ever description of MDV-1 genomes sequenced directly from a field
588 setting. This work builds on recent efforts to sequence VZV and HCMV genomes directly from
589 human clinical samples, but importantly the approaches presented here do not employ either the
590 oligo-enrichment used for VZV or the PCR-amplicon strategy used for HCMV (32, 34, 36, 41,
591 70). This makes our technique accessible to a wider number of labs and reduces the potential

592 methodological bias. We compared these genomes to one another, and to previously sequenced
593 isolates of MDV, at the level of consensus genomes. We further explored the potential for
594 variation within each consensus genome, by searching for minority variants that were hidden
595 beneath the dominant majority allele found in the consensus genome. Our time-series exploration
596 of one such locus in ICP4 demonstrates the power of combining NGS approaches with rapid
597 single-gene analysis of a large number of field samples.
598
599 Together, these data and approaches provide powerful new tools to measure viral diversity in
600 field settings, and to track changes in pathogen populations over time in the field. For instance,
601 these approaches can be used to guide targeted surveillance of one or more sites of variation in
602 the viral genome, with a personalized approach that is tailored to match the variations circulating
603 in a given area. In the case of MDV-1, this could be used to track viral spread across a
604 geographic area, or between multiple end-users associated with a single parent corporation
605 (**Figure 1B**). It could also be used as part of a public-health or animal-health program, for
606 instance to guide management decisions on how to limit pathogen spread or accurately contain
607 airborne pathogens such as MDV-1. For the poultry industry, and in human health, host
608 resistance to a given pathogen is a major area of interest. The ability to sequence and compare
609 large viral genomes directly from individual hosts and field sites will allow a new level of
610 interrogation of host-virus fitness interactions (**Figure 1B**). Finally, analysis of viral genomes
611 from single feather follicles, as from single VZV vesicles, enable our first insights into naturally-
612 occurring within-host variations for these pathogens (**Figure 1B**). For MDV-1, we can now
613 address whether the virus found in the spleen, which has traditionally been used as the source of
614 lab-grown viral stocks and prior MDV-1 genome sequencing, is genetically identical to that shed

615 from the feather follicle epithelium. Evidence from tissue compartmentalization studies in
616 HCMV suggests that viral genomes differ in distinct body niches (34, 36). If the same were true
617 for MDV-1, it is would suggest that prior studies of spleen-derived virus may not fully reflect the
618 virus that is shed from feather-follicle epithelia and transmitted to new hosts.
619
620 Our comparison of new field-isolated MDV-1 genomes revealed a distinct genetic clustering of
621 these genomes, separate from other previously sequenced MDV-1 genomes (**Figure 7**). We
622 cannot yet distinguish whether this pattern results more from geographic and temporal drift in
623 these strains, or from the genetic adaptation(s) to tissue culture, which has been used to
624 propagate the virus for all previously genome sequencing MDV-1 efforts. Prior studies have
625 shown that when MDV is passaged for multiple generations in cell-culture, the virus
626 accumulated a series of mutations, including several that affected virulence (28). The same is
627 true for the betaherpesvirus HCMV (24). Cultured viruses can undergo bottlenecks during initial
628 adaptation to cell culture, and they may accumulate variations by genetic drift or positive
629 selection during the passaging required to grow stocks or prepare DNA for sequencing. We thus
630 anticipate that these field-isolated viral genomes more accurately reflect the genomes of wild
631 MDV-1 that are circulating in the field. The observed genetic distance of the new field MDV
632 genomes is also likely influenced by geographic and temporal separation. The impact of
633 geography on the genetic relatedness of herpesvirus genomes has been previously shown for
634 related alphaherpesviruses such as VZV and HSV-1 (66–69, 26). Phenomena such as
635 recombination can have an impact on clustering pattern of MDV isolates. It is worth noting that
636 the genetic distance dendrogram constructed here included genomes from isolates that were
637 collected over a 40 year span (13, 16–20, 53–55). Agricultural and farming practices have

evolved significantly during this time, and we presume that pathogens have kept pace. To truly understand the global diversity of MDV, future studies will need to include the impacts of recombination and polymorphisms within samples, in addition to the overall consensus-genome differences reflected by static genetic distance analyses.

MDV-1 has proven to be a valuable model for studying virus-host interactions in a natural system. Understanding the genetic variation present in populations of MDV-1 may help us make predictions about the extent of genetic variation in human herpesviruses and lead to the creation of new or improved therapeutics. In combination with statistical models, identification of genetic markers of virulence will aid in tracking the spread of MDV-1 in and between different farms. Epidemiological insights gained from these findings will be of great value in understanding dynamics of disease spread in animal and human populations. Since mutations are the raw material for evolution, understanding the extent and types of viral variations that are present in a field setting will provides valuable insights into the possible future evolution of these pathogens.

Acknowledgements

The authors would also like to thank Michael DeGiorgio, Peter Kerr, Sue Baigent, and members of Szpara and Read labs for helpful feedback and discussion. This work was supported and inspired by the Center for Infectious Disease Dynamics and the Huck Institutes for the Life Sciences at the Pennsylvania State University.

Table 1: Field sample statistics and assembly of MDV-1 consensus genomes

Line #	Category[1]	Farm A - dust 1	Farm A - dust 2	Farm B - dust	Farm B - feather 1	Farm B - feather 2
1	**Nanograms of DNA**	120	127	144	12	27
2	**% MDV-1**	2.4%	1.3%	0.6%	40.6%	5.7%
3	**% MDV-2**	4.6%	2.7%	5.9%	0.1%	0%
4	**Total Reads[2]**	1.4×10^{7}	2.5×10^{7}	2.7×10^{7}	3.9×10^{5}	3.4×10^{5}
5	**MDV-specific reads[2]**	3.7×10^{5}	5.1×10^{5}	1.4×10^{6}	1.0×10^{5}	1.7×10^{5}
6	**% MDV specific reads**	2.6%	2.0%	5.2%	26.9%	48.3%
7	**Average depth (X-fold)**	271	333	597	44	68
8	**Genome length**	177,967	178,049	178,169	178,327	178,540
9	**NCBI accession number**	KU173116	KU173115	KU173119	KU173117	KU173118

[1]Lines 1-3 refer to sample preparation, lines 4-6 to Illumina MiSeq output, and lines 7-9 to new viral genomes.
[2] Sequence read counts in line 4 and 5 are the sum of forward and reverse reads for each sample.

Table 2. Pair-wise DNA identity and variant proteins between pairs of consensus genomes

Comparisons	% DNA identity	Total # bp different	Intergenic		Genic		
			INDELs (# events)	SNPs	INDELs (# events)	Synonymous SNPs	Non-synonymous SNPs
Different farms: Dust vs. dust							
Farm B-dust vs. Farm A-dust 1	99.73	353	143 (22)	140	66 (1) in MDV055	1 in MDV020	3 (one each in MDV010, 012, 056)
Farm B-dust vs. Farm A-dust 2	99.87	195	49 (14)	76	66 (1) in MDV055	1 MDV020	3 (one each in MDV010, 012,056)
Same farm: Dust vs. host							
Farm B-dust vs. Farm B-feather 1	99.64	552	476 (11)	6	66 (1) in MDV055	1 in MDV020	3 (one each in MDV010, 012, 056)
Farm B-dust vs. Farm B-feather 2	99.52	687	572 (19)	45	66 (1) in MDV055	1 in MDV020	3 (one each in MDV010, 012, 056)
Same farm: Separated in time and space							
Farm A-dust 1 vs. Farm A-dust 2	99.76	338	170 (20)	168	0	0	0
Same farm: one host vs. another							
Farm B-feather 1 vs. Farm B-feather 2	99.38	973	972 (9)	1	0	0	0

Supplemental Tables

Supplemental Table S1: Yield and percent MDV1+MDV2 and total nanograms of DNA in each sample for Farm A-dust 1

Samples	Washes on 0.1 µm filter*	%MDV-1	%MDV-2	% MDV1 + MDV2	DNA (ng)
1	0	2.88	5.44	8.3	6.94
2	0	2.03	5.12	7.2	6.59
3	0	4.16	8.39	12.5	6.73
4	0	2.51	4.73	7.2	4.71
5	0	1.66	3.30	4.96	6.97
6	**1**	**9.13**	**13.99**	**23.12**	**2.69**
7	**1**	**9.29**	**15.70**	**24.99**	**2.16**
8	**1**	**5.86**	**10.91**	**16.77**	**3.36**
9	0	1.89	2.98	4.9	9.81
10	0	1.76	2.90	4.7	17.35
11	0	2.69	5.33	8.02	8.96
12	0	4.49	7.80	12.29	4.14
13	0	1.16	2.49	3.65	20.00
14	0	1.36	2.83	4.19	19.47

*Samples that were washed before lysis (bold) yielded a higher percent MDV DNA, but less DNA overall.

Supplemental Table S2: Yield and percent MDV1+MDV2 and total nanograms of DNA in each sample for Farm A-dust 2

Samples	Washes on 0.1 µM filter*	%MDV-1	% MDV-2	%MDV1 + MDV2	DNA (ng)
1	0	1.50	3.16	4.66	10.69
2	0	2.55	5.62	8.17	7.18
3	0	1.36	3.68	5.04	7.62
4	0	1.38	2.94	4.32	9.84
5	**1**	**2.71**	**6.19**	**8.90**	**4.11**
6	**1**	**3.08**	**5.87**	**8.95**	**4.37**
7	**1**	**2.68**	**4.91**	**7.59**	**5.88**
8	**1**	**3.49**	**6.24**	**9.73**	**4.88**
9	**1**	**4.09**	**7.94**	**12.03**	**2.66**
10	**1**	**6.42**	**10.52**	**16.94**	**3.15**
11	0	0.26	0.91	1.17	20.35
12	0	0.19	0.56	0.75	26.09
13	0	0.24	0.93	1.17	15.13
14	0	0.36	1.21	1.57	5.62

*Samples that were washed before lysis (bold) yielded a higher percent MDV DNA, but less DNA overall.

Supplemental Table S3: Yield and percent MDV1+MDV2 and total nanograms of DNA in each sample for Farm B-dust

Samples	Washes on 0.1 µM filter*	% MDV-1	% MDV-2	% MDV-1 +MDV-2	DNA (ng)
1	0	0.84	6.68	7.52	14.10
2	0	0.46	5.20	5.66	26.64
3	0	0.65	4.85	5.50	19.43
4	0	0.75	5.91	6.66	16.84
5	0	0.23	3.67	3.90	25.90
6	0	0.53	4.65	5.18	23.50
7	**1**	**1.10**	**14.50**	**15.60**	**4.59**
8	**1**	**0.95**	**15.77**	**16.72**	**4.29**
9	**1**	**0.95**	**14.40**	**15.35**	**4.81**
10	**1**	**1.02**	**10.69**	**11.71**	**3.59**

*Samples that were washed before lysis (bold) yielded a higher percent MDV DNA, but less DNA overall.

Supplemental Table S4: Yield and percent MDV1+MDV2 and total nanograms of DNA in each sample for Farm B-feathers

Samples	% MDV-1	% MDV-2	% MDV-1 +MDV-2	DNA (ng)
Feather 1	40.59	0.12	40.72	11.97
Feather 2	5.68	0.02	5.70	27.36

Supplemental Table S5: Polymorphic loci detected within all consensus genomes using high and low-stringency variant-filtering criteria.

(see attached Excel spreadsheet)

Supplemental Table S6: Chi-squared values from pairwise comparisons of different categories of polymorphisms.

Sample*	Intergenic vs. synonymous	Intergenic vs. non-synonymous	Intergenic vs. genic untranslated	Synonymous vs. non-synonymous	Synonymous vs. genic untranslated	Non-synonymous vs. genic untranslated
Farm A-dust 1	χ^2=16.6 (p = <0.001)	χ^2=55.47 (p = <0.001)	χ^2=3.74 (p = 0.053)	χ^2=0.03 (p=0.873)	χ^2=0.83 (p = 0.361)	χ^2=1.73 (p = 0.189)
Farm A-dust 2	χ^2=31.76 (p = <0.001)	χ^2=94.93 (p = <0.001)	χ^2=9.48 (p = 0.002)	χ^2=1.11 (p = 0.292)	χ^2=2.72 (p = 0.099)	χ^2=0.69 (p = 0.407)
Farm B-dust	χ^2=25.27 (p = <0.001)	χ^2=47.32 (p = <0.001)	χ^2=5.39 (p = 0.020)	χ^2=1.83 (p = 0.176)	χ^2=1.61 (p = 0.205)	χ^2=0.09 (p = 0.759)

*Degrees of freedom (d.f.) = 1 for all comparisons; p indicates p-value.

Supplemental Table S7: Taxonomic diversity present in all five samples.

(see attached Excel spreadsheet)

Figures

Figure 1. Diagram of samples collected for genome sequencing of field isolates of MDV.

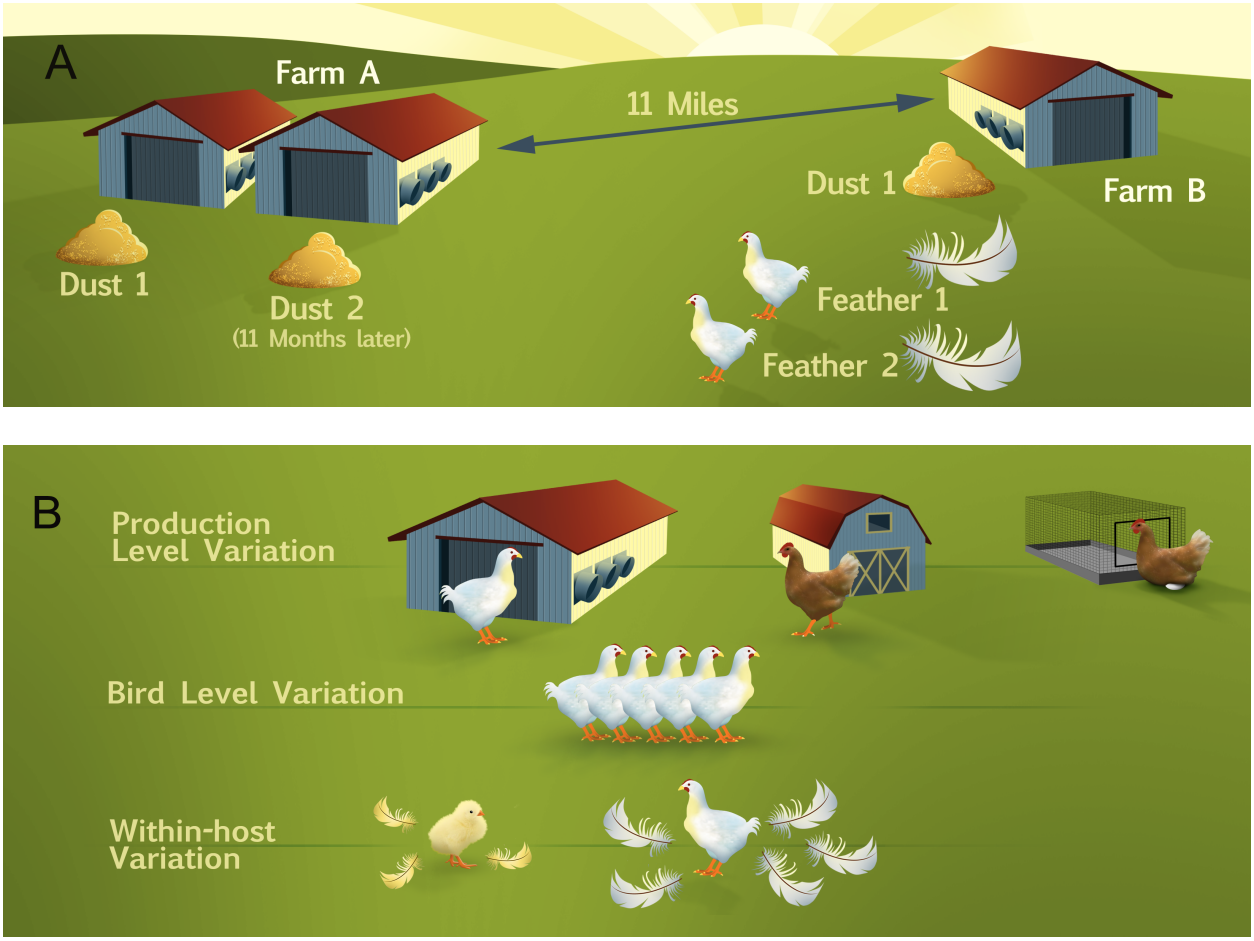


Figure 1. Diagram of samples collected for genome sequencing of field isolates of MDV.

(A) Samples collected for genome sequencing were sourced from two Pennsylvania farms with large-scale operations that house approximately 25,000-30,000 individuals per building. These farms were separated by 11 miles. On Farm A, two separate collections of poultry dust were made 11 months apart. On Farm B we collected one dust sample, along with individual feathers from two chickens. In total, three dust collections and two feathers were used to generate five consensus genomes of MDV field isolates (**Tables 1-3**). (**B**) Future comparisons using these approaches could include an analysis of viral genome variation between scales or types of production (top row), between different animals within any field site (middle row), or between different body locations or temporal stages of a single host (bottom row). (Artwork by Nick Sloff, Penn State University, Department of Entomology).

Figure 2. Procedures for enrichment and isolation of MDV DNA from (A) poultry dust or (B) individual feather follicles.

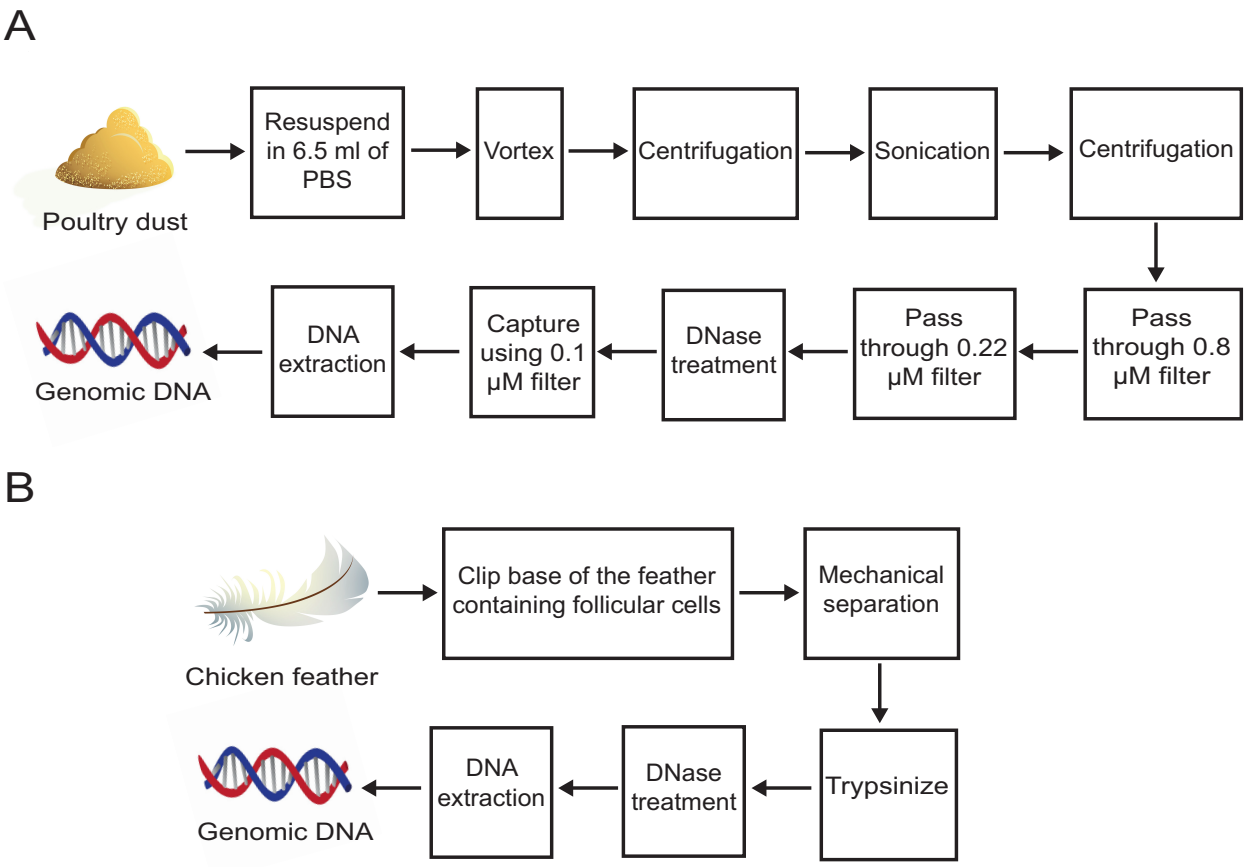


Figure 2. Procedures for enrichment and isolation of MDV DNA from poultry dust or individual feather follicles. (A) Procedure for enrichment of MDV DNA using poultry dust as the source of viral DNA. Vortexing, centrifugation and sonication were essential to release cell-associated virus into the solution. The virus-containing supernatant was then passed through 0.8 µM and 0.22 µM filters for removal of larger contaminants. The flow-thorough was treated with DNase and the viral particles were captured using 0.1 µM filter. The membrane of the 0.1 µM filter was then excised and used for extraction of the viral DNA. (**B**) Procedure for enrichment of MDV DNA using chicken feather follicle as the source of viral DNA. A feather was

mechanically disrupted (bead-beating) and treated with trypsin to break open host cells and release cell-free virus into the solution. The sample was then treated with DNase to remove contaminant DNA. Finally, the viral capsids were lysed to obtain viral genomic DNA.

Figure 3. Complete and trimmed MDV-1 genome

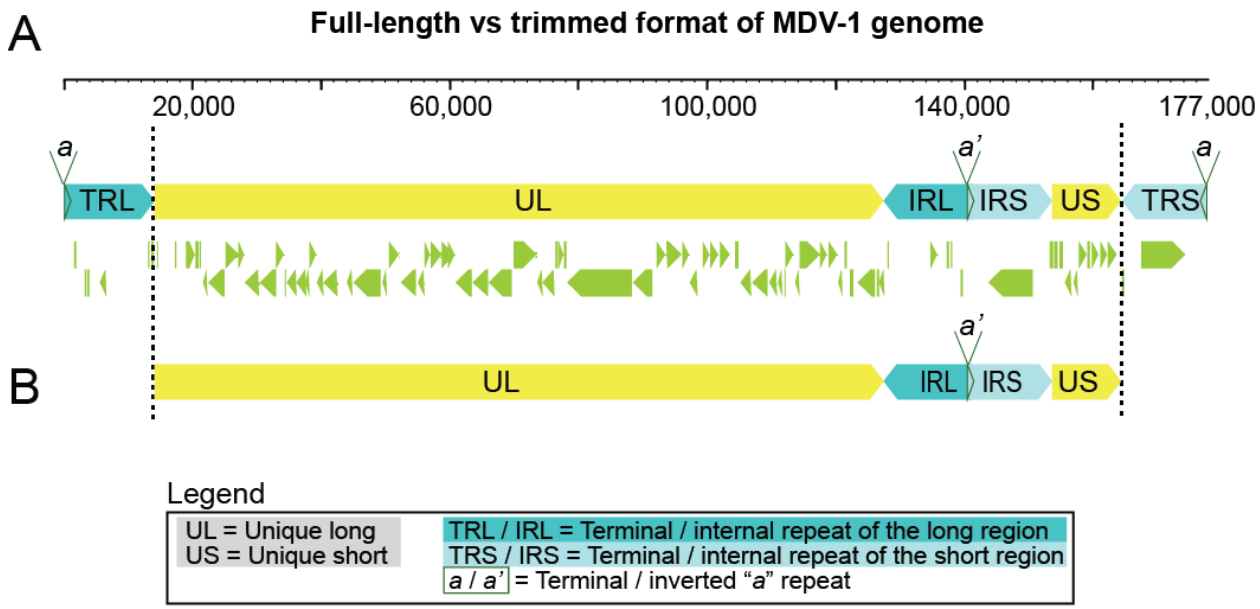


Figure 3. The complete MDV-1 genome includes two unique regions and two sets of large inverted repeats. (A) The full structure of the MDV-1 genome includes a unique long region (UL) and a unique short regions (US), each of which are flanked large repeats known as the terminal and internal repeats of the long region (TRL and IRL) and the short region (TRS and IRS). Most of the ORFs (green arrows) lie in the unique regions of the genome. **(B)** A trimmed genome format without the terminal repeat regions was used for analyses in order to not over-represent the repeat regions.

Figure 4: Genome-wide distribution of polymorphic bases within each consensus genome

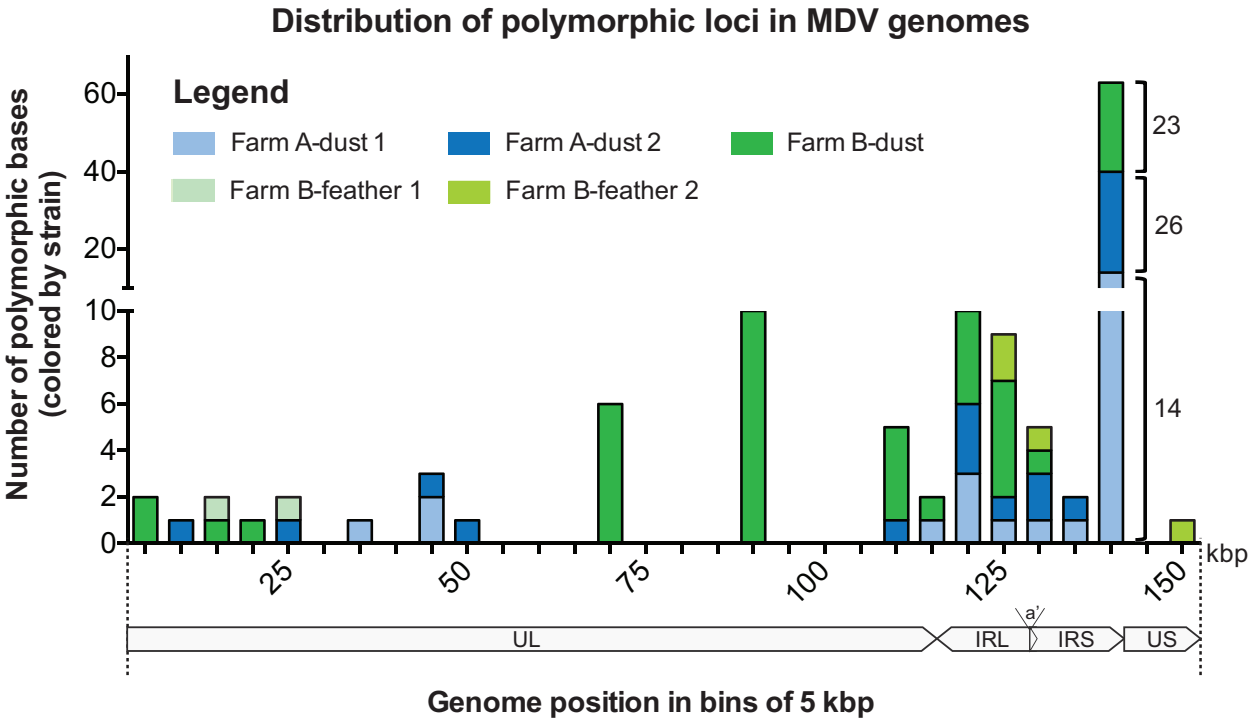


Figure 4: Genome-wide distribution of polymorphic bases within each consensus genome. Polymorphic base calls from each MDV genome were grouped in bins of 5 kb and the sum of polymorphisms in each bin was plotted. Farm B-dust (dark green bars) contained the largest number of polymorphic bases, with the majority occurring in the repeat region (IRL/IRS). Farm A-dust 1 (light blue) and Farm A-dust 2 (dark blue) harbored fewer polymorphic bases, with similar distribution to Farm B-dust. Polymorphic bases detected in feather genomes were more rare, although this likely reflects their lower coverage depth (see **Table 1**). Note that the upper and lower segments of the y-axis have different scales; the number of polymorphic bases per genome for the split column on the right are labeled for clarity.

Figure 5. Observed vs. expected polymorphism categories for each consensus genome

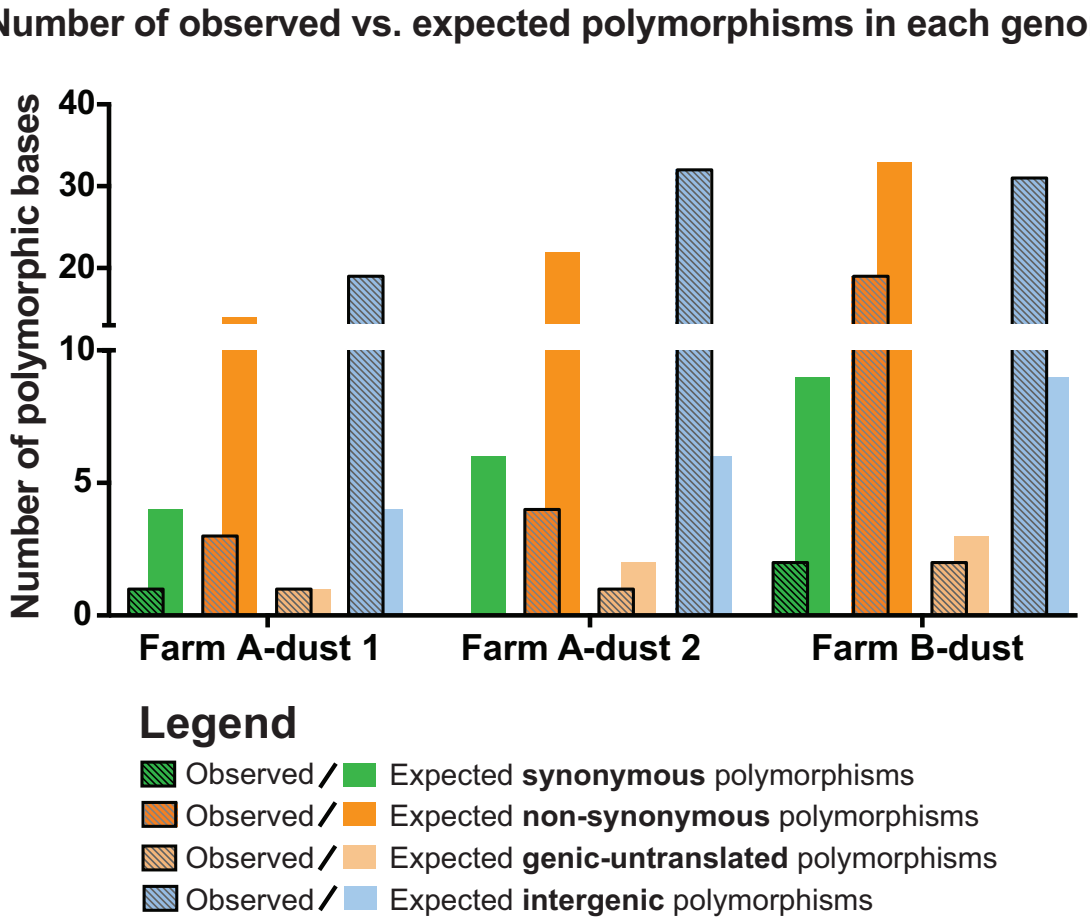


Figure 5. Observed vs. expected polymorphism categories for each consensus genome. Each consensus genome was analyzed for the presence of polymorphic loci (see Methods for details). Observed polymorphic loci (striped bars) were categorized as causing synonymous (green), or non-synonymous (dark orange) mutations, or as genic-untranslated (light orange), or intergenic (blue). The expected outcomes (open colored bars) for a random distribution of polymorphisms is plotted behind the observed outcomes (striped bars) for each category. For all genomes there was a significant difference of the observed-vs.-expected intergenic polymorphisms, relative to those of other categories.

Figure 6: A new polymorphic locus in ICP4, and its shifting allele frequency over time

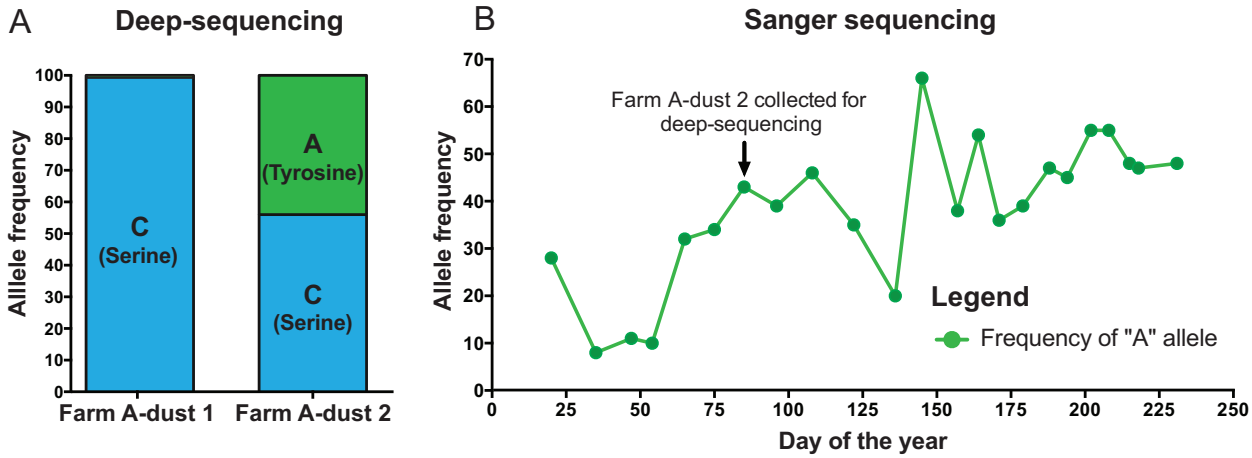


Figure 6: A new polymorphic locus in ICP4, and its shifting allele frequency over time. (A) NGS data revealed a new polymorphic locus in ICP4 (MDV084) at nucleotide position 5,495. In the spatially- and temporally-separated dust samples from Farm A (see **Figure 1A** and **Methods** for details), we observed a different prevalence of C (encoding serine) and A (encoding tyrosine) alleles. **(B)** Using targeted Sanger sequencing of this locus, time-separated dust samples spanning nine months were Sanger-sequenced to track polymorphism frequency at this locus over time. The major and minor allele frequencies at this locus varied widely across time, and the major allele switched from C to A more than twice during this time.

Figure 7: Dendrogram of genetic distances among all sequenced MDV-1 genomes

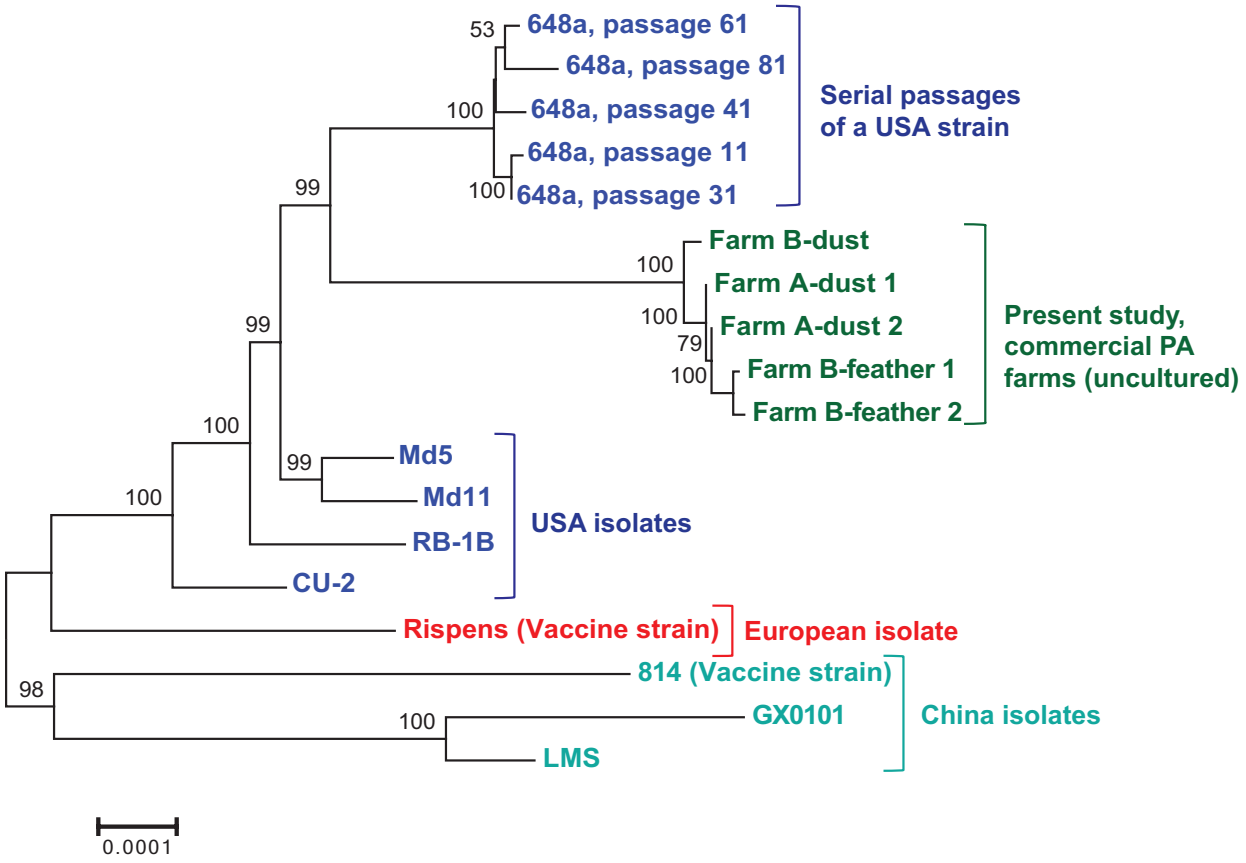
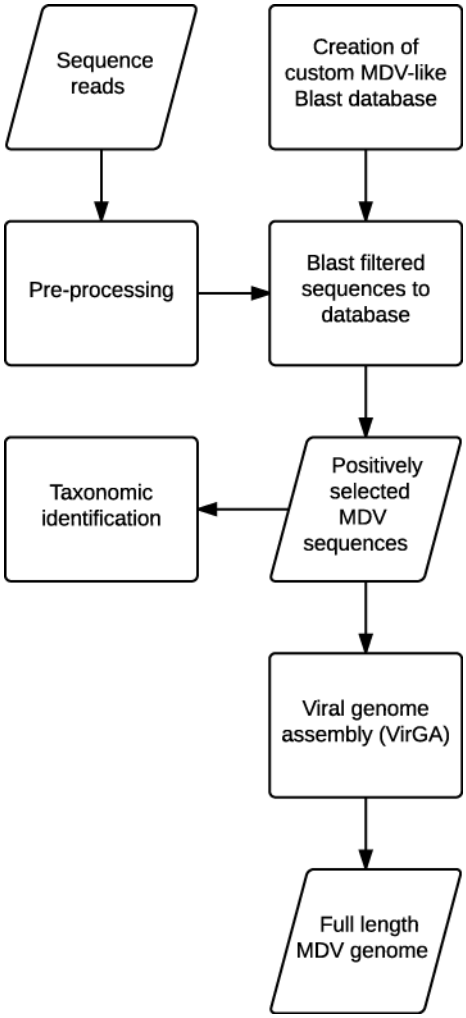


Figure 7: Dendrogram of genetic distances among all sequenced MDV-1 genomes. Using a multiple-genome alignment of all available complete MDV-1 genomes, we calculated the evolutionary distances between genomes using the Jukes-Cantor model. A dendrogram was then created using the neighbor-joining method in MEGA with 1000 bootstraps. The five new field-sampled MDV-1 genomes (green) formed a separate group between the two clusters of USA isolates (blue). The European vaccine strain (Rispens) formed a separate clade, as did the three Chinese MDV-1 genomes (aqua). GenBank Accessions for all strains: new genomes, Table 1; Passage 11-648a, JQ806361; Passage 31-648a, JQ806362; Passage 61-648a, JQ809692; Passage 41-648a, JQ809691; Passage 81-648a, JQ820250; CU-2, EU499381; RB-1B, EF523390; Md11, 170950; Md5, AF243438; Rispens (CVI988), DQ530348; 814, JF742597; GX0101, JX844666; LMS, JQ314003.

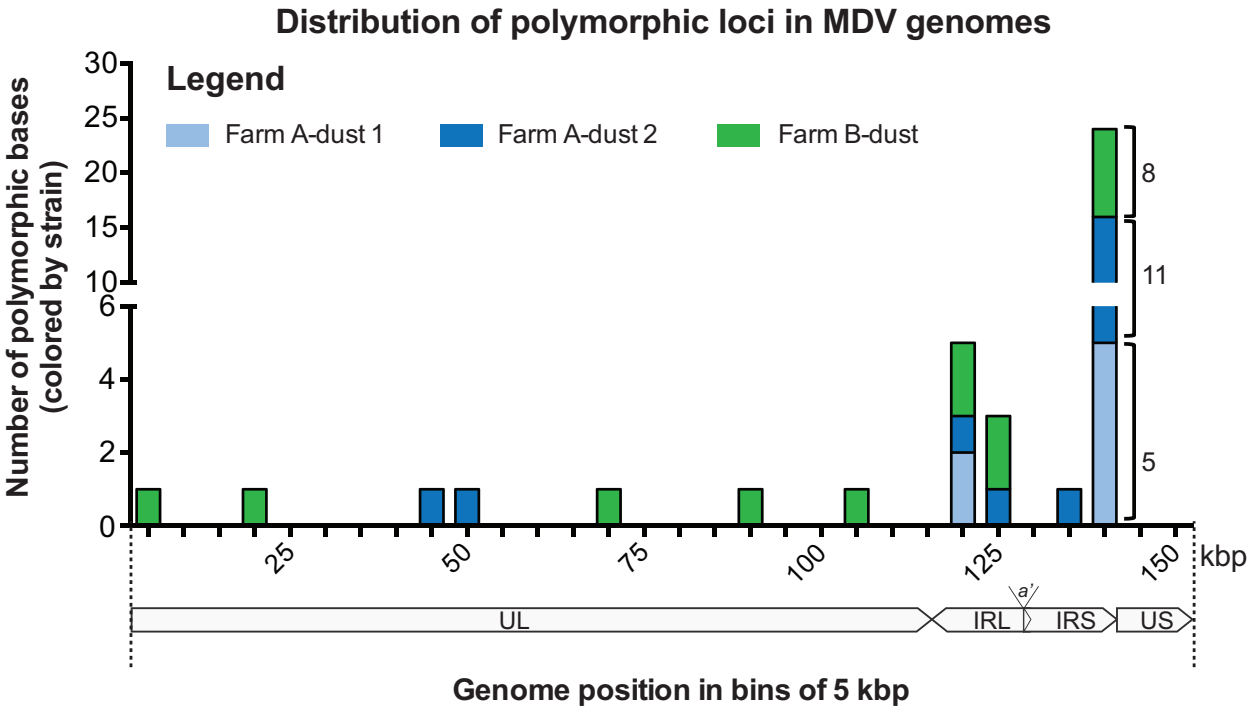
Supplemental Figures

Supplemental Figure S1. Workflow for computational enrichment for MDV sequences and subsequent viral genome assembly and taxonomic profiling



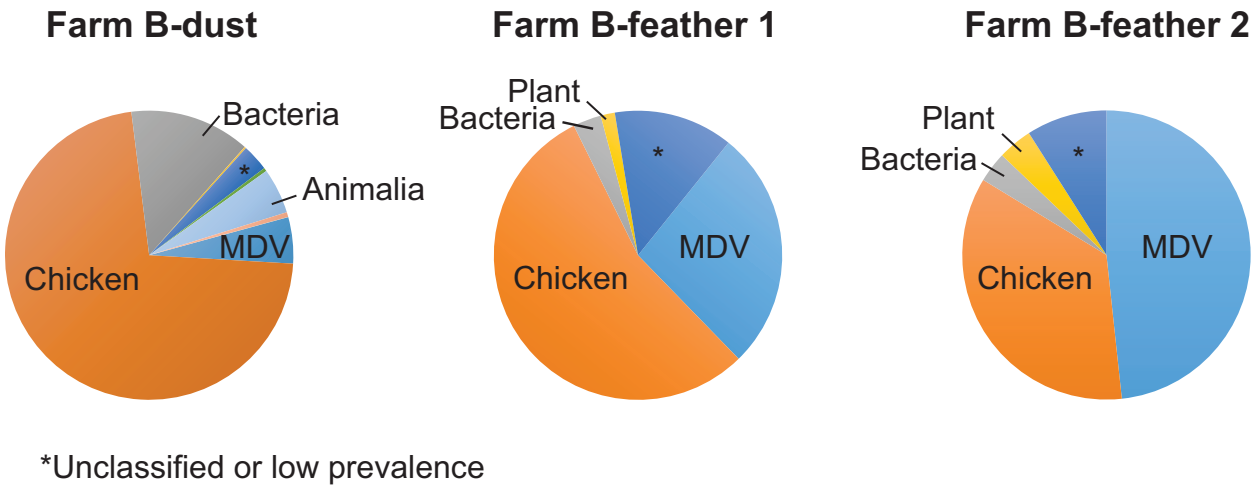
Supplemental Figure S1. Workflow for computational enrichment for MDV sequences and subsequent viral genome assembly and taxonomic profiling. The VirGA workflow (44) requires an input of high-quality NGS data from the viral genome of interest. For this study we added an additional step that selected MDV-like sequence reads from the milieu of dust and feather samples. The sequence reads of interest were obtained by using BLAST to compare all reads against a custom MDV database with an E-value of 10^{-2}; these were then submitted to VirGA for assembly. Taxonomic profiling followed a similar path using NCBI's all-nucleotide database to identify the taxonomic kingdom for each sequence read. In this workflow diagram, parallelograms represent data outputs while rectangles represent computational actions.

Supplemental Figure S2. Genome-wide distribution of polymorphisms within each consensus genome, using high-stringency criteria



Supplemental Figure S2: Genome-wide distribution of polymorphisms within each consensus genome, using high-stringency criteria. Polymorphic base calls from each MDV genome were grouped by position in bins of 5 kb and the sum of polymorphisms in each bin was plotted. Stricter parameters of polymorphism detection (see **Methods** for details) revealed a similar distribution to those in **Figure 4**. No polymorphisms were detected in feather-derived genomes using high-stringency criteria, due to their lower coverage depth (see **Table** 1). Note that the upper and lower segments of the y-axis have different scales; the number of polymorphic bases per segment for the split column on the right are thus labeled on the graph.

Supplemental Figure S3. Taxonomic diversity in poultry dust and chicken feathers from Farm B



Supplemental Figure S3. Taxonomic diversity in poultry dust and chicken feathers from Farm B. We used an iterative BLASTN workflow to generate taxonomic profiles for all samples from Farm B (see **Methods** for details). Major categories are shown here, with a full list of taxa (to family level) in **Supplemental Table S7**. Farm B-feather 1 and Farm B-feather 2 show less overall diversity, as would be expected from direct host-sampling, vs. the environmental mixture of the dust samples. Since the viral DNA enrichment procedures remove variable amounts of host and environmental contaminants, the proportion of taxa present is representative but not fully descriptive of those present initially.

References

1. **Biggs PM**. 2004. Marek's disease: Long and difficult beginnings, p. 8–16. *In* Davison, F, Nair, V (eds.), Marek's Disease. Academic Press, Oxford.

2. **Baigent SJ**, **Davison F**. 2004. 6 - Marek's disease virus: Biology and life cycle, p. 62–ii. *In* Davison, F, Nair, V (eds.), Marek's Disease. Academic Press, Oxford.

3. **Gimeno IM**. 2008. Marek's disease vaccines: a solution for today but a worry for tomorrow? Vaccine **26 Suppl 3**:C31–41.

4. **Spatz SJ**, **Schat KA**. 2011. Comparative genomic sequence analysis of the Marek's disease vaccine strain SB-1. Virus Genes **42**:331–338.

5. **Atkins KE**, **Read AF**, **Walkden-Brown SW**, **Savill NJ**, **Woolhouse MEJ**. 2013. The effectiveness of mass vaccination on Marek's disease virus (MDV) outbreaks and detection within a broiler barn: a modeling study. Epidemics **5**:208–217.

6. **Beasley JN**, **Patterson LT**, **McWade DH**. 1970. Transmission of Marek's disease by poultry house dust and chicken dander. Am J Vet Res **31**:339–344.

7. **Couteaudier M**, **Denesvre C**. 2014. Marek's disease virus and skin interactions. Vet Res **45**:36.

8. **Witter RL**, **Lee LF**. 1984. Polyvalent Marek's disease vaccines: Safety, efficacy and protective synergism in chickens with maternal antibodies. Avian Pathol **13**:75–92.

9. **Atkins KE**, **Read AF**, **Savill NJ**, **Renz KG**, **Islam AFMF**, **Walkden-Brown SW**, **Woolhouse MEJ**. 2013. Vaccination and reduced cohort duration can drive virulence evolution: Marek's disease virus and industrialized agriculture. Evol Int J Org Evol **67**:851–860.

10. **Read AF**, **Baigent SJ**, **Powers C**, **Kgosana LB**, **Blackwell L**, **Smith LP**, **Kennedy DA**, **Walkden-Brown SW**, **Nair VK**. 2015. Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens. PLoS Biol **13**.

11. **Witter RL**. 1997. Increased virulence of Marek's disease virus field isolates. Avian Dis **41**:149–163.

12. **Nair V**. 2013. Latency and Tumorigenesis in Marek's Disease. Avian Dis **57**:360–365.

13. **Tulman ER**, **Afonso CL**, **Lu Z**, **Zsak L**, **Rock DL**, **Kutish GF**. 2000. The Genome of a Very Virulent Marek's Disease Virus. J Virol **74**:7980–7988.

14. **Niikura M**, **Dodgson J**, **Cheng H**. 2006. Direct evidence of host genome acquisition by the alphaherpesvirus Marek's disease virus. Arch Virol **151**:537–549.

15. **Spatz SJ**, **Silva RF**. 2007. Sequence determination of variable regions within the genomes of gallid herpesvirus-2 pathotypes. Arch Virol **152**:1665–1678.

16. **Spatz SJ**, **Zhao Y**, **Petherbridge L**, **Smith LP**, **Baigent SJ**, **Nair V**. 2007. Comparative sequence analysis of a highly oncogenic but horizontal spread-defective clone of Marek's disease virus. Virus Genes **35**:753–766.

17. **Zhang F**, **Liu C-J**, **Zhang Y-P**, **Li Z-J**, **Liu A-L**, **Yan F-H**, **Cong F**, **Cheng Y**. 2011. Comparative full-length sequence analysis of Marek's disease virus vaccine strain 814. Arch Virol **157**:177–183.

18. **Cheng Y**, **Cong F**, **Zhang Y**, **Li Z**, **Xu N**, **Hou G**, **Liu C**. 2012. Genome sequence determination and analysis of a Chinese virulent strain, LMS, of Gallid herpesvirus type 2. Virus Genes **45**:56–62.

19. **Spatz SJ**, **Volkening JD**, **Gimeno IM**, **Heidari M**, **Witter RL**. 2012. Dynamic equilibrium of Marek's disease genomes during in vitro serial passage. Virus Genes **45**:526–536.

20. **Su S**, **Cui N**, **Cui Z**, **Zhao P**, **Li Y**, **Ding J**, **Dong X**. 2012. Complete Genome Sequence of a Recombinant Marek's Disease Virus Field Strain with One Reticuloendotheliosis Virus Long Terminal Repeat Insert. J Virol **86**:13818–13819.

21. **Peters G a**, **Tyler SD**, **Grose C**, **Severini A**, **Gray MJ**, **Upton C**, **Tipples G a**. 2006. A full-genome phylogenetic analysis of varicella-zoster virus reveals a novel origin of replication-based genotyping scheme and evidence of recombination between major circulating clades. J Virol **80**:9850–9860.

22. **Tyler SD**, **Peters GA**, **Grose C**, **Severini A**, **Gray MJ**, **Upton C**, **Tipples GA**. 2007. Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. Virology **359**:447–58.

23. **Bradley AJ**, **Lurain NS**, **Ghazal P**, **Trivedi U**, **Cunningham C**, **Baluchova K**, **Gatherer D**, **Wilkinson GWG**, **Dargan DJ**, **Davison AJ**. 2009. High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. J Gen Virol **90**:2375–80.

24. **Dargan DJ**, **Douglas E**, **Cunningham C**, **Jamieson F**, **Stanton RJ**, **Baluchova K**, **McSharry BP**, **Tomasec P**, **Emery VC**, **Percivalle E**, **Sarasini A**, **Gerna G**, **Wilkinson GWG**, **Davison AJ**. 2010. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. J Gen Virol **91**:1535–46.

25. **Szpara ML**, **Tafuri YR**, **Parsons L**, **Shamim SR**, **Verstrepen KJ**, **Legendre M**, **Enquist LW**. 2011. A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. PLoS Pathog **7**:1–23.

26. **Szpara ML**, **Gatherer D**, **Ochoa A**, **Greenbaum B**, **Dolan A**, **Bowden RJ**, **Enquist LW**, **Legendre M**, **Davison AJ**. 2014. Evolution and diversity in human herpes simplex virus genomes. J Virol **88**:1209–27.

27. **Newman RM**, **Lamers SL**, **Weiner B**, **Ray SC**, **Colgrove RC**, **Diaz F**, **Jing L**, **Wang K**, **Saif S**, **Young S**, **Henn M**, **Laeyendecker O**, **Tobian AAR**, **Cohen JI**, **Koelle DM**, **Quinn TC**, **Knipe DM**. 2015. Genome Sequencing and Analysis of Geographically Diverse Clinical Isolates of Herpes Simplex Virus 2. J Virol JVI.01303–15.

28. **Spatz SJ**. 2010. Accumulation of attenuating mutations in varying proportions within a high passage very virulent plus strain of Gallid herpesvirus type 2. Virus Res **149**:135–142.

29. **Hildebrandt E**, **Dunn JR**, **Perumbakkam S**, **Niikura M**, **Cheng HH**. 2014. Characterizing the molecular basis of attenuation of Marek's disease virus via in vitro serial passage identifies de novo mutations in the helicase-primase subunit gene UL5 and other candidates associated with reduced virulence. J Virol **88**:6232–6242.

30. **Dix RD**, **McKendall RR**, **Baringer JR**. 1983. Comparative neurovirulence of herpes simplex virus type 1 strains after peripheral or intracerebral inoculation of BALB/c mice. Infect Immun **40**:103–112.

31. **Cunningham C**, **Gatherer D**, **Hilfrich B**, **Baluchova K**, **Derrick J**, **Thomson M**, **Griffiths PD**, **Wilkinson GWG**, **Schulz TF**, **Dargan DJ**, **Davison AJ**. 2010. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. J Gen Virol **91**:605–15.

32. **Depledge DP**, **Palser AL**, **Watson SJ**, **Lai IY-C**, **Gray ER**, **Grant P**, **Kanda RK**, **Leproust E**, **Kellam P**, **Breuer J**. 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. PloS One **6**:1–7.

33. **Renzette N**, **Bhattacharjee B**, **Jensen JD**, **Gibson L**, **Kowalik TF**. 2011. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog **7**:e1001344.

34. **Renzette N**, **Gibson L**, **Bhattacharjee B**, **Fisher D**, **Schleiss MR**, **Jensen JD**, **Kowalik TF**. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. PLoS Genet **9**:e1003735.

35. **Lei H**, **Li T**, **Hung G-C**, **Li B**, **Tsai S**, **Lo S-C**. 2013. Identification and characterization of EBV genomes in spontaneously immortalized human peripheral blood B lymphocytes by NGS technology. BMC Genomics **14**:804.

36. **Renzette N**, **Pokalyuk C**, **Gibson L**, **Bhattacharjee B**, **Schleiss MR**, **Hamprecht K**, **Yamamoto AY**, **Mussi-Pinhata MM**, **Britt WJ**, **Jensen JD**, **Kowalik TF**. 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci 201501880.

37. **Renzette N**, **Gibson L**, **Jensen JD**, **Kowalik TF**. 2014. Human cytomegalovirus intrahost evolution—a new avenue for understanding and controlling herpesvirus infections. Curr Opin Virol **8**:109–115.

38. **Sijmons S**, **Van Ranst M**, **Maes P**. 2014. Genomic and Functional Characteristics of Human Cytomegalovirus Revealed by Next-Generation Sequencing. Viruses **6**:1049–1072.

39. **Holland J**, **Spindler K**, **Horodyski F**, **Grabau E**, **Nichol S**, **VandePol S**. 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.

40. **Domingo E**, **Martín V**, **Perales C**, **Grande-Pérez A**, **García-Arriaza J**, **Arias A**. 2006. Viruses as Quasispecies: Biological Implications, p. 51–82. *In* Domingo, E (ed.), Quasispecies: Concept and Implications for Virology. Springer Berlin Heidelberg.

41. **Depledge DP**, **Kundu S**, **Jensen NJ**, **Gray ER**, **Jones M**, **Steinberg S**, **Gershon A**, **Kinchington PR**, **Schmid DS**, **Balloux F**, **Nichols RA**, **Breuer J**. 2014. Deep Sequencing of Viral Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans. Mol Biol Evol **31**:397–409.

42. **Islam A**, **Harrison B**, **Cheetham BF**, **Mahony TJ**, **Young PL**, **Walkden-Brown SW**. 2004. Differential amplification and quantitation of Marek's disease viruses using real-time polymerase chain reaction. J Virol Methods **119**:103–113.

43. **Altschul SF**, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ**. 1990. Basic local alignment search tool. J Mol Biol **215**:403–410.

44. **Parsons LR**, **Tafuri YR**, **Shreve JT**, **Bowen CD**, **Shipley MM**, **Enquist LW**, **Szpara ML**. 2015. Rapid genome assembly and comparison decode intrastrain variation in human alphaherpesviruses. mBio **6**.

45. **Kuhn JH**, **Bao Y**, **Bavari S**, **Becker S**, **Bradfute S**, **Brister JR**, **Bukreyev AA**, **Chandran K**, **Davey RA**, **Dolnik O**, **Dye JM**, **Enterlein S**, **Hensley LE**, **Honko AN**, **Jahrling PB**, **Johnson KM**, **Kobinger G**, **Leroy EM**, **Lever MS**, **Mühlberger E**, **Netesov SV**, **Olinger GG**, **Palacios G**, **Patterson JL**, **Paweska JT**, **Pitt L**, **Radoshitzky SR**, **Saphire EO**, **Smither SJ**, **Swanepoel R**, **Towner JS**, **van der Groen G**, **Volchkov VE**, **Wahl-Jensen V**, **Warren TK**, **Weidmann M**, **Nichol ST**. 2013. Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. Arch Virol **158**:301–311.

46. **Szpara ML**, **Parsons L**, **Enquist LW**. 2010. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. J Virol **84**:5303–13.

47. **Watson G**, **Xu W**, **Reed A**, **Babra B**, **Putman T**, **Wick E**, **Wechsler SL**, **Rohrmann GF**, **Jin L**. 2012. Sequence and comparative analysis of the genome of HSV-1 strain McKrae. Virology **433**:528–37.

48. **Koboldt DC**, **Zhang Q**, **Larson DE**, **Shen D**, **McLellan MD**, **Lin L**, **Miller CA**, **Mardis ER**, **Ding L**, **Wilson RK**. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res **22**:568–576.

49. **Nakamura K**, **Oshima T**, **Morimoto T**, **Ikeda S**, **Yoshikawa H**, **Shiwa Y**, **Ishikawa S**, **Linak MC**, **Hirai A**, **Takahashi H**, **Altaf-Ul-Amin M**, **Ogasawara N**, **Kanaya S**. 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res **39**:1–13.

50. **Cingolani P**, **Platts A**, **Wang LL**, **Coon M**, **Nguyen T**, **Wang L**, **Land SJ**, **Lu X**, **Ruden DM**. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) **6**:80–92.

51. **Cingolani P**, **Patel VM**, **Coon M**, **Nguyen T**, **Land SJ**, **Ruden DM**, **Lu X**. 2012. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet **3**.

52. **Roy S**, **Schreiber E**. 2014. Detecting and Quantifying Low Level Gene Variants in Sanger Sequencing Traces Using the ab1 Peak Reporter Tool. J Biomol Tech JBT **25**:S13–S14.

53. **Spatz SJ**, **Rue CA**. 2008. Sequence determination of a mildly virulent strain (CU-2) of Gallid herpesvirus type 2 using 454 pyrosequencing. Virus Genes **36**:479–489.

54. **Niikura M**, **Dodgson J**, **Cheng H**. 2005. Direct evidence of host genome acquisition by the alphaherpesvirus Marek's disease virus. Arch Virol **151**:537–549.

55. **Spatz SJ**, **Petherbridge L**, **Zhao Y**, **Nair V**. 2007. Comparative full-length sequence analysis of oncogenic and vaccine (Rispens) strains of Marek's disease virus. J Gen Virol **88**:1080–1096.

56. **Katoh K**, **Misawa K**, **Kuma K**, **Miyata T**. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**:3059–3066.

57. **JUKES TH**, **CANTOR CR**. 1969. CHAPTER 24 - Evolution of Protein Molecules, p. 21–132. *In* MUNRO, HN (ed.), Mammalian Protein Metabolism. Academic Press.

58. **Saitou N**, **Nei M**. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4**:406–425.

59. **Tamura K**, **Stecher G**, **Peterson D**, **Filipski A**, **Kumar S**. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol **30**:2725–2729.

60. **Felsenstein J**. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution **39**:783–791.

61. 2013. Fields virology6th ed. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia, PA.

62. **Pellett PE**, **Roizman B**. 2013. Herpesviridae, p. 1802–1822. *In* Fields Virology, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

63. **Knipe DM**, **Howley P**. 2013. Fields Virology. Lippincott Williams & Wilkins.

64. **Osterrieder K**, **Vautherot J-F**. 2004. 3 - The genome content of Marek's disease-like viruses, p. 17–31. *In* Davison, F, Nair, V (eds.), Marek's Disease. Academic Press, Oxford.

65. **Xie Q**, **Anderson AS**, **Morgan RW**. 1996. Marek's disease virus (MDV) ICP4, pp38, and meq genes are involved in the maintenance of transformation of MDCC-MSB1 MDV-transformed lymphoblastoid cells. J Virol **70**:1125–1131.

66. **Norberg P**, **Tyler S**, **Severini A**, **Whitley R**, **Liljeqvist J-A**, **Bergstrom T**. 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. PloS One **6**:1–8.

67. **Grose C**. 2012. Pangaea and the Out-of-Africa Model of Varicella-Zoster Virus Evolution and Phylogeography. J Virol **86**:9558–9565.

68. **Kolb AW**, **Ané C**, **Brandt CR**. 2013. Using HSV-1 genome phylogenetics to track past human migrations. PloS One **8**:1–9.

69. **Chow VT**, **Tipples GA**, **Grose C**. 2013. Bioinformatics of varicella-zoster virus: single nucleotide polymorphisms define clades and attenuated vaccine genotypes. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis **18**:351–356.

70. **Renzette N**, **Bhattacharjee B**, **Jensen JD**, **Gibson L**, **Kowalik TF**. 2011. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog **7**:1–14.

Farm A/dust 1 (low stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm A/dust 1	30905	G	A	7.61%	53	32	2	5	29%	71%	Synonymous variant	MDV030
Farm A/dust 1	40519	G	A	17.78%	56	18	12	4	75%	25%	Non-synonymous variant	MDV034
Farm A/dust 1	43053	G	A	17.24%	48	24	12	3	80%	20%	Genic_UTR	MDV035
Farm A/dust 1	113439	G	T	7.61%	66	19	6	1	86%	14%	Non-synonymous variant	MDV073
Farm A/dust 1	115376	C	A	8.16%	93	42	6	6	50%	50%	Intergenic	N/A
Farm A/dust 1	115377	C	A	29.20%	51	29	21	12	64%	36%	Intergenic	N/A
Farm A/dust 1	117527	G	T	4.55%	132	57	6	3	67%	33%	Intergenic	N/A
Farm A/dust 1	121803	C	T	26.11%	92	41	39	8	83%	17%	Genic_UTR	MDV076
Farm A/dust 1	126256	T	G	12.96%	43	4	4	3	57%	43%	Intergenic	N/A
Farm A/dust 1	132086	T	C	11.97%	108	17	14	3	82%	18%	Non-synonymous variant	MDV084
Farm A/dust 1	137099	A	C	34.74%	42	20	18	15	55%	45%	Intergenic	N/A
Farm A/dust 1	137100	A	C	8.41%	58	40	8	1	89%	11%	Intergenic	N/A
Farm A/dust 1	137101	A	C	5.93%	75	36	4	3	57%	43%	Intergenic	N/A
Farm A/dust 1	137264	A	G	5.22%	87	40	5	2	71%	29%	Intergenic	N/A
Farm A/dust 1	138209	C	A	8.27%	88	34	8	3	73%	27%	Intergenic	N/A
Farm A/dust 1	138212	C	G	6.29%	93	41	8	1	89%	11%	Intergenic	N/A
Farm A/dust 1	138213	T	A	7.87%	77	40	9	1	90%	10%	Intergenic	N/A
Farm A/dust 1	138281	A	C	12.82%	47	21	6	4	60%	40%	Intergenic	N/A
Farm A/dust 1	138377	G	C	8.33%	60	50	1	9	10%	90%	Intergenic	N/A
Farm A/dust 1	138379	A	C	9.71%	39	54	1	9	10%	90%	Intergenic	N/A
Farm A/dust 1	138381	A	T	10.78%	44	47	2	9	18%	82%	Intergenic	N/A
Farm A/dust 1	138490	A	T	12.64%	56	20	3	8	27%	73%	Intergenic	N/A
Farm A/dust 1	138492	C	T	10.84%	52	22	2	7	22%	78%	Intergenic	N/A
Farm A/dust 1	138523	A	G	35.48%	31	9	12	10	55%	45%	Intergenic	N/A

Farm A/dust 2 (low stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm A/dust 2	7251	G	A	5.14%	63	103	1	8	11%	88.89%	Intergenic	N/A
Farm A/dust 2	22829	A	G	4.35%	105	71	6	2	75%	25.00%	Non-synonymous variant	MDV025
Farm A/dust 2	40519	G	A	19.75%	51	79	14	18	44%	56.25%	Non-synonymous variant	MDV034
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	27.78%	Non-synonymous variant	MDV040
Farm A/dust 2	109048	G	A	5.57%	186	102	2	15	12%	88.24%	Genic_UTR	MDV072
Farm A/dust 2	115422	C	T	2.47%	178	134	1	7	13%	87.50%	Intergenic	N/A
Farm A/dust 2	115441	A	C	2.20%	129	225	6	2	75%	25.00%	Intergenic	N/A
Farm A/dust 2	116937	G	A	12.67%	86	176	11	27	29%	71.05%	Intergenic	N/A
Farm A/dust 2	121872	T	C	36.76%	52	108	35	58	38%	62.37%	Genic_UTR	MDV076
Farm A/dust 2	126692	C	G	26.35%	108	1	33	6	85%	15.38%	Intergenic	N/A
Farm A/dust 2	126693	C	T	27.21%	105	1	33	7	83%	17.50%	Intergenic	N/A
Farm A/dust 2	130968	G	T	43.48%	77	105	52	88	37%	62.86%	Non-synonymous variant	MDV084
Farm A/dust 2	137156	C	A	5.17%	81	139	4	8	33%	66.67%	Intergenic	N/A
Farm A/dust 2	137320	T	A	9.94%	120	41	3	15	17%	83.33%	Intergenic	N/A
Farm A/dust 2	138433	C	A	6.50%	215	85	13	8	62%	38.10%	Intergenic	N/A
Farm A/dust 2	138434	T	A	4.43%	202	89	7	7	50%	50.00%	Intergenic	N/A
Farm A/dust 2	138436	C	G	7.06%	221	93	18	6	75%	25.00%	Intergenic	N/A
Farm A/dust 2	138437	T	A	8.12%	219	96	20	8	71%	28.57%	Intergenic	N/A
Farm A/dust 2	138451	G	A	2.42%	244	110	4	5	44%	55.56%	Intergenic	N/A
Farm A/dust 2	138505	C	A	28.26%	75	139	58	33	64%	36.26%	Intergenic	N/A
Farm A/dust 2	138506	A	C	8.97%	122	162	15	13	54%	46.43%	Intergenic	N/A
Farm A/dust 2	138593	C	G	6.17%	81	222	10	10	50%	50.00%	Intergenic	N/A
Farm A/dust 2	138594	G	A	12.81%	76	203	13	28	32%	68.29%	Intergenic	N/A
Farm A/dust 2	138595	G	C	5.02%	84	219	7	9	44%	56.25%	Intergenic	N/A
Farm A/dust 2	138596	T	C	5.61%	81	220	10	8	56%	44.44%	Intergenic	N/A
Farm A/dust 2	138599	A	C	5.38%	87	211	5	12	29%	70.59%	Intergenic	N/A
Farm A/dust 2	138600	G	C	2.59%	90	210	1	7	13%	87.50%	Intergenic	N/A
Farm A/dust 2	138600	G	C	2.59%	90	210	1	7	13%	87.50%	Intergenic	N/A
Farm A/dust 2	138601	G	T	2.97%	81	213	6	3	67%	33.33%	Intergenic	N/A
Farm A/dust 2	138601	G	T	2.97%	81	213	6	3	67%	33.33%	Intergenic	N/A
Farm A/dust 2	138602	G	C	4.15%	77	197	3	9	25%	75.00%	Intergenic	N/A
Farm A/dust 2	138602	G	C	4.15%	77	197	3	9	25%	75.00%	Intergenic	N/A
Farm A/dust 2	138604	A	C	4.96%	77	191	5	9	36%	64.29%	Intergenic	N/A
Farm A/dust 2	138604	A	C	4.96%	77	191	5	9	36%	64.29%	Intergenic	N/A
Farm A/dust 2	138606	A	T	4.63%	75	192	3	10	23%	76.92%	Intergenic	N/A
Farm A/dust 2	138606	A	T	4.63%	75	192	3	10	23%	76.92%	Intergenic	N/A
Farm A/dust 2	138748	A	G	17.82%	12	65	5	13	28%	72.22%	Intergenic	N/A
Farm A/dust 2	138748	A	G	17.82%	12	65	5	13	28%	72.22%	Intergenic	N/A

Farm B/dust (low stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm B/dust	2072	T	G	44%	90	43	66	37	64%	36%	Non-synonymous variant	MDV010
Farm B/dust	4411	G	T	44%	18	80	14	64	18%	82%	Non-synonymous variant	MDV012
Farm B/dust	13809	G	C	3%	173	133	4	4	50%	50%	Non-synonymous variant	MDV019
Farm B/dust	15775	C	T	46%	94	53	78	46	63%	37%	Synonymous variant	MDV020
Farm B/dust	65764	G	A	10%	66	93	3	14	18%	82%	Non-synonymous variant	MDV049
Farm B/dust	65773	G	T	16%	62	95	4	26	13%	87%	Non-synonymous variant	MDV049
Farm B/dust	65796	A	G	35%	53	85	9	66	12%	88%	Synonymous variant	MDV049
Farm B/dust	65804	A	T	34%	61	89	11	67	14%	86%	Non-synonymous variant	MDV049
Farm B/dust	65821	G	T	34%	53	83	8	63	11%	89%	Non-synonymous variant	MDV049
Farm B/dust	65843	A	G	11%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049
Farm B/dust	85939	A	C	95%	0	2	15	21	42%	58%	Non-synonymous variant	MDV055
Farm B/dust	85954	C	T	29%	14	18	11	2	85%	15%	Non-synonymous variant	MDV055
Farm B/dust	85959	C	A	36%	22	17	16	6	73%	27%	Non-synonymous variant	MDV055
Farm B/dust	85961	C	A	39%	28	16	20	8	71%	29%	Non-synonymous variant	MDV055
Farm B/dust	85962	G	A	40%	25	10	19	4	83%	17%	Non-synonymous variant	MDV055
Farm B/dust	85963	T	C	56%	7	12	18	6	75%	25%	Non-synonymous variant	MDV055
Farm B/dust	85966	G	A	9%	46	22	3	4	43%	57%	Non-synonymous variant	MDV055
Farm B/dust	85971	G	C	11%	47	18	4	4	50%	50%	Non-synonymous variant	MDV055
Farm B/dust	85974	G	C	9%	49	21	2	5	29%	71%	Non-synonymous variant	MDV055
Farm B/dust	86626	T	C	40%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056
Farm B/dust	108743	T	C	42%	173	95	119	73	62%	38%	Genic_UTR	MDV072
Farm B/dust	108856	G	A	10%	254	164	6	41	13%	87%	Genic_UTR	MDV072
Farm B/dust	108899	A	G	11%	227	192	7	45	13%	87%	Genic_UTR	MDV072
Farm B/dust	109012	T	C	5%	184	176	2	16	11%	89%	Genic_UTR	MDV072
Farm B/dust	114927	T	G	2%	331	276	13	2	87%	13%	Intergenic	N/A
Farm B/dust	115231	A	C	22%	127	221	36	61	37%	63%	Intergenic	N/A
Farm B/dust	115232	A	C	15%	161	270	19	57	25%	75%	Intergenic	N/A
Farm B/dust	115241	C	A	4%	155	360	17	6	74%	26%	Intergenic	N/A
Farm B/dust	116288	T	A	2%	309	248	3	9	25%	75%	Intergenic	N/A
Farm B/dust	120327	A	G	29%	4	64	3	25	11%	89%	Intergenic	N/A
Farm B/dust	121181	A	C	2%	203	194	4	6	40%	60%	Non-synonymous variant	MDV076
Farm B/dust	121656	C	T	37%	76	118	43	72	37%	63%	Genic_UTR	MDV076
Farm B/dust	122052	A	T	3%	381	231	14	3	82%	18%	Genic_UTR	MDV076
Farm B/dust	124841	T	C	42%	188	151	130	113	53%	47%	Intergenic	N/A
Farm B/dust	127347	G	T	42%	35	178	22	130	14%	86%	Intergenic	N/A
Farm B/dust	137081	A	C	4%	97	66	1	6	14%	86%	Intergenic	N/A
Farm B/dust	137101	A	G	2%	137	198	7	1	88%	13%	Intergenic	N/A
Farm B/dust	137102	A	G	4%	119	199	10	2	83%	17%	Intergenic	N/A
Farm B/dust	137249	A	G	3%	112	193	7	1	88%	13%	Intergenic	N/A
Farm B/dust	137449	A	C	45%	149	62	103	70	60%	40%	Intergenic	N/A
Farm B/dust	138195	C	A	4%	402	137	15	5	75%	25%	Intergenic	N/A
Farm B/dust	138196	T	A	3%	383	138	10	7	59%	41%	Intergenic	N/A
Farm B/dust	138198	C	G	5%	405	142	18	8	69%	31%	Intergenic	N/A
Farm B/dust	138199	T	A	6%	393	142	24	8	75%	25%	Intergenic	N/A
Farm B/dust	138266	C	A	4%	266	386	19	9	68%	32%	Intergenic	N/A
Farm B/dust	138267	C	A	37%	145	231	81	142	36%	64%	Intergenic	N/A
Farm B/dust	138268	A	C	8%	188	396	25	24	51%	49%	Intergenic	N/A
Farm B/dust	138355	C	G	5%	83	407	9	18	33%	67%	Intergenic	N/A
Farm B/dust	138356	G	A	13%	74	368	15	51	23%	77%	Intergenic	N/A
Farm B/dust	138357	G	C	4%	83	397	5	14	26%	74%	Intergenic	N/A
Farm B/dust	138358	T	C	4%	82	399	8	12	40%	60%	Intergenic	N/A
Farm B/dust	138361	A	C	5%	89	381	6	20	23%	77%	Intergenic	N/A
Farm B/dust	138362	G	C	3%	95	379	2	11	15%	85%	Intergenic	N/A
Farm B/dust	138364	G	C	4%	80	355	5	15	25%	75%	Intergenic	N/A
Farm B/dust	138366	A	C	5%	79	343	8	14	36%	64%	Intergenic	N/A
Farm B/dust	138368	A	T	3%	79	345	2	13	13%	87%	Intergenic	N/A
Farm B/dust	138476	A	C	4%	111	79	2	5	29%	71%	Intergenic	N/A
Farm B/dust	138510	A	G	28%	25	73	10	28	26%	74%	Intergenic	N/A

Farm B/feather 1 (low stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm B/feather 1	12176	C	A	15.56%	17	20	6	1	86%	14%	Non-synonymous variant	MDV018
Farm B/feather 1	23126	G	T	12.28%	28	21	2	5	29%	71%	Non-synonymous variant	MDV025

Farm B/feather 2 (low stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm B/feather 2	122174	C	T	10.94%	32	25	1	6	14%	85.71%	Genic_UTR	MDV076
Farm B/feather 2	122204	C	T	15.79%	21	27	2	7	22%	77.78%	Genic_UTR	MDV076
Farm B/feather 2	128489	G	T	10.13%	38	33	2	6	25%	75.00%	Intergenic	N/A
Farm B/feather 2	144479	C	A	14.00%	33	10	5	2	71%	28.57%	Non-synonymous variant	MDV092

Farm A/dust 1 (high stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm A/dust 1	115376	C	A	8.16%	93	42	6	6	50%	50.00%	Intergenic	N/A
Farm A/dust 1	115377	C	A	29.20%	51	29	21	12	64%	36.36%	Intergenic	N/A
Farm A/dust 1	137099	A	C	34.74%	42	20	18	15	55%	45.45%	Intergenic	N/A
Farm A/dust 1	137101	A	C	5.93%	75	36	4	3	57%	42.86%	Intergenic	N/A
Farm A/dust 1	137264	A	G	5.22%	87	40	5	2	71%	28.57%	Intergenic	N/A
Farm A/dust 1	138209	C	A	8.27%	88	34	8	3	73%	27.27%	Intergenic	N/A
Farm A/dust 1	138281	A	C	12.82%	47	21	6	4	60%	40.00%	Intergenic	N/A

Farm A/dust 2 (high stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm A/dust 2	40519	G	A	19.88%	51	78	14	18	44%	56%	Non-synonymous variant	MDV034
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	28%	Non-synonymous variant	MDV040
Farm A/dust 2	116937	G	A	12.67%	86	176	11	27	29%	71%	Intergenic	N/A
Farm A/dust 2	121872	T	C	36.76%	52	108	35	58	38%	62%	Genic_UTR	MDV076
Farm A/dust 2	130968	G	T	43.48%	77	105	52	88	37%	63%	Non-synonymous variant	MDV084
Farm A/dust 2	137156	C	A	5.17%	81	139	4	8	33%	67%	Intergenic	N/A
Farm A/dust 2	138433	C	A	6.54%	215	85	13	8	62%	38%	Intergenic	N/A
Farm A/dust 2	138436	C	G	7.12%	221	92	18	6	75%	25%	Intergenic	N/A
Farm A/dust 2	138437	T	A	8.19%	218	96	20	8	71%	29%	Intergenic	N/A
Farm A/dust 2	138505	C	A	28.91%	73	136	52	33	61%	39%	Intergenic	N/A
Farm A/dust 2	138506	A	C	9.15%	117	161	15	13	54%	46%	Intergenic	N/A
Farm A/dust 2	138593	C	G	6.19%	81	222	10	10	50%	50%	Intergenic	N/A
Farm A/dust 2	138594	G	A	12.81%	76	203	13	28	32%	68%	Intergenic	N/A
Farm A/dust 2	138596	T	C	5.35%	81	220	9	8	53%	47%	Intergenic	N/A
Farm A/dust 2	138599	A	C	5.40%	87	211	5	12	29%	71%	Intergenic	N/A
Farm A/dust 2	138748	A	G	19.15%	12	64	5	13	28%	72%	Intergenic	N/A

Farm B/dust (high stringency)

Isolate	Position in the genome	Major allele	Minor allele	Minor allele frequency	Reads supporting major allele on forward strand	Reads supporting major allele on reverse strand	Reads supporting minor allele on forward strand	Reads supporting minor allele on reverse strand	Percent reads supporting minor allele on forward strand	Percent reads supporting minor allele on reverse strand	Type of variation	Gene
Farm B/dust	2072	T	G	43.64%	90	43	66	37	64%	36%	Non-synonymous variant	MDV010
Farm B/dust	15775	C	T	45.76%	94	53	78	46	63%	37%	Synonymous variant	MDV020
Farm B/dust	65843	A	G	11.30%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049
Farm B/dust	86626	T	C	40.19%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056
Farm B/dust	108743	T	C	41.74%	173	95	119	73	62%	38%	Genic_UTR	MDV072
Farm B/dust	115231	A	C	21.80%	127	221	36	61	37%	63%	Intergenic	N/A
Farm B/dust	115232	A	C	14.99%	161	270	19	57	25%	75%	Intergenic	N/A
Farm B/dust	121656	C	T	37.22%	76	118	43	72	37%	63%	Genic_UTR	MDV076
Farm B/dust	124841	T	C	41.75%	188	151	130	113	53%	47%	Intergenic	N/A
Farm B/dust	137449	A	C	45.05%	149	62	103	70	60%	40%	Intergenic	N/A
Farm B/dust	138199	T	A	5.64%	393	142	24	8	75%	25%	Intergenic	N/A
Farm B/dust	138267	C	A	37.23%	145	231	81	142	36%	64%	Intergenic	N/A
Farm B/dust	138268	A	C	7.74%	188	396	25	24	51%	49%	Intergenic	N/A
Farm B/dust	138355	C	G	5.22%	83	407	9	18	33%	67%	Intergenic	N/A
Farm B/dust	138356	G	A	12.99%	74	368	15	51	23%	77%	Intergenic	N/A
Farm B/dust	138361	A	C	5.24%	89	381	6	20	23%	77%	Intergenic	N/A
Farm B/dust	138510	A	G	27.94%	25	73	10	28	26%	74%	Intergenic	N/A

Family Name	Farm B-dust	Farm B-feather 1	Farm B-feather 2	Farm A-dust 1	Farm A-dust 2
Chicken	19,119,306	212,314	121,060	11,783,342	21,215,016
Herpesviridae	1,372,838	103,939	165,216	370,393	512,531
Bovidae	1,131,177	-	-	8,055	39,873
Bradyrhizobiaceae	986,082	-	-	682	99,710
Dermabacteraceae	605,631	-	-	75,916	375,290
<100	570,555	50,009	27,895	794,467	533,986
Staphylococcaceae	337,057	-	-	178,747	240,507
Corynebacteriaceae	229,208	-	-	52,055	129,345
unclassified	196,911	-	-	147,502	109,899
Lactobacillaceae	196,157	-	-	147,351	195,604
Babesiidae	141,039	-	-	3,930	10,966
Bacillaceae	104,791	-	-	77,052	97,930
Vira	102,266	-	-	7,700	31,685
Sphingobacteriaceae	77,220	-	-	36,775	68,077
Bacteroidaceae	72,040	-	-	32,391	64,454
Streptococcaceae	63,088	-	-	23,633	42,208
Actinoplanaceae	55,273	-	-	19,379	52,946
Lachnospiraceae	54,841	-	-	35,767	100,136
Siphoviridae	54,286	-	-	28,164	74,766
Clostridiaceae	53,824	-	-	18,827	43,915
biota	53,379	-	-	20,321	64,013
Peptostreptococcaceae	49,713	-	-	4,585	29,781
Gramineae	48,512	6,145	13,189	17,909	59,709
Meleagrididae	47,292	9,570	6,352	16,493	76,814
Ruminococcaceae	36,850	-	-	22,980	63,335
Methylobacteriaceae	35,567	2,660	5,205	356	58,844
Enterobacteraceae	35,496	-	-	11,304	72,340
Micrococcaceae	30,584	-	-	13,502	28,318
other sequences	27,430	1,678	3,012	10,400	14,568
Nocardiaceae	26,664	-	-	6,869	20,515
Myoviridae	25,744	-	-	26,447	56,563
Actinosynnemataceae	24,644	-	-	7,354	21,047
Enterococcaceae	24,224	-	-	13,055	23,442
Mycobacteriaceae	23,366	-	-	6,151	18,818
Pseudomonadaceae	21,552	-	-	2,478	11,341
Burkholderiaceae	21,347	-	-	-	5,960
Rhizobiaceae	16,069	-	-	103	4,745
Propionibacteriaceae	15,405	-	148	6,197	12,459
Nocardiopsaceae	14,995	-	-	11,804	22,330
Sphingomonadaceae	14,945	-	-	407	15,444
Campylobacter group	13,811	-	-	6,182	5,466
Porphyromonadaceae	13,008	-	-	6,759	25,153
Comamonadaceae	11,483	-	-	623	6,288
Rikenellaceae	11,272	-	-	13,944	61,763
Cellulomonadaceae	10,982	-	-	3,894	9,435
Xanthobacteraceae	10,979	-	-	389	2,367
Rhodospirillaceae	10,498	-	-	499	3,578
Microbacteriaceae	9,978	-	-	3,217	8,655
Cervidae	9,471	-	-	-	-
Lysobacteraceae	9,311	-	-	1,026	5,336
Phyllobacteriaceae	9,070	-	-	217	2,283
Promicromonosporaceae	8,756	-	-	2,975	7,802
Dermacoccaceae	8,713	-	-	2,667	7,232
Bifidobacteriaceae	8,667	-	-	5,241	52,059
Coriobacteriaceae	7,818	-	-	6,884	12,377
Geodermatophilaceae	7,635	-	-	2,153	6,594
Paenibacillaceae	6,921	-	-	4,357	7,627
Rhodobacteraceae	6,764	-	-	331	3,802
Caulobacteraceae	6,531	-	-	685	2,499
Listeriaceae	6,379	-	-	6,555	8,296
Nocardioidaceae	6,376	-	-	2,592	5,286
Gordoniaceae	6,318	-	-	1,779	4,944
Pasteurellaceae	6,071	-	-	2,431	1,853
Eubacteriaceae	5,647	-	-	3,392	8,423
Flavobacteriaceae	5,492	-	-	3,837	4,432
Erysipelothrix group	5,484	-	-	3,451	8,925
Alcaligenaceae	5,223	-	-	561	2,550
Frankiaceae	5,149	-	-	2,071	4,817
Peptococcaceae	4,819	-	-	3,381	6,319

Family Name	Farm B-dust	Farm B-feather 1	Farm B-feather 2	Farm A-dust 1	Farm A-dust 2
Beutenbergiaceae	4,723	-	-	1,786	4,091
Sanguibacteraceae	4,681	-	-	1,619	3,950
Carnobacteriaceae	4,362	-	-	3,804	4,445
Kineosporiaceae	4,278	-	-	1,351	4,031
Catenulisporaceae	4,007	-	-	973	2,611
Megasphaera group	3,758	-	-	2,066	4,526
Rhodocyclaceae	3,754	-	-	484	2,155
Caryophanaceae	3,378	-	-	2,287	3,492
Oscillospiraceae	3,268	-	-	2,227	5,981
Delphinidae	3,118	-	-	-	-
Intrasporangiaceae	3,105	-	-	822	2,230
Verrucomicrobia subdivision 1	2,960	-	-	134	2,274
Aspergillaceae	2,918	-	-	-	-
Thermoanaerobacterales Family III. Incerta	2,695	-	-	1,907	2,999
Leuconostoc group	2,675	-	-	2,382	3,363
Acetobacteraceae	2,675	-	-	-	1,164
Oxalobacteraceae	2,523	-	-	154	1,090
Ancylobacter group	2,446	-	-	-	977
Microsphaeraceae	2,427	-	-	773	2,196
Desulfovibrionaceae	2,370	-	-	605	3,042
Actinomycetaceae	2,328	-	-	1,052	1,947
Tsukamurellaceae	2,252	-	-	585	1,746
Fusobacteriaceae	2,232	-	-	872	1,807
Acinetobacteraceae	2,138	-	-	10,446	4,512
Borrelomycetaceae	2,006	-	-	546	1,011
Cytophaga-Flexibacter group	1,999	-	-	490	1,636
Alcanivorax/Fundibacter group	1,945	-	-	484	1,729
Sapromycetaceae	1,943	-	-	1,103	1,606
Spirochaetaceae	1,866	-	-	1,069	3,796
Muridae	1,847	-	-	1,525	3,953
Ectothiorhodospira group	1,784	-	-	121	1,441
Deinococcaceae	1,768	-	-	733	1,637
Nymphalidae	1,764	-	-	-	112
Glycomycetaceae	1,732	-	-	575	1,515
Thermoanaerobacteraceae	1,659	-	-	850	2,119
Prevotellaceae	1,573	-	-	244	2,229
Aeromonadaceae	1,534	-	-	273	1,471
Anaeromyxobacteraceae	1,496	-	-	326	1,337
Myxococcaceae	1,458	-	-	443	1,417
Microviridae	1,430	-	-	-	758
Geobacteraceae	1,371	-	-	383	1,653
Peptoniphilaceae	1,357	-	-	820	1,317
Chromobacteriaceae	1,305	-	-	398	943
Leptotrichiaceae	1,302	-	-	674	1,496
Sorangiaceae	1,280	-	-	145	943
Chromatiaceae	1,264	-	-	127	963
Rhodobiaceae	1,239	-	-	-	245
Spiroplasmataceae	1,234	-	-	824	1,207
Beijerinckiaceae	1,198	-	-	-	338
Halanaerobiaceae	1,153	-	-	605	936
Suidae	1,106	-	-	294	1,088
Thermaceae	1,089	-	-	220	1,568
Haloarchaeaceae	1,084	-	-	-	215
Sporolactobacillaceae	1,084	-	-	885	1,221
Conexibacteraceae	1,070	-	-	235	832
Brachyspiraceae	1,065	-	-	572	1,003
Acidobacteriaceae	1,001	-	-	-	887
Acidimicrobiaceae	980	-	-	236	478
Cercopithecidae	882	-	-	600	1,146
Hominidae	849	-	-	39,199	3,817
Nolanaceae	847	-	-	4,110	926
Acidaminococcaceae	818	-	-	515	1,206
Thermotogaceae	794	-	-	456	624
Chlorobiacea	774	-	-	-	954
Clostridiales Family XVIII. Incertae Sedis	771	-	-	375	1,055
Dietziaceae	755	-	-	129	338
Methanobacteriaceae	750	-	-	180	114
Methylocystaceae	739	-	-	-	202
Aerococcaceae	716	-	-	531	631

Family Name	Farm B-dust	Farm B-feather 1	Farm B-feather 2	Farm A-dust 1	Farm A-dust 2
Jonesiaceae	675	-	-	377	574
Helicobacteraceae	671	-	-	-	393
Clostridiales Family XVII. Incertae Sedis	664	-	-	313	949
Cyprinidae	650	-	-	108	174
Methanomassiliicoccaceae	632	-	-	435	690
Haliangiaceae	628	-	-	146	435
Clostridiales Family XV. Incertae Sedis	626	-	-	244	812
Eremotheciaceae	622	-	-	146	312
Brucellaceae	614	-	-	105	185
Podoviridae	570	-	-	262	2,117
Giraffidae	569	-	-	-	-
Alteromonadaceae	568	-	-	113	469
Mustelidae	568	-	-	-	-
Rubrobacteraceae	548	-	-	118	556
Leporidae	506	-	-	103	129
Mimiviridae	500	-	-	-	309
Alicyclobacillaceae	498	-	-	122	564
Equidae	494	-	-	-	-
Opitutaceae	492	-	-	109	524
Euphorbiaceae	483	-	-	-	-
Hyphomonadaceae	472	-	-	-	153
Alcanivoracaceae	471	-	-	132	289
Halobacteroidaceae	465	-	-	408	648
Sphaerobacteraceae	458	-	-	-	330
Archangiaceae	458	-	-	149	372
Candidatus Brocadiaceae	457	-	-	249	236
Segniliparaceae	423	-	-	132	332
Schistosomatidae	416	-	-	296	524
Desulfobulbaceae	415	-	-	126	406
Erythrobacteraceae	400	-	-	-	382
Piscirickettsia group	399	-	-	102	248
Pelobacteraceae	396	-	-	249	588
Solibacteraceae	390	-	-	-	286
Dasyuridae	372	-	-	400	358
Heliobacteriaceae	371	-	-	198	502
Acaridae	341	-	-	111	217
Rhodothermaceae	334	-	-	-	722
Iridoviridae	333	-	-	-	-
Cyclobacteriaceae	328	-	-	184	264
Desulfarculaceae	317	-	-	149	440
Trueperaceae	309	-	-	104	200
Dictyoglomaceae	298	-	-	147	166
Phycisphaeraceae	296	-	-	135	370
Desulfobacteraceae	282	-	-	-	343
Vibrionaceae	277	-	-	-	400
Gallionella group	275	-	-	-	311
Hydrogenothermaceae	275	-	-	118	156
Ignavibacteriaceae	268	-	-	-	155
Methylococcaceae	256	-	-	-	296
Shewanellaceae	252	-	-	102	161
Nostocaceae	240	-	-	-	113
Nitrosomonadaceae	229	-	-	104	271
Hydrogenophilaceae	227	-	-	-	209
Orbaceae	223	-	-	158	189
Acidithermaceae	218	-	-	-	188
Caviidae	218	-	-	-	-
Parvularculaceae	218	-	-	-	137
Planctomycetaceae	217	-	-	111	206
Tetrahymenidae	214	-	-	-	-
Natranaerobiaceae	212	-	-	132	209
Nitrospiraceae	211	-	-	-	136
African mole-rats	207	-	-	104	158
Deferribacteraceae	207	-	-	279	268
Caldilineaceae	207	-	-	100	224
Syntrophaceae	205	-	-	115	216
Brevibacteriaceae	200	-	-	-	-
Gemmantimonadaceae	199	-	-	-	139
Desulfomicrobiaceae	194	-	-	103	219
Thermoanaerobacterales Family IV. Incerta	186	-	-	109	222

Family Name	Farm B-dust	Farm B-feather 1	Farm B-feather 2	Farm A-dust 1	Farm A-dust 2
Strongylocentrotidae	186	-	-	-	-
Fabaceae	184	-	-	-	389
Chitinophagaceae	178	-	-	-	239
Thermodesulfobacteriaceae	170	-	-	154	203
Desulfurobacteriaceae	168	-	-	-	104
Sulfuricellaceae	163	-	-	-	-
Legionellaceae	160	-	-	-	-
Colwelliaceae	159	-	-	-	144
Entomoplasma group	159	-	-	-	152
Syntrophomonadaceae	158	-	-	-	123
Camelidae	154	-	-	-	209
Debaryomycetaceae	152	-	-	174	-
Cryomorphaceae	151	-	-	113	273
Bdellovibrionaceae	145	-	-	-	142
Ferrimonadaceae	145	-	-	-	105
Saprospiraceae	144	-	-	-	166
Roseiflexaceae	140	-	-	-	324
Chrysiogenaceae	139	-	-	-	196
Tetraodontidae	139	-	-	-	-
Chloroflexaceae	133	-	-	-	-
Draconibacteriaceae	130	-	-	-	117
Cneoraceae	127	-	-	200	300
Culicidae	124	-	-	-	131
Flammeovirgaceae	117	-	-	-	-
Hypocreaceae	113	-	-	-	-
Leeaceae	112	-	-	216	218
Rivulariaceae	111	-	-	-	-
Arthrodermataceae	110	-	-	-	-
Hahellaceae	107	-	-	-	107
Desulfurellaceae	107	-	-	-	117
Pinaceae	103	-	-	-	-
Noelaerhabdaceae	103	-	-	512	379
Acidithiobacillaceae	103	-	-	-	247
Sphaeriaceae	103	-	-	-	-
Costariaceae	-	-	-	130	-
Chaetomiaceae	-	-	-	241	274
Alligatoridae	-	-	-	-	253
Onchocercidae	-	-	-	-	119
Mamiellaceae	-	-	-	-	161
Drosophilidae	-	-	-	-	178
Loliginidae	-	-	-	-	171
Hydridae	-	-	-	-	181
Thermodesulfobiaceae	-	-	-	-	136
Mycosphaerellaceae	-	-	-	349	1,122
Balaenopteridae	-	-	-	-	754
Francisella group	-	-	-	-	146
Lagriidae	-	-	-	-	109
Dipodascaceae	-	-	-	220	-
Syntrophobacteraceae	-	-	-	-	119
Rhabditidae	-	-	-	126	-
Pseudoalteromonadaceae	-	-	-	127	231
Retroviridae	-	-	-	296	174
Sarcocystidae	-	-	-	261	140
Fibrobacteraceae	-	-	-	-	140
Mycosyringaceae	-	-	-	-	228
Magnetococcaceae	-	-	-	-	102
	26,519,857	386,315	342,077	14,241,186	25,157,407