DNA from dust: comparative genomics of large DNA viruses in field surveillance samples

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

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

[2] Center for Infectious Disease Dynamics, Departments 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 (Present address: Indiana University School of Medicine)
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

Running title: Genomic comparison of large DNA viruses from the field

Word counts: Abstract, 219; Importance, 135; Intro/Results/Discussion, 4552; Methods, 2923.

Abstract

The intensification of the poultry industry over the last sixty years facilitated the evolution of increased virulence and vaccine breaks in Marek's disease virus (MDV-1). Full genome sequences are essential for understanding why and how this evolution occurred, but what is known about genome-wide variation in MDV comes from laboratory culture. To rectify this, we developed methods for obtaining high quality genome sequences directly from field samples without the need for sequence-based enrichment strategies prior to sequencing. We applied this to the first characterization of MDV-1 genomes from the field, without prior culture. These viruses were collected from vaccinated hosts that acquired naturally circulating field strains of MDV-1, in the absence of a disease outbreak. This reflects the current issue afflicting the poultry industry, where virulent field strains continue to circulate despite vaccination, and can remain undetected due to the lack of overt disease symptoms. We found that viral genomes from adjacent field sites had high levels of overall DNA identity, and despite strong evidence of purifying selection, had coding variations in proteins associated with virulence and manipulation of host immunity. Our methods empower ecological field surveillance, make it possible to determine the basis of viral virulence and vaccine breaks, and can be used to obtain full genomes from clinical samples of other large DNA viruses, known and unknown.

Importance

Despite both clinical and laboratory data that show increased virulence in field isolates of MDV-1 over the last half century, we do not yet understand the genetic basis of its pathogenicity. Our knowledge of genome-wide variation between strains of this virus comes exclusively from isolates that have been cultured in the laboratory. MDV-1 isolates tend to lose virulence during

59 repeated cycles of replication in the laboratory, raising concerns about the ability of cultured
60 isolates to accurately reflect virus in the field. The ability to directly sequence and compare field
61 isolates of this virus is critical to understanding the genetic basis of rising virulence in the wild.
62 Our approaches remove the prior requirement for cell culture, and allow direct measurement of
63 viral genomic variation within and between hosts, over time, and during adaptation to changing
64 conditions.
65

66 **Introduction**

67 Marek's disease virus (MDV), a large DNA alphaherpesvirus of poultry, became increasingly
68 virulent over the second half of the 20th century, evolving from a virus that caused relatively mild
69 disease to one that can kill unvaccinated hosts lacking maternal antibodies in as little as ten days
70 (1–5). Today, mass immunizations with live-attenuated vaccines help to control production
71 losses, which are mainly associated with immunosuppression and losses due to condemnation of
72 carcasses (4, 6). Almost 9 billion broiler chickens are vaccinated against MD each year in the US
73 alone (7). MD vaccines prevent host animals from developing disease symptoms, but do not
74 prevent them from becoming infected, nor do they block transmission of the virus (6, 8). Perhaps
75 because of that, those vaccines may have created conditions favoring the evolutionary emergence
76 of the hyperpathogenic strains which dominate the poultry industry today (5). Certainly, virus
77 evolution undermined two generations of MD vaccines (1–4). However, the genetics underlying
78 MDV-1 evolution into more virulent forms and vaccine breaks are not well understood (4, 9).
79 Likewise, the nature of the vaccine-break lesions that can result from human immunization with
80 live-attenuated varicella zoster virus (VZV) vaccine is an area of active study (10–13).

Remarkably, our understanding of MDV-1 (genus: *Mardivirus*, species: Gallid alphaherpesvirus type 2) genomics and genetic variation comes exclusively from the study of 10 different laboratory-grown strains (14–21). Most herpesviruses share this limitation, where the large genome size and the need for high-titer samples has led to a preponderance of genome studies on cultured virus, rather than clinical or field samples (22–28). Repeated observations about the loss of virulence during serial passage of MDV-1 and other herpesviruses raises concerns about the ability of cultured strains to accurately reflect the genetic basis of virulence in wild populations of virus (25, 29–31). The ability to capture and sequence viral genomes directly from host infections and sites of transmission is the necessary first step to reveal when and where variations associated with vaccine-breaks arise, which one(s) spread into future host generations, and to begin to understand the evolutionary genetics of virulence and vaccine failure.

Recent high-throughput sequencing (HTS) applications have demonstrated that herpesvirus genomes can be captured from human clinical samples using genome amplification techniques such as oligonucleotide enrichment and PCR amplicon-based approaches (32–36). Here we present a method for the enrichment and isolation of viral genomes from dust and feather follicles, without the use of either of these solution-based enrichment methods. Chickens become infected with MDV by the inhalation of dust contaminated with virus shed from the feather follicles of infected birds. Although these vaccinated hosts were infected by and shedding wild MDV-1, there were no overt disease outbreaks. Deep sequencing of viral DNA from dust and feather follicles enabled us to observe, for the first time, the complete genome of MDV-1 directly from field samples of naturally infected hosts. This revealed variations in both new and

104 known candidates for virulence and modulation of host immunity. These variations were
105 detected both within and between the virus populations at different field sites, and during
106 sequential sampling. One of the new loci potentially associated with virulence, in the viral
107 transactivator ICP4 (MDV084 / MDV100), was tracked using targeted gene surveillance of
108 longitudinal field samples. These findings confirm the genetic flexibility of this large DNA virus
109 in a field setting, and demonstrate how a new combination of HTS and targeted Sanger-based
110 surveillance approaches can be combined to understand viral evolution in the field.
111

112 ## Results

113 **Sequencing, assembly and annotation of new MDV-1 consensus genomes from the field**

114 To assess the level of genomic diversity within and between field sites that are under real world
115 selection, two commercial farms in central Pennsylvania (11 miles apart) with a high prevalence
116 of MDV-1 were chosen **(Figure 1A)**. These operations raise poultry for meat (also known as
117 broilers), and house 25,000-30,000 individuals per house. The poultry were vaccinated with a
118 bivalent vaccine composed of MDV-2 (strain SB-1) and HVT (strain FC126). In contrast to the
119 Rispens vaccine, which is an attenuated MDV-1 strain, MDV-2 and HVT can be readily
120 distinguished from MDV-1 across the length of the genome, which allowed us to differentiate
121 wild MDV-1 from concomitant shedding of vaccine strains. These farms are part of a
122 longitudinal study of MDV-1 epidemiology and evolution in modern agricultural settings (37,
123 38).
124

125 To obtain material for genomic surveillance, we isolated MDV nucleocapsids from dust or
126 epithelial tissues from the individual feather follicles of selected hosts (see Methods and
127 **Supplemental Figure S1-S2** for overviews, and **Supplemental Tables S1-S2** for DNA yields).
128 A total of five uncultured wild-type samples of MDV were sequenced using an in-house Illumina
129 MiSeq sequencer (**Table 1**, lines 4-6; see **Methods** for details). The sequence read data derived
130 from dust contained approximately 2-5% MDV-1 DNA, while the feather samples ranged from
131 ~27%-48% MDV-1 (**Table 1**, line 6). Since dust represents the infectious material that transmits
132 MDV from host to host, and across generations of animals that pass through a farm or field site,
133 we pursued analysis of wild MDV-1 genomes from both types of source material.
134
135 Consensus genomes were created for each of the five samples in **Table 1**, using a recently
136 described combination of *de novo* assembly and reference-guided alignment of large sequence
137 blocks, or contigs (**Figure 2A**) (39). Nearly complete genomes were obtained for all five samples
138 (**Table 1**). The coverage depth for each genome was directly proportional to the number of
139 MDV-1 specific reads obtained from each sequencing library (**Table 1**, line 5,7). The dust
140 sample from Farm B had the highest coverage depth, at an average of almost 600X across the
141 viral genome. Feather 1 from Farm B had the lowest coverage depth, averaging 44X genome-
142 wide, which still exceeds that of most bacterial or eukaryotic genome assemblies. The genome
143 length for all 5 samples was approximately 180 kilobases (**Table 1**), which is comparable to all
144 previously sequenced MDV-1 isolates (14–21).
145
146 For each field sample collected and analyzed here, we assembled a consensus viral genome. We
147 anticipated that the viral DNA present in a single feather follicle might be homotypic, based on

148 similar results found for individual vesicular lesions of the alphaherpesvirus VZV (10, 33). We
149 further expected that the genomes assembled from a dust sample would represent a mix of viral
150 genomes, summed over time and space. Viral genomes assembled from dust represent the most
151 common genome sequence, or alleles therein, from all of the circulating MDV-1 on a particular
152 farm. The comparison of consensus genomes provided a view into the amount of sequence
153 variation between Farm A and Farm B, or between two individuals on the same Farm (**Table 2**).
154 In contrast, examining the polymorphic loci within each consensus genome assembly allowed us
155 to observe the level of variation within the viral population at each point source (**Figures 3-4,**
156 **Supplemental Figure S3, Supplemental Table S3**).
157

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

159 We began our assessment of genetic diversity by determining the extent of DNA and amino acid
160 variations between the five different consensus genomes. We found that the five genomes are
161 highly similar to one another at the DNA level, with the percent homology ranging from 99.4%
162 to 99.9% in pairwise comparisons (**Figure 2C**, **Table 2**). These comparisons used a trimmed
163 genome format (**Figure 2B**) where the terminal repeat regions had been removed, so that these
164 sequences were not over-represented in the analyses. The level of identity between samples is
165 akin to that observed in closely related isolates of herpes simplex virus 1 (HSV-1) (39).
166 Observed nucleotide differences were categorized as genic or intergenic, and further sub-divided
167 based on whether the differences were insertions or deletions (INDELs) or single-nucleotide
168 polymorphism (SNPs) **(Table 2)**. The number of nucleotide differences was higher in intergenic
169 regions than in genic regions for all genomes. For the INDEL differences, we also calculated the
170 minimum number of events that could have led to the observed differences, to provide context on

171 the relative frequency of these loci in each genome. We anticipate that these variations include
172 silent mutations, as well as potentially advantageous or deleterious coding differences.
173
174 To understand the effect(s) of these nucleotide variations on protein coding and function, we
175 next compared the amino acid (AA) sequences of all open reading frames (ORFs) for the five
176 isolates. The consensus protein coding sequences of all five isolates were nearly identical, with
177 just a few differences (**Table 2**). In comparison to the other four samples, Farm B-dust harbored
178 AA substitutions in four proteins. A single non-synonymous mutation was seen in each of the
179 following: the virulence-associated lipase homolog vLIP (MDV010; Farm B-dust, S501A) (40),
180 the MHC class I immune evasion protein LORF2 (MDV012; Farm B-dust, L311W) (41), and the
181 probable membrane protein UL43 (MDV056; Farm B-dust, S74L). A single synonymous
182 mutation was observed in the DNA helicase-primase protein UL8 (MDV020; Farm B-dust,
183 L253L). Finally, a 22 AA insertion unique to Farm B-dust was observed in the DNA polymerase
184 processivity subunit protein UL42 (MDV055; Farm B-dust, insertion at AA 277). We did not
185 observe any coding differences between temporally separated dust isolates from Farm A or
186 between feather isolates from different hosts in Farm B, although both of these comparisons
187 (**Table 2**, bottom) revealed hundreds of noncoding differences.
188

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

190 Comparing viral genomes found in different sites provides a macro-level assessment of viral
191 diversity. We next investigated the presence of polymorphic viral populations within each
192 consensus genome, to reveal how much diversity might exist within a field site (as reflected in
193 dust-derived genomes) or within a single host (as reflected in feather genomes).

194

195 For each consensus genome, we used polymorphism detection analysis to examine the depth and
196 content of the sequence reads at every nucleotide position in each genome (see **Methods** for
197 details). Rather than detecting differences between isolates, as in **Table 2**, this approach revealed
198 polymorphic sites within the viral population that contributed to each consensus genome. We
199 detected 2-58 polymorphic sites within each consensus genome (**Figure 3**) (see **Methods** for
200 details). The feather genomes had a lower number of polymorphisms compared to the dust
201 genomes, which may be due to low within-follicle diversity or the relatively low sequence
202 coverage. INDELs were not included in this polymorphism analysis, but clearly contributed to
203 between-sample variation (**Table 2**), suggesting that this may be an underestimate of the overall
204 amount of within-sample variation. Viral polymorphisms were distributed across the entire
205 length of the genome (**Figure 3**), with the majority concentrated in the repeat regions.
206 Application of a more stringent set of parameters (see **Methods** for details) yielded a similar
207 distribution of polymorphisms, albeit with no polymorphisms detected in feather samples due to
208 their lower depth of coverage (**Supplemental Figure S3**). These data reveal that polymorphic
209 alleles are present in field isolates, including in viral genomes collected from single sites of
210 shedding in infected animals.

211

212 To address the potential effect(s) of these polymorphisms on MDV biology, we divided the
213 observed polymorphisms into categories of synonymous, non-synonymous, genic-untranslated,
214 or intergenic (**Supplemental Table S3**). The majority of all polymorphisms were located in
215 intergenic regions (**Supplemental Table S3**). We next investigated whether evidence of
216 selection could be detected from the distribution of polymorphisms in our samples. One way to

217 assess this is to determine whether the relative frequencies of synonymous, non-synonymous,
218 genic-untranslated, and intergenic polymorphisms can be explained by random chance. If the
219 observed frequencies differ from those expected from a random distribution, it would suggest
220 genetic selection. After calculating the expected distribution in each sample (as described in
221 **Methods**), we determined that the distribution of variants differed from that expected by chance
222 in each of our dust samples (**Figure 4**, Farm A-Dust 1: χ^2=68.16, d.f. =3, p<0.001; Farm A-Dust
223 2: χ^2=128.57, d.f. =3, p<0.001; Farm B-Dust 1: χ^2=63.42, d.f. =3, p<0.001). In addition, we
224 found in pairwise tests that the number of observed intergenic polymorphisms was significantly
225 higher than the observed values for other categories (**Supplemental Table S4**). This suggests
226 that the mutations that occurred in the intergenic regions were better tolerated and more likely to
227 be maintained in the genome; *i.e.* that purifying selection was acting on coding regions.
228

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

230 In addition to observing polymorphic SNPs in each sample at a single moment in time, we
231 explored whether any shifts in polymorphic allele frequency were detected in the two sequential
232 dust samples from Farm A. We found one locus in the ICP4 (MDV084 / MDV100) gene
233 (nucleotide position 5,495) that was polymorphic in the Farm A-dust 2 sample, with nearly equal
234 proportions of sequence reads supporting the major allele (C) and the minor allele (A) (**Figure**
235 **5A**). In contrast, this locus had been 99% A and only 1% C in Farm A-dust 1 (collected 11
236 months earlier in another house on the same farm), such that it was not counted as polymorphic
237 in that sample by our parameters (see **Methods** for details). At this polymorphic locus, the
238 nucleotide C encodes a serine, while nucleotide A encodes a tyrosine. The encoded AA lies in
239 the C-terminal domain of ICP4 (AA position 1,832). ICP4 is an important immediate-early

protein in all herpesviruses, where it serves as a major regulator of viral transcription (42–44). The role of ICP4 in MDV pathogenesis is also considered crucial because of its proximity to the latency associated transcripts (LAT) and recently described miRNAs (44–46). In a previous study of MDV-1 attenuation through serial passage *in vitro*, mutations in ICP4 appeared to coincide with attenuation (31).

Given the very different allele frequencies at this ICP4 locus between two houses on the same farm 11 months apart, we examined dust samples from one of the houses over 9 months with targeted Sanger sequencing of this SNP (**Figure 5B**). We found that this locus was highly polymorphic in time-separated dust samples. The A (Tyrosine) allele rose to almost 50% frequency in the 9 month period. In four of the dust samples, the A (Tyrosine) allele was dominant over the C (serine) allele. This reversible fluctuation in allele frequencies over a short period of time is unprecedented for alphaherpesviruses so far as we know. However, recent studies on human cytomegalovirus (HCMV) have shown that selection can cause viral populations to evolve in short periods of time (34, 35). While this is only one example of a polymorphic locus that shifts in frequency over time, similar approaches could be used at any of the hundreds of polymorphic loci detected here (**Supplemental Table S3**).

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

To compare these new field-based MDV genomes to previously sequenced isolates of MDV, we created a multiple sequence alignment of all available MDV-1 genomes (14, 15, 17–21, 47, 48). The multiple sequence alignment was used to generate a dendrogram depicting genetic relatedness (see **Methods**). We observed that the five new isolates form a separate group when

263 compared to all previously sequenced isolates (**Figure 6**). This may result from geographic
264 differences as previously seen for HSV-1 and VZV (27, 49–52), or from temporal differences in
265 the time of sample isolations, or from the lack of cell-culture adaptation in these new genomes.
266
267 We also noted a distinctive mutation in the genes encoding glycoprotein L (gL; also known as
268 UL1 or MDV013). All of the field isolates had a 12 nucleotide deletion in gL that has been
269 described previously in strains from the Eastern USA. This deletion is found predominantly in
270 very virulent or hypervirulent strains (vv and vv+, in MDV-1 pathotyping nomenclature (1))
271 (53–56). This deletion falls in the putative cleavage site of gL, which is necessary for its post-
272 translational modification in the endoplasmic reticulum (54). Glycoprotein L forms a complex
273 with another glycoprotein H (gH). The gH/gL dimer is conserved across the *Herpesviridae*
274 family and has been associated with virus entry (57, 58).
275
276 These field-isolated genomes also contain a number of previously characterized variations in the
277 oncogenesis-associated Marek's EcoRI-Q-encoded protein (Meq; also known as MDV005,
278 MDV076, and RLORF7). We observed three substitutions in the C-terminal (transactivation)
279 domain of Meq (P153Q, P176A, P217A) (59). The first two of these variations have been
280 previously associated with MDV-1 strains of very virulent and hypervirulent pathotypes (vv and
281 vv+) (53, 60, 61), while the third mutation has been shown to enhance transactivation (62). In
282 contrast, the field isolates lacked the 59 AA insertion in the Meq proline repeats that is often
283 associated with attenuation, as seen in the vaccine strain CVI988 and the mildly-virulent strain
284 CU-2 (47, 48, 63, 64). We also observed a C119R substitution in all five field-derived genomes,
285 which is absent from attenuated and mildly-virulent isolates. This C119R mutation falls in the

286 LXCXE motif of Meq, which normally binds to the tumor suppressor protein Rb to regulate cell
287 cycle progression (53, 65). Although comprehensive *in vitro* and *in vivo* studies will be required
288 to fully understand the biological implications of these variations, sequence comparisons of both
289 gL and Meq from dust and feather genomes suggest that these closely resemble highly virulent
290 (vv and vv+) variants of MDV-1 (47, 48). This is corroborated by the dendrogram (**Figure 6**),
291 where the dust- and feather-derived genomes cluster closely with 648a, which is a highly virulent
292 (vv+) MDV isolate.
293

294 **Assessment of taxonomic diversity in dust and feathers**

295 As noted in **Table 1**, only a fraction of the reads obtained from each sequencing library were
296 specific to MDV-1. We analyzed the remaining sequences to gain insight into the taxonomic
297 diversity found in poultry dust and feathers. Since our enrichment for viral capsids removed most
298 host and environmental contaminants, the taxa observed here represent only a fraction of the
299 material present initially. However it provides useful insight into the overall complexity of each
300 sample type. The results of the classification for Farm B- dust, Farm B-feather 1, and Farm B-
301 feather 2 are shown in **Supplemental Figure S4**. We divided the sequence reads by the different
302 kingdoms they represent. Complete lists of taxonomic diversity for all samples to the family
303 level are listed in **Supplemental Table S5**. As expected, the taxonomic diversity of dust is
304 greater than that of feather samples. The majority of sequences in the dust samples mapped to the
305 chicken genome, and only about 2-5% were MDV specific (see also **Table 1**, line 6). We found
306 that single feathers were a better source of MDV DNA, due to their reduced level of taxonomic
307 diversity and higher percentage of MDV-specific reads (**Table 1**, line 6 and **Supplemental**
308 **Figure S4**).

309

310 **Discussion**

311 This study presents the first description of MDV-1 genomes sequenced directly from a field
312 setting. This work builds on recent efforts to sequence VZV and HCMV genomes directly from
313 human clinical samples, but importantly the approaches presented here do not employ either the
314 oligo-enrichment used for VZV or the PCR-amplicon strategy used for HCMV (10, 33–35, 66).
315 This makes our technique widely accessible and reduces potential methodological bias. It is also
316 more rapid to implement and is applicable to the isolation of unknown large DNA viruses, since
317 it does not rely on sequence-specific enrichment strategies. These five genomes were
318 interrogated at the level of comparing consensus genomes – between-host variation – as well as
319 within each consensus genome – within-host variation. By following up with targeted PCR and
320 Sanger sequencing, we demonstrate that HTS can rapidly empower molecular epidemiological
321 field surveillance of loci undergoing genetic shifts.

322

323 Although a limited number of non-synonymous differences were detected between the field
324 samples compared here, it is striking that several of these (vLIP, LORF2, UL42) have been
325 previously demonstrated to have roles in virulence and immune evasion. The N-glycosylated
326 protein viral lipase (vLIP; MDV010) encodes a 120 kDa protein that is required for lytic virus
327 replication in chickens (40, 3). The vLIP gene of MDV-1 is homologous to other viruses in the
328 *Mardivirus* genus as well as to avian adenoviruses (67–69). The S501A mutation in the second
329 exon of vLIP protein is not present in the conserved region that bears homology to other
330 pancreatic lipases (40). The viral protein LORF2 (MDV012) is a viral immune evasion gene that

suppresses MHC class I expression by inhibiting TAP transporter delivery of peptides to the endoplasmic reticulum (41). LORF2 is unique to the non-mammalian *Mardivirus* clade, but its function is analogous to that of the mammalian alphaherpesvirus product UL49.5 (41, 70, 71). Another study has shown that LORF2 is an essential phosphoprotein with a potential role as a nuclear/cytoplasmic shuttling protein (72). Interestingly, we also observed a 22 amino acid insertion in the DNA polymerase processivity subunit protein UL42 (MDV055). In herpesviruses, UL42 has been recognized as an integral part of the DNA polymerase complex, interacting directly with DNA and forming a heterodimer with the catalytic subunit of the polymerase (73–75). In HSV-1, the N-terminal two thirds of UL42 have been shown to be sufficient for all known functions of UL42; the insertion in Farm B-dust falls at the edge of this N-terminal region (76). The non-synonymous mutations and insertions detected here warrant further study to evaluate their impacts on protein function and viral fitness *in vivo*. The fact that any coding differences were observed in this small sampling of field-derived genomes suggests that the natural ecology of MDV-1 may include mutations and adaptations in protein function, in addition to genetic drift.

Drug resistance and vaccine failure have been attributed to the variation present in viral populations (10, 33, 77). Polymorphic populations allow viruses to adapt to diverse environments and withstand changing selective pressures, such as evading the host immune system, adapting to different tissue compartments, and facilitating transmission between hosts (10, 26, 33–35, 66, 77, 78). Polymorphisms that were not fully penetrant in the consensus genomes, but that may be fodder for future selection, include residues in genes associated with virulence and immune evasion, such as ICP4 (**Figure 5**), Meq, pp38, vLIP, LORF2, and others

354 (**Supplemental Table S3**). The non-synonymous polymorphism that we observed in Meq is a
355 low-frequency variant present in the C-terminal domain (I201L) (Farm B-dust, **Supplemental**
356 **Table S3**). However a comparison of 88 different Meq sequences from GenBank and
357 unpublished field isolates (37) did not reveal any examples where leucine was the dominant
358 allele; all sequenced isolates to date have isoleucine at position 201.
359
360 Previous studies have examined the accumulation of polymorphic loci in MDV-1 genomes after
361 serial passage *in vitro* (30, 31). Overall, we found a similar quantity of polymorphisms in field-
362 derived genomes as found in these prior studies, but we did not find any specific polymorphic
363 loci that were identical between field-derived and *in vitro*-passaged genomes (30, 31). The genes
364 ICP4 (MDV084 / MDV100), LORF2 (MDV012), UL42 and MDV020 contain polymorphic in
365 both field and serially-passaged isolates, albeit at different loci (30, 31). It is noteworthy that
366 these coding variations are detected despite signs of clearance of polymorphisms from coding
367 regions (**Figure 4**), as indicated by the higher-than-expected ratios of intergenic to coding
368 polymorphisms in these genomes. Together these findings suggest that MDV-1 exhibits genetic
369 variation and undergoes rapid selection in the field, which may demonstrate the basis of its
370 ability to overcome vaccine induced host-resistance to infection (3, 79, 80, 5).
371
372 For the viral transactivation protein ICP4, we explored the penetrance of a polymorphic locus
373 (nucleotide position 5,495) both in full-length genomes, and also via targeted sequencing over
374 time. Most of the work on this region in the MDV-1 genome has actually focused on the LAT
375 transcripts that lie antisense to the ICP4 gene (44, 45). This polymorphic locus could thus impact
376 either ICP4's coding sequence (AA 1,832 serine vs. tyrosine) or the sequence of the LAT

377 transcripts. This variation in ICP4 lies in the C-terminal domain, which in HSV-1 has been
378 implicated in the DNA synthesis, late gene expression, and intranuclear localization functions of
379 ICP4 (81, 82). This combination of deep-sequencing genomic approaches to detect new
380 polymorphic loci, and fast gene-specific surveillance to track changes in SNP frequency over a
381 larger number of samples, illustrate the power of high-quality full genome sequences from field
382 samples to provide powerful new markers for field ecology.
383
384 Our comparison of new field-isolated MDV-1 genomes revealed a distinct genetic clustering of
385 these genomes, separate from other previously sequenced MDV-1 genomes (**Figure 6**). This
386 pattern may results from geographic and temporal drift in these strains, or from the wild, virulent
387 nature of these strains vs. the adaptation(s) to tissue culture in all prior MDV-1 genome
388 sequences. The impact of geography on the genetic relatedness of herpesvirus genomes has been
389 previously shown for related alphaherpesviruses such as VZV and HSV-1 (27, 49–52).
390 Phenomena such as recombination can also have an impact on the clustering pattern of MDV
391 isolates. It is worth noting that the genetic distance dendrogram constructed here included
392 genomes from isolates that were collected over a 40 year span, which introduces the potential for
393 temporal drift (14, 15, 17–21, 47, 48). Agricultural and farming practices have evolved
394 significantly during this time, and we presume that pathogens have kept pace. To truly
395 understand the global diversity of MDV, future studies will need to include the impacts of
396 recombination and polymorphisms within samples, in addition to the overall consensus-genome
397 differences reflected by static genetic distance analyses.
398

399 Prior studies have shown that when MDV is passaged for multiple generations in cell-culture, the
400 virus accumulates a series of mutations, including several that affect virulence (30). The same is
401 true for the betaherpesvirus HCMV (25). Extended passage *in vitro* forms the basis of vaccine
402 attenuation strategies, as for the successful vaccine strain (vOka) of the alphaherpesvirus VZV
403 (83). Cultured viruses can undergo bottlenecks during initial adaptation to cell culture, and they
404 may accumulate variations and loss of function mutations by genetic drift or positive selection.
405 The variations and mutations thus accumulated may have little relationship to virulence and the
406 balance of variation and selection in the field. We thus anticipate that these field-isolated viral
407 genomes more accurately reflect the genomes of wild MDV-1 that are circulating in the field.
408 The ability to access and compare virus from virulent infections in the field will enable future
409 analyses of vaccine-break viruses.
410
411 Our data and approaches provide powerful new tools to measure viral diversity in field settings,
412 and to track changes in large DNA virus populations over time in hosts and ecosystems. In the
413 case of MDV-1, targeted surveillance based on an initial genomic survey could be used to track
414 viral spread across a geographic area, or between multiple end-users associated with a single
415 parent corporation (**Figure 1B**). Similar approaches could be implemented for public-or animal-
416 health programs, for instance to guide management decisions on how to limit pathogen spread
417 and contain airborne pathogens. The ability to sequence and compare large viral genomes
418 directly from individual hosts and field sites will allow a new level of interrogation of host-virus
419 fitness interactions, which form the basis of host resistance to infection (**Figure 1B**). Finally, the
420 analysis of viral genomes from single feather follicles, as from single VZV vesicles, enables our
421 first insights into naturally-occurring within-host variation during infection and transmission

422 (**Figure 1B**). Evidence from tissue compartmentalization studies in HCMV and VZV suggests
423 that viral genomes differ in distinct body niches (10, 35, 66). These new technique will enable us
424 to ask similar questions about MDV-1, and to begin exploring the relative fitness levels of
425 viruses found in different tissue compartments.

426 Materials and Methods

427 Collection of dust and feathers

428 Samples were collected from two commercial-scale farms in central Pennsylvania, where each
429 poultry building housed 25,000-30,000 individuals **(Figure 1A)**. The poultry on both farms were
430 the same breed and strain of colored ("red") commercial broiler chicken from the same hatchery
431 and company. Dust samples were collected into 1.5 ml tubes from fan louvers. This location
432 contains less moisture and contaminants than floor-collected samples, and represents a mixture
433 of air-borne virus particles and feather dander. Sequential samples from Farm A (**Table 1**) were
434 collected 11 months apart, from adjacent houses on the same farm (**Figure 1A**). Samples from
435 Farm B (**Table 1**) were collected from a single house, at a single point in time. Feathers were
436 collected just before hosts were transported from the farms for sale, to maximize the potential for
437 infection and high viral titer. At the time of collection the animals were 10-12 weeks old. Ten
438 individuals were chosen randomly throughout the entirety of one house for feather collection.
439 Two feathers from each animal were collected from the axillary track (breast feathers). The distal
440 0.5-1.0 cm proximal shaft or feather tip, which contains the feather pulp, was snipped into a
441 sterile 1.5 ml micro-tube containing a single sterile 5 mm steel bead (Qiagen). On return to the
442 laboratory, tubes were stored at -80°C until processing. One feather from each animal was tested
443 for the presence and quantity and MDV-1 present (see below for quantitative PCR details). The

444 remaining feathers from the two animals with highest apparent MDV-1 titer were used for a
445 more thorough DNA extraction (see below for details) and next-generation sequencing. Animal
446 procedures were approved by the Institutional Animal Care and Use Committee of the
447 Pennsylvania State University (IACUC Protocol#: 46599).
448

449 **Viral DNA isolation from dust**

450 MDV nucleocapsids were isolated from dust as indicated in **Supplemental Figure S1A**. Dust
451 collected from poultry houses was stored in 50 ml polypropylene Falcon® tubes (Corning) at 4°C
452 until required. 500 mg of dust was suspended in 6.5 ml of 1X phosphate buffered saline (PBS).
453 To distribute the dust particles into solution and help release cell-associated virus, the mixture
454 was vortexed vigorously until homogenous and centrifuged at 2000 × g for 10 minutes. This
455 supernatant was further agitated on ice for 30 sec. using a Sonica Ultrasonic Processor Q125
456 (probe sonicator with 1/8th inch microtip) set to 20% amplitude. It was then vortexed before
457 being centrifuged for a further 10 minutes at 2000 × g. To enrich viral capsids away from the
458 remaining contaminants, the supernatant (approximately 5 ml in volume) was subjected to a
459 series of filtration steps. First, we used a Corning® surfactant-free cellulose acetate (SFCA) filter
460 (0.8 μM) that had been soaked overnight in fetal bovine serum (FBS) to remove particles at the
461 level of eukaryotic cells, and bacteria. To remove smaller contaminants, the flow-through was
462 then passed through a Millipore Express® PLUS Membrane vacuum filter (0.22 μM) and the
463 membrane subsequently washed twice with 2.5 ml of PBS. To remove contaminant DNA, the
464 final filtrate (approximately 10 ml in volume) was treated with DNase (Sigma) at a concentration
465 of 0.1 mg/ml for 30 minutes at room temperature. In the absence of DNase treatment we
466 observed a higher yield of viral DNA, but with much lower purity (data not shown). The MDV

467 nucleocapsids present in the DNase-treated solution were captured on a polyethersulfone (PES)
468 membrane (VWR) filter (0.1 μM). This filter membrane trapped the viral nucleocapsids, which
469 are between 0.1-0.2 μm (84). An increased MDV purity, but ultimately reduced total nanograms
470 of DNA yield, may be achieved by washing this membrane once with 2.5 ml PBS (see
471 **Supplemental Table S1**). In the future, samples with a higher percentage of MDV DNA could
472 be obtained by applying these wash steps to all components of the sample pool. The membrane
473 was then carefully excised using a sterile needle and forceps, and laid – exit side downwards – in
474 a sterile 5 cm diameter plastic petri-dish where it was folded twice lengthwise. The "rolled"
475 membrane was then placed into a 2 ml micro-tube containing 1.8 ml of lysis solution (ATL
476 buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen). Digestion was
477 allowed to proceed at 56°C for 1 hour on an incubating microplate shaker (VWR) set to 1100
478 rpm. The membrane was then removed, held vertically over a tilted sterile 5 cm diameter plastic
479 petri-dish and washed with a small volume of the lysis solution (from the 2 ml micro-tube). This
480 wash was subsequently returned to the 2 ml micro-tube and the tube replaced on the heated
481 shaker where it was allowed to incubate overnight. The following day, the DNA was isolated as
482 per manufacturer's instructions using the DNeasy® Blood and Tissue kit (Qiagen). DNA was
483 eluted in 200 μl DNase-free water. Ten to fourteen aliquots of 500 mg each were used to obtain
484 sufficient DNA for each dust sample (see **Supplemental Table S1**). Quantitative PCR was used
485 to assess the copy number of viral genomes in the resulting DNA. Total yield and percent MDV-
486 1 vs. MDV-2 DNA are listed in **Supplemental Table S1**.
487

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

489 The protocol for extraction of MDV DNA from feather follicles was optimized for the smaller
490 input material and an expectation of higher purity (**Supplemental Figure S1B**). Sequential size
491 filters were not used to filter out contaminants from feather follicles, since these direct host
492 samples have fewer impurities than the environmental samples of dust. However, the feather
493 follicle cells were encased inside the keratinaceous shell of the feather tip, which required
494 disruption to release the cells. Each tube containing a single feather tip and one sterile 5 mm
495 diameter steel bead was allowed to thaw, and then 200 µl of PBS was added and the sample
496 bead-beaten for 30 seconds at 30 Hz using a Tissuelyser (Qiagen) (**Supplemental Figure S1B**).
497 Vigorous bead-beating achieved the desired destruction of the follicle tip. To dissociate the cells,
498 80 µl of 2.5 mg/ml trypsin (Sigma) and 720 µl of PBS were then added (final trypsin
499 concentration: 0.8 mg/ml), and the solution was transferred to a new sterile 2 ml micro-tube and
500 incubated for 2 hours at 37°C on a heated microplate shaker (VWR) set to 700 rpm. To release
501 cell-associated virus, the suspension was then sonicated on ice for 30 seconds using a Sonica
502 Ultrasonic Processor Q125 (probe sonicator with 1/8th inch microtip) set to 50% amplitude.
503 DNase I was added to a final concentration of 0.1 mg/ml and allowed to digest for 1 hour at
504 room temperature to remove non-encapsidated DNA. An equal volume of lysis solution (ATL
505 buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen) was added and the
506 sample was incubated over night at 56°C on an Incubating Microplate Shaker (VWR) set to 1100
507 rpm. The following day, the DNA was isolated as per manufacturer's instructions using the
508 DNeasy® Blood and Tissue kit (Qiagen). While the overall amount of DNA obtained from
509 feather follicles was lower than that obtained from pooled dust samples (**Supplemental Table**

510 **S2)**, it was of higher purity and was sufficient to generate libraries for sequencing (**Table 1**, lines
511 1-3).
512

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

514 The total amount of DNA present in the samples was quantified by fluorescence analysis using a
515 Qubit® fluorescence assay (Invitrogen) following the manufacturer's recommended protocol.
516 MDV genome copy numbers were determined using serotype-specific quantitative PCR (qPCR)
517 primers and probes, targeting either the MDV-1 pp38 (MDV073; previously known as
518 LORF14a) gene or MDV-2 (SB-1 strain) DNA polymerase (UL42, MDV055) gene. The MDV-1
519 assay was designed by Sue Baigent: forward primer (Spp38for) 5'-
520 GAGCTAACCGGAGAGGGAGA-3'; reverse primer (Spp38rev) 5'-
521 CGCATACCGACTTTCGTCAA-3'; probe (MDV-1) 6FAM-CCCACTGTGACAGCC-BHQ1
522 (S. Baigent, *pers. comm.*). The MDV-2 assay is that of Islam et al. (85), but with a shorter MGB
523 probe (6FAM-GTAATGCACCCGTGAC-MGB) in place of their BHQ-2 probe. Real-time
524 quantitative PCRs were performed on an ABI Prism 7500 Fast System with an initial
525 denaturation of 95°C for 20 seconds followed by 40 cycles of denaturation at 95°C for 3 seconds
526 and annealing and extension at 60°C for 30 seconds. Both assays included 4 µl of DNA in a total
527 PCR reaction volume of 20 µl with 1X PerfeCTa™ qPCR FastMix™ (Quanta Biosciences),
528 forward and reverse primers at 300 nM and TaqMan ® BHQ (MDV-1) or MGB (MDV-2)
529 probes (Sigma and Life Sciences, respectively) at 100 nM and 200 nM, respectively. In addition
530 each qPCR reaction incorporated 2 µl BSA (Sigma). Absolute quantification of genomes was
531 based on a standard curve of serially diluted plasmids cloned from the respective target genes.
532 The absolute quantification obtained was then converted to concentration. Once the

533 concentration of the total DNA, MDV-1, and MDV-2 DNA present in the sample were known,
534 we calculated the percentage of MDV-1 and MDV-2 genomic DNA in the total DNA pool (see
535 **Supplemental Tables S1-S2**).
536

537 **Illumina next-generation sequencing**

538 Sequencing libraries for each of the isolates were prepared using the Illumina TruSeq Nano DNA
539 Sample Prep Kit, according to the manufacturer's recommended protocol for sequencing of
540 genomic DNA. Genomic DNA inputs used for each sample are listed in **Table 1**. The DNA
541 fragment size selected for library construction was 550 base pairs (bp). All the samples were
542 sequenced on an in-house Illumina MiSeq using version 3 sequencing chemistry to obtain
543 paired-end sequences of 300 × 300 bp. Base calling and image analysis was performed with the
544 MiSeq Control Software (MCS) version 2.3.0.
545

546 **Consensus genome assembly**

547 As our samples contained DNA from many more organisms than just MDV, we developed a
548 computational workflow (**Supplemental Figure S2**) to preprocess our data prior to assembly. A
549 local BLAST database was created from every *Gallid herpesvirus* genome available in GenBank.
550 All sequence reads for each sample were then compared to this database using BLASTN (86)
551 with a loose e-value less than or equal to 10^{-2} in order to computationally enrich for sequences
552 related to MDV. These "MDV-like" reads were then processed for downstream genome
553 assembly. The use of bivalent vaccine made it possible for us to readily distinguish sequence
554 reads that resulted from the shedding of virulent MDV-1 vs. vaccine virus (MDV-2 or HVT)

strains. The overall DNA identity of MDV-1 and MDV-2 is just 61% (87). In a comparison of strains MDV-1 Md5 (NC_002229) and MDV-2 SB-1 (HQ840738), we found no spans of identical DNA greater than 50 bp (data not shown). This allowed us to accurately distinguish these 300 x 300 bp MiSeq sequence reads as being derived from either MDV-1 or MDV-2.

MDV genomes were assembled using the viral genome assembly VirGA (39) workflow which combines quality control preprocessing of reads, *de novo* assembly, genome linearization and annotation, and post-assembly quality assessments. For the reference-guided portion of viral genome assembly in VirGA, the *Gallid herpesvirus 2* (MDV-1) strain MD5 was used (GenBank Accession: NC_002229.3). These new genomes were named according to recent recommendations, as outlined by Kuhn et al (88). We use shortened forms of these names throughout the manuscript (see **Table 1** for short names). The full names for all five genomes are as follows: MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 1; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 2; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm B-dust; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm B-feather 1; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm B-feather 2. GenBank Accessions are listed below and in **Table 1**. Annotated copies of each genome, in a format compatible with genome- and sequence browsers, are available at the Pennsylvania State University ScholarSphere data repository: https://scholarsphere.psu.edu/collections/1544bp14j.

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

577 Clustalw2 (43) was used to construct pairwise global nucleotide alignments between whole
578 genome sequences, and pairwise global amino acid alignments between open reading frames.
579 These alignments were utilized by downstream custom Python scripts to calculate percent
580 identity, protein differences, and variation between samples.
581
582 The proline-rich region of UL36 (also known as VP1/2 or MDV049), which contains an
583 extended array of tandem repeats, was removed from all five consensus genomes prior to
584 comparison. The amount of polymorphism seen in this region of UL36 is driven by fluctuations
585 in the length of these tandem repeats, as has been seen in prior studies with other
586 alphaherpesviruses such as HSV, VZV, and pseudorabies virus (PRV) (32,48–50). Since the
587 length of extended arrays of perfect repeats cannot be precisely determined by *de novo* assembly
588 (22, 23, 26, 27), we excluded this region from pairwise comparisons of genome-wide variation.
589 Genome alignments with and without the UL36 region removed are archived at the
590 ScholarSphere site: https://scholarsphere.psu.edu/collections/1544bp14j.
591

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

593 VarScan v2.2.11 (91) was used to detect variants present within each consensus genome. To aid
594 in differentiating true variants from potential sequencing errors (92), two separate variant calling
595 analyses were explored. (10). Our main polymorphism-detection parameters (used in **Figures 3-**
596 **4** and **Supplemental Tables S3-S4**) were as follows: minimum variant allele frequency ≥ 0.02;
597 base call quality ≥ 20; read depth at the position ≥ 10; independent reads supporting minor allele
598 ≥ 2. Directional strand bias ≥ 90% was excluded; a minimum of two reads in opposing directions

599 was required. For comparison and added stringency, we also explored a second set of parameters
600 (used in **Supplemental Figure S3**): minimum variant allele frequency ≥ 0.05; base call quality ≥
601 20; read depth at the position ≥ 100; independent reads supporting minor allele ≥5. Directional
602 strand bias ≥ 80% was excluded. The variants obtained from VarScan were then mapped back to
603 the genome to understand their distribution and mutational impact using SnpEff and SnpSift (93,
604 94). Polymorphisms in the proline-rich region of UL36 were excluded, as noted above.
605

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

607 For each of our five consensus genomes, which each represent a viral population, we classified
608 the polymorphisms detected into categories of synonymous, non-synonymous, genic-
609 untranslated, or intergenic, based on where each polymorphism was positioned in the genome.
610 For these analyses (**Figure 4**), we were only able to include polymorphisms detected in the three
611 dust genomes, since the total number of polymorphisms obtained from feather genomes was too
612 low for chi-square analysis. First, we calculated the total possible number of single nucleotide
613 mutations that could be categorized as synonymous, non-synonymous, genic-untranslated or
614 intergenic. To remove ambiguity when mutations in overlapping genes could be classified as
615 either synonymous or non-synonymous, genes with alternative splice variants or overlapping
616 reading frames were excluded from these analyses. This removed 25 open reading frames
617 (approximately 21% of the genome). These tallies of potential mutational events were used to
618 calculate the expected fraction of mutations in each category. We preformed chi-squared tests on
619 each dataset to assess whether the observed distribution of polymorphisms matched the expected
620 distribution. We also performed a similar analysis in pairwise fashion (**Supplemental Table S4**),
621 to assess whether the fraction of variants differed from what would be expected by random

622 chance. Pairwise combinations included the following: synonymous vs. non-synonymous,
623 synonymous vs. intergenic, synonymous vs. genic-untranslated, non-synonymous vs. intergenic,
624 non-synonymous vs. genic-untranslated, and intergenic vs. genic-untranslated. Statistically
625 significant outcomes would suggest that recent or historical selection differed between those
626 categories of variants.
627

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

629 A potential locus of active selection within the ICP4 (MDV084 / MDV100) gene was detected
630 during deep-sequencing of Farm B-dust. This locus was examined using Sanger sequencing. An
631 approximately 400 bp region of the ICP4 gene was amplified using a Taq PCR Core Kit
632 (Qiagen) and the following primers at 200 nM: forward primer (ICP4selF)
633 5'AACACCTCTTGCCATGGTTC 3'; reverse primer (ICP4selR)
634 5'GGACCAATCATCCTCTCTGG 3'. Cycling conditions included an initial denaturation of
635 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at
636 55°C for 30 seconds and extension at 72°C for 1 minute, with a terminal extension at 72°C for 10
637 minutes. The total reaction volume of 50 µl included 10 µl of DNA and 4 µl BSA (final
638 concentration 0.8 mg/ml). Amplification products were visualized on a 1.5% agarose gel, the
639 target amplicon excised and then purified using the E.Z.N.A. Gel Extraction Kit (Omega Bio-
640 tek). Sanger sequencing was performed by the Penn State Genomics Core Facility utilizing the
641 same primers as used for DNA amplification. The relative peak height of each base call at the
642 polymorphic position was analyzed using the ab1PeakReporter tool (95).
643

Genetic distance and dendrogram

Multiple sequence alignments of complete MDV-1 (*Gallid herpesvirus 2*) genomes from GenBank and those assembled by our lab were generated using MAFFT (96). The evolutionary distances were computed using the Jukes-Cantor method (97) and the evolutionary history was inferred using the neighbor-joining method (98) in MEGA6 (99), with 1,000 bootstrap replicates (100). Positions containing gaps and missing data were excluded. The 18-strain genome alignment is archived at ScholarSphere: https://scholarsphere.psu.edu/collections/1544bp14j.

Taxonomic estimation of non-MDV sequences in dust and feathers

All sequence reads from each sample were submitted to a quality control preprocessing method to remove sequencing primers, artifacts, and areas of low confidence (39). Sequence annotation was performed using a massively iterative all-vs.-all BLASTN (E-value $\leq 10^{-2}$) approach using the all-nucleotide-database from NCBI. Only a portion of the total sequence read pool could be identified with confidence using this method. We then used *de novo* assembly to extend the length of these unidentified sequences, therefore elongating them into contigs. These were iterated through BLASTN again, which revealed alignment to repetitive regions of the *Gallus domesticus* (chicken) genome. Since the viral DNA enrichment procedures include a level of stochasticity in removal of host and environmental contaminants, the proportion of taxa present is not a definitive outline of those present initially. The results of these classifications are shown in **Supplemental Figure S4** and listed in **Supplemental Table S5**.

665 ## GenBank accession numbers and availability of materials

666 GenBank Accessions are listed here and in **Table 1**: Farm A - dust 1, KU173116; Farm A - dust
667 2, KU173115; Farm B – dust, KU173119; Farm B - feather 1, KU173117; Farm B - feather 2,
668 KU173118. Additional files used in this manuscript, such as multiple-sequence alignments of
669 these genomes, are archived and available at ScholarSphere:
670 https://scholarsphere.psu.edu/collections/1544bp14j

671 ## Acknowledgements

672 We thank Sue Baigent, Michael DeGiorgio, Peter Kerr and members of Szpara and Read labs for
673 helpful feedback and discussion. This work was supported and inspired by the Center for
674 Infectious Disease Dynamics and the Huck Institutes for the Life Sciences, as well as by startup
675 funds (MLS) from the Pennsylvania State University. This work was part funded by the Institute
676 of General Medical Sciences, National Institutes of Health (R01GM105244; AFR) as part of the
677 joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program.

678 ## References

679 1. **Witter RL**. 1997. Increased virulence of Marek's disease virus field isolates. Avian Dis
680 149–163.

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

683 3. **Osterrieder N**, **Kamil JP**, **Schumacher D**, **Tischer BK**, **Trapp S**. 2006. Marek's disease
684 virus: from miasma to model. Nat Rev Microbiol **4**:283–294.

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

687 5. **Read AF**, **Baigent SJ**, **Powers C**, **Kgosana LB**, **Blackwell L**, **Smith LP**, **Kennedy DA**,
688 **Walkden-Brown SW**, **Nair VK**. 2015. Imperfect Vaccination Can Enhance the
689 Transmission of Highly Virulent Pathogens. PLOS Biol **13**:e1002198.

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

692 7. Poultry Slaughter Annual Summary. U S Dep Agric Econ Stat Mark Inf Syst. Government
693 data.

694 8. **Islam AFMF**, **Walkden-Brown SW**, **Groves PJ**, **Underwood GJ**. 2008. Kinetics of
695 Marek's disease virus (MDV) infection in broiler chickens 1: effect of varying vaccination
696 to challenge interval on vaccinal protection and load of MDV and herpesvirus of turkey in
697 the spleen and feather dander over time. Avian Pathol **37**:225–235.

698 9. **Nair V**. 2005. Evolution of Marek's disease – A paradigm for incessant race between the
699 pathogen and the host. Vet J **170**:175–183.

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

704 11. **Quinlivan M**, **Breuer J**. 2014. Clinical and molecular aspects of the live attenuated Oka
705 varicella vaccine: Studies of the Oka varicella vaccine. Rev Med Virol **24**:254–273.

706 12. **Zerboni L**, **Sen N**, **Oliver SL**, **Arvin AM**. 2014. Molecular mechanisms of varicella zoster
707 virus pathogenesis. Nat Rev Microbiol **12**:197–210.

708 13. **Weinert LA**, **Depledge DP**, **Kundu S**, **Gershon AA**, **Nichols RA**, **Balloux F**, **Welch JJ**,
709 **Breuer J**. 2015. Rates of Vaccine Evolution Show Strong Effects of Latency: Implications
710 for Varicella Zoster Virus Epidemiology. Mol Biol Evol **32**:1020–1028.

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

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

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

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

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

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

726 20. **Spatz SJ**, **Volkening JD**, **Gimeno IM**, **Heidari M**, **Witter RL**. 2012. Dynamic
727 equilibrium of Marek's disease genomes during in vitro serial passage. Virus Genes
728 **45**:526–536.

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

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

736 23. **Tyler SD**, **Peters GA**, **Grose C**, **Severini A**, **Gray MJ**, **Upton C**, **Tipples GA**. 2007.
737 Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain
738 variability with implications for attenuation and phenotypic differences. Virology
739 **359**:447–58.

740 24. **Bradley AJ**, **Lurain NS**, **Ghazal P**, **Trivedi U**, **Cunningham C**, **Baluchova K**, **Gatherer**
741 **D**, **Wilkinson GWG**, **Dargan DJ**, **Davison AJ**. 2009. High-throughput sequence analysis
742 of variants of human cytomegalovirus strains Towne and AD169. J Gen Virol **90**:2375–
743 80.

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

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

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

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

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

761 30. **Spatz SJ**. 2010. Accumulation of attenuating mutations in varying proportions within a
762 high passage very virulent plus strain of Gallid herpesvirus type 2. Virus Res **149**:135–
763 142.

764 31. **Hildebrandt E**, **Dunn JR**, **Perumbakkam S**, **Niikura M**, **Cheng HH**. 2014.
765 Characterizing the molecular basis of attenuation of Marek's disease virus via in vitro
766 serial passage identifies de novo mutations in the helicase-primase subunit gene UL5 and
767 other candidates associated with reduced virulence. J Virol **88**:6232–42.

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

772 33. **Depledge DP**, **Palser AL**, **Watson SJ**, **Lai IY-C**, **Gray ER**, **Grant P**, **Kanda RK**,
773 **Leproust E**, **Kellam P**, **Breuer J**. 2011. Specific Capture and Whole-Genome
774 Sequencing of Viruses from Clinical Samples. PLoS ONE **6**:e27805.

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

778 35. **Renzette N**, **Gibson L**, **Bhattacharjee B**, **Fisher D**, **Schleiss MR**, **Jensen JD**, **Kowalik**
779 **TF**. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography
780 and positive selection. PLoS Genet **9**:1–14.

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

784 37. **Bell AS**, **Jones MJ**, **Christopher CL**, **Kennedy DA**, **Pandey U**, **Dunn PA**, **Szpara ML**,
785 **Read AF**. Marek's disease: MDV-1 diversity and dynamics of infection in the field, a
786 Sanger sequencing-based study in Pennsylvania, USA (In prep).

787 38. **Kennedy DA**, **Cairns CL**, **Jones MJ**, **Bell AS**, **Salathe RM**, **Baigent SJ**, **Nair VK**, **Dunn**
788 **PA**, **Read AF**. Industry-wide surveillance of Marek's disease virus on commercial poultry
789 farms: underlying potential for virulence evolution and vaccine escape (In review).

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

793 40. **Kamil JP**, **Tischer BK**, **Trapp S**, **Nair VK**, **Osterrieder N**, **Kung H-J**. 2005. vLIP, a
794 Viral Lipase Homologue, Is a Virulence Factor of Marek's Disease Virus. J Virol
795 **79**:6984–6996.

796 41. **Hearn C**, **Preeyanon L**, **Hunt HD**, **York IA**. 2015. An MHC class I immune evasion gene
797 of Marek′s disease virus. Virology **475**:88–95.

798 42. **Ahlers SE**, **Feldman LT**. 1987. Immediate-early protein of pseudorabies virus is not
799 continuously required to reinitiate transcription of induced genes. J Virol **61**:1258–1260.

800 43. **Wu CL**, **Wilcox KW**. 1991. The conserved DNA-binding domains encoded by the herpes
801 simplex virus type 1 ICP4, pseudorabies virus IE180, and varicella-zoster virus ORF62
802 genes recognize similar sites in the corresponding promoters. J Virol **65**:1149–59.

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

806 45. **Cantello JL**, **Parcells MS**, **Anderson AS**, **Morgan RW**. 1997. Marek's disease virus
807 latency-associated transcripts belong to a family of spliced RNAs that are antisense to the
808 ICP4 homolog gene. J Virol **71**:1353–1361.

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

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

812 48. **Spatz SJ**, **Petherbridge L**, **Zhao Y**, **Nair V**. 2007. Comparative full-length sequence
813 analysis of oncogenic and vaccine (Rispens) strains of Marek's disease virus. J Gen Virol
814 **88**:1080–1096.

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

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

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

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

825 53. **Shamblin CE**, **Greene N**, **Arumugaswami V**, **Dienglewicz RL**, **Parcells MS**. 2004.
826 Comparative analysis of Marek's disease virus (MDV) glycoprotein-, lytic antigen pp38-
827 and transformation antigen Meq-encoding genes: association of meq mutations with
828 MDVs of high virulence. Vet Microbiol **102**:147–167.

829 54. **Santin ER**, **Shamblin CE**, **Prigge JT**, **Arumugaswami V**, **Dienglewicz RL**, **Parcells MS**.
830 2006. Examination of the Effect of a Naturally Occurring Mutation in Glycoprotein L on
831 Marek's Disease Virus Pathogenesis. Avian Dis **50**:96–103.

832 55. **Tavlarides-Hontz P**, **Kumar PM**, **Amortegui JR**, **Osterrieder N**, **Parcells MS**. 2009. A
833 deletion within glycoprotein L of Marek's disease virus (MDV) field isolates correlates
834 with a decrease in bivalent MDV vaccine efficacy in contact-exposed chickens. Avian Dis
835 **53**:287–296.

836 56. **Shaikh SAR**, **Katneni UK**, **Dong H**, **Gaddamanugu S**, **Tavlarides-Hontz P**, **Jarosinski**
837 **KW**, **Osterrieder N**, **Parcells MS**. 2013. A Deletion in the Glycoprotein L (gL) Gene of
838 U.S. Marek's Disease Virus (MDV) Field Strains Is Insufficient to Confer Increased
839 Pathogenicity to the Bacterial Artificial Chromosome (BAC)–Based Strain, RB-1B. Avian
840 Dis **57**:509–518.

841 57. **Gianni T**, **Massaro R**, **Campadelli-Fiume G**. 2015. Dissociation of HSV gL from gH by
842 αvβ6- or αvβ8-integrin promotes gH activation and virus entry. Proc Natl Acad Sci U S A
843 **112**:E3901–E3910.

844 58. **Wu P**, **Reed WM**, **Lee LF**. 2001. Glycoproteins H and L of Marek's disease virus form a
845 hetero-oligomer essential for translocation and cell surface expression. Arch Virol
846 **146**:983–992.

847 59. **Qian Z**, **Brunovskis P**, **Rauscher F**, **Lee L**, **Kung HJ**. 1995. Transactivation activity of
848 Meq, a Marek's disease herpesvirus bZIP protein persistently expressed in latently
849 infected transformed T cells. J Virol **69**:4037–4044.

850 60. **Nair V**, **Kung H-J**. 2004. 4 - Marek's disease virus oncogenicity: Molecular mechanisms,
851 p. 32–48. *In* Davison, F, Nair, V (eds.), Marek's Disease. Academic Press, Oxford.

852 61. **Tian M**, **Zhao Y**, **Lin Y**, **Zou N**, **Liu C**, **Liu P**, **Cao S**, **Wen X**, **Huang Y**. 2011.
853 Comparative analysis of oncogenic genes revealed unique evolutionary features of field
854 Marek's disease virus prevalent in recent years in China. Virol J **8**:121.

855 62. **Murata S**, **Okada T**, **Kano R**, **Hayashi Y**, **Hashiguchi T**, **Onuma M**, **Konnai S**, **Ohashi**
856 **K**. 2011. Analysis of transcriptional activities of the Meq proteins present in highly
857 virulent Marek's disease virus strains, RB1B and Md5. Virus Genes **43**:66–71.

858 63. **Spatz SJ**, **Silva RF**. 2007. Polymorphisms in the repeat long regions of oncogenic and
859 attenuated pathotypes of Marek's disease virus 1. Virus Genes **35**:41–53.

860 64. **Renz KG**, **Cooke J**, **Clarke N**, **Cheetham BF**, **Hussain Z**, **Fakhrul Islam AFM**,
861 **Tannock GA**, **Walkden-Brown SW**. 2012. Pathotyping of Australian isolates of Marek's
862 disease virus and association of pathogenicity with *meq* gene polymorphism. Avian Pathol
863 **41**:161–176.

864 65. **Giacinti C**, **Giordano A**. 2006. RB and cell cycle progression. Oncogene **25**:5220–5227.

865 66. **Renzette N**, **Pokalyuk C**, **Gibson L**, **Bhattacharjee B**, **Schleiss MR**, **Hamprecht K**,
866 **Yamamoto AY**, **Mussi-Pinhata MM**, **Britt WJ**, **Jensen JD**, **Kowalik TF**. 2015. Limits
867 and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci
868 201501880.

869 67. **Afonso CL**, **Tulman ER**, **Lu Z**, **Zsak L**, **Rock DL**, **Kutish GF**. 2001. The Genome of
870 Turkey Herpesvirus. J Virol **75**:971–978.

871 68. **Izumiya Y**, **Jang H-K**, **Ono M**, **Mikami T**. 2001. A Complete Genomic DNA Sequence of
872 Marek's Disease Virus Type 2, Strain HPRS24, p. 191–221. *In* Hirai, PDK (ed.), Marek's
873 Disease. Springer Berlin Heidelberg.

874 69. **Ojkic D**, **Nagy É**. 2000. The complete nucleotide sequence of fowl adenovirus type 8. J
875 Gen Virol **81**:1833–1837.

876 70. **Koppers-Lalic D**, **Verweij MC**, **Lipińska AD**, **Wang Y**, **Quinten E**, **Reits EA**, **Koch J**,
877 **Loch S**, **Rezende MM**, **Daus F**, **Bieńkowska-Szewczyk K**, **Osterrieder N**, **Mettenleiter**
878 **TC**, **Heemskerk MHM**, **Tampé R**, **Neefjes JJ**, **Chowdhury SI**, **Ressing ME**, **Rijsewijk**
879 **FAM**, **Wiertz EJHJ**. 2008. Varicellovirus UL49.5 Proteins Differentially Affect the
880 Function of the Transporter Associated with Antigen Processing, TAP. PLoS Pathog
881 **4**:e1000080.

882 71. **Verweij MC**, **Lipinska AD**, **Koppers-Lalic D**, **van Leeuwen WF**, **Cohen JI**,
883 **Kinchington PR**, **Messaoudi I**, **Bienkowska-Szewczyk K**, **Ressing ME**, **Rijsewijk**
884 **FAM**, **Wiertz EJHJ**. 2011. The Capacity of UL49.5 Proteins To Inhibit TAP Is Widely
885 Distributed among Members of the Genus Varicellovirus. J Virol **85**:2351–2363.

886 72. **Schippers T**, **Jarosinski K**, **Osterrieder N**. 2014. The ORF012 Gene of Marek's Disease
887 Virus Type 1 Produces a Spliced Transcript and Encodes a Novel Nuclear Phosphoprotein
888 Essential for Virus Growth. J Virol **89**:1348–1363.

889 73. **Zuccola HJ**, **Filman DJ**, **Coen DM**, **Hogle JM**. 2000. The crystal structure of an unusual
890 processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate
891 polymerase. Mol Cell **5**:267–78.

892 74. **Wang Y-P**, **Du W-J**, **Huang L-P**, **Wei Y-W**, **Wu H-L**, **Feng L**, **Liu C-M**. 2016. The
893 Pseudorabies Virus DNA Polymerase Accessory Subunit UL42 Directs Nuclear Transport
894 of the Holoenzyme. Front Microbiol **7**.

895 75. **Zhukovskaya NL**, **Guan H**, **Saw YL**, **Nuth M**, **Ricciardi RP**. 2015. The processivity
896 factor complex of feline herpes virus-1 is a new drug target. Antiviral Res **115**:17–20.

897 76. **Digard P**, **Chow CS**, **Pirrit L**, **Coen DM**. 1993. Functional analysis of the herpes simplex
898 virus UL42 protein. J Virol **67**:1159–1168.

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

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

904 79. **Atkins KE**, **Read AF**, **Savill NJ**, **Renz KG**, **Islam AF**, **Walkden-Brown SW**, **Woolhouse**
905 **MEJ**. 2013. Vaccination and reduced cohort duration can drive virulence evolution:
906 Marek's disease virus and industrialized agriculture. Evolution **67**:851–860.

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

910 81. **DeLuca NA**, **Schaffer PA**. 1988. Physical and functional domains of the herpes simplex
911 virus transcriptional regulatory protein ICP4. J Virol **62**:732–743.

912 82. **Wagner LM**, **Lester JT**, **Sivrich FL**, **DeLuca NA**. 2012. The N Terminus and C Terminus
913 of Herpes Simplex Virus 1 ICP4 Cooperate To Activate Viral Gene Expression. J Virol
914 **86**:6862–6874.

915 83. **Arvin AM**, **Gershon AA**. 1996. Live Attenuated Varicella Vaccine. Annu Rev Microbiol
916 **50**:59–100.

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

919 85. **Islam A**, **Harrison B**, **Cheetham BF**, **Mahony TJ**, **Young PL**, **Walkden-Brown SW**.
920 2004. Differential amplification and quantitation of Marek's disease viruses using real-
921 time polymerase chain reaction. J Virol Methods **119**:103–113.

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

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

88. **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.

89. **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.

90. **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.

91. **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.

92. **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.

93. **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.

94. **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**.

95. **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.

96. **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.

97. **Jukes TH**, **Cantor CR**. 1969. Evolution of Protein Molecules (Chapter 24), p. 21–132. *In* Munro, HN (ed.), Mammalian Protein Metabolism. Academic Press.

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

99. **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.

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

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

Line #	Category[a]	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[b]**	1.4×10^{7}	2.5×10^{7}	2.7×10^{7}	3.9×10^{5}	3.4×10^{5}
5	**MDV-specific reads[b]**	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

966 [a]Lines 1-3 refer to sample preparation, lines 4-6 to Illumina MiSeq output, and lines 7-9 to new
967 viral genomes.
968 [b] Sequence read counts in line 4 and 5 are the sum of forward and reverse reads for each sample.
969

970

971 **Table 2. Pair-wise DNA identity and variant proteins between pairs of**

972 **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 DNA-pol[a]	1 in helicase-primase[a]	3 (one each in vLIP, LORF2, UL43)[a]
Farm B-dust vs. Farm A-dust 2	99.87	195	49 (14)	76	66 (1) in DNA-pol[a]	1 in helicase-primase[a]	3 (one each in vLIP, LORF2, UL43)[a]
Same farm, same time: Dust vs. host							
Farm B-dust vs. Farm B-feather 1	99.64	552	476 (11)	6	66 (1) in DNA-pol[a]	1 in helicase-primase[a]	3 (one each in vLIP, LORF2, UL43)[a]
Farm B-dust vs. Farm B-feather 2	99.52	687	572 (19)	45	66 (1) in DNA-pol[a]	1 in helicase-primase [a]	3 (one each in vLIP, LORF2, UL43)[a]
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, same time: one host vs. another							
Farm B-feather 1 vs. Farm B-feather 2	99.38	973	972 (9)	1	0	0	0

973 [a] Abbreviations refer to DNA polymerase processivity subunit protein UL42 (MDV055);
974 helicase-primase subunit UL8 (MDV020); vLIP, lipase homolog (MDV010); LORF2, immune
975 evasion protein (MDV012); UL43 membrane protein (MDV056)

.

976 ## Figure legends

977 **Figure 1. Diagram of samples collected for genome sequencing of field isolates of MDV.**
978 **(A)** Samples collected for genome sequencing were sourced from two Pennsylvania farms with
979 large-scale operations that house approximately 25,000-30,000 individuals per building. These
980 farms were separated by 11 miles. On Farm A, two separate collections of dust were made 11
981 months apart. On Farm B, we collected one dust sample and individual feathers from several
982 hosts, all at a single point in time. In total, three dust collections and two feathers were used to
983 generate five consensus genomes of MDV field isolates (**Table 1**). **(B)** These methods can be
984 used to explore additional aspects of variation in future studies. (Artwork by Nick Sloff, Penn
985 State University, Department of Entomology).
986

987 **Figure 2. The complete MDV-1 genome includes two unique regions and two sets of large**
988 **inverted repeats. (A)** The full structure of the MDV-1 genome includes a unique long region
989 (UL) and a unique short regions (US), each of which are flanked large repeats known as the
990 terminal and internal repeats of the long region (TRL and IRL) and the short region (TRS and
991 IRS). Most ORFs (pale green arrows) are located in the unique regions of the genome. ORFs
992 implicated in MDV pathogenesis are outlined and labeled; these include ICP4 (MDV084 /
993 MDV100), UL36 (MDV049), and Meq (MDV005 / MDV076) (see Results for complete list).
994 **(B)** A trimmed genome format without the terminal repeat regions was used for analyses in order
995 to not over-represent the repeat regions. **(C)** Percent identity from mean pairwise comparison of
996 five consensus genomes, plotted spatially along the length of the genome. Darker colors indicate
997 lower percent identity (see Legend).

.

998

999 **Figure 3: Genome-wide distribution of polymorphic bases within each consensus genome.**
1000 Polymorphic base calls from each MDV genome were grouped in bins of 5 kb and the sum of
1001 polymorphisms in each bin was plotted. Farm B-dust (aqua) contained the largest number of
1002 polymorphic bases, with the majority occurring in the repeat region (IRL/IRS). Farm A-dust 1
1003 (brown) and Farm A-dust 2 (gray) harbored fewer polymorphic bases, with similar distribution to
1004 Farm B-dust. Polymorphic bases detected in feather genomes were more rare, although this
1005 likely reflects their lower coverage depth (see **Table 1**). Note that the upper and lower segments
1006 of the y-axis have different scales; the number of polymorphic bases per genome for the split
1007 column on the right are labeled for clarity.

1008

1009 F**igure 4. Observed vs. expected polymorphism categories for each consensus genome.** Each
1010 consensus genome was analyzed for the presence of polymorphic loci (see Methods for details).
1011 Observed polymorphic loci (solid bars) were categorized as causing synonymous (green) or non-
1012 synonymous (aqua) mutations, or as genic-untranslated (gray) or intergenic (brown). The
1013 expected outcomes (striped bars) for a random distribution of polymorphisms is plotted behind
1014 the observed outcomes (solid bars) for each category. For all genomes there was a significant
1015 difference of the observed-vs.-expected intergenic polymorphisms, relative to those of other
1016 categories.

1017

1018 **Figure 5: A new polymorphic locus in ICP4, and its shifting allele frequency over time. (A)**
1019 HTS data revealed a new polymorphic locus in ICP4 (MDV084) at nucleotide position 5,495. In
1020 the spatially- and temporally-separated dust samples from Farm A (see **Figure 1A** and **Methods**

.

1021 for details), we observed a different prevalence of C (encoding serine) and A (encoding tyrosine)
1022 alleles. **(B)** Using targeted Sanger sequencing of this locus, time-separated dust samples
1023 spanning nine months were Sanger-sequenced to track polymorphism frequency at this locus
1024 over time. The major and minor allele frequencies at this locus varied widely across time, and the
1025 major allele switched from C to A more than twice during this time.
1026
1027 **Figure 6: Dendrogram of genetic distances among all sequenced MDV-1 genomes.** Using a
1028 multiple-genome alignment of all available complete MDV-1 genomes, we calculated the
1029 evolutionary distances between genomes using the Jukes-Cantor model. A dendrogram was then
1030 created using the neighbor-joining method in MEGA with 1000 bootstraps. The five new field-
1031 sampled MDV-1 genomes (green) formed a separate group between the two clusters of USA
1032 isolates (blue). The European vaccine strain (Rispens) formed a separate clade, as did the three
1033 Chinese MDV-1 genomes (aqua). GenBank Accessions for all strains: new genomes, Table 1;
1034 Passage 11-648a, JQ806361; Passage 31-648a, JQ806362; Passage 61-648a, JQ809692; Passage
1035 41-648a, JQ809691; Passage 81-648a, JQ820250; CU-2, EU499381; RB-1B, EF523390; Md11,
1036 170950; Md5, AF243438; Rispens (CVI988), DQ530348; 814, JF742597; GX0101, JX844666;
1037 LMS, JQ314003.
1038

.

1039 ## Supplemental Material Legends and Descriptions

1040 **Supplemental Figure S1. Procedures for enrichment and isolation of MDV DNA from dust**
1041 **or individual feather follicles. (A)** Procedure for enrichment of MDV DNA using dust as the
1042 source of viral DNA. Vortexing, centrifugation and sonication were essential to release cell-
1043 associated virus into the solution. The virus-containing supernatant was then passed through 0.8
1044 μM and 0.22 μM filters for removal of larger contaminants. The flow-thorough was treated with
1045 DNase and the viral particles were captured using 0.1 μM filter. The membrane of the 0.1 μM
1046 filter was then excised and used for extraction of the viral DNA. (**B)** Procedure for enrichment of
1047 MDV DNA using chicken feather follicle as the source of viral DNA. A feather was
1048 mechanically disrupted (bead-beating) and treated with trypsin to break open host cells and
1049 release cell-free virus into the solution. The sample was then treated with DNase to remove
1050 contaminant DNA. Finally, the viral capsids were lysed to obtain viral genomic DNA.
1051

1052 **Supplemental Figure S2. Workflow for computational enrichment for MDV sequences and**
1053 **subsequent viral genome assembly and taxonomic profiling.** The VirGA workflow (39)
1054 requires an input of high-quality HTS data from the viral genome of interest. For this study we
1055 added an additional step that selected MDV-like sequence reads from the milieu of dust and
1056 feather samples. The sequence reads of interest were obtained by using BLAST to compare all
1057 reads against a custom MDV database with an E-value of 10^{-2}; these were then submitted to
1058 VirGA for assembly. Taxonomic profiling followed a similar path using NCBI's all-nucleotide
1059 database to identify the taxonomic kingdom for each sequence read. In this workflow diagram,
1060 parallelograms represent data outputs while rectangles represent computational actions.
1061

.

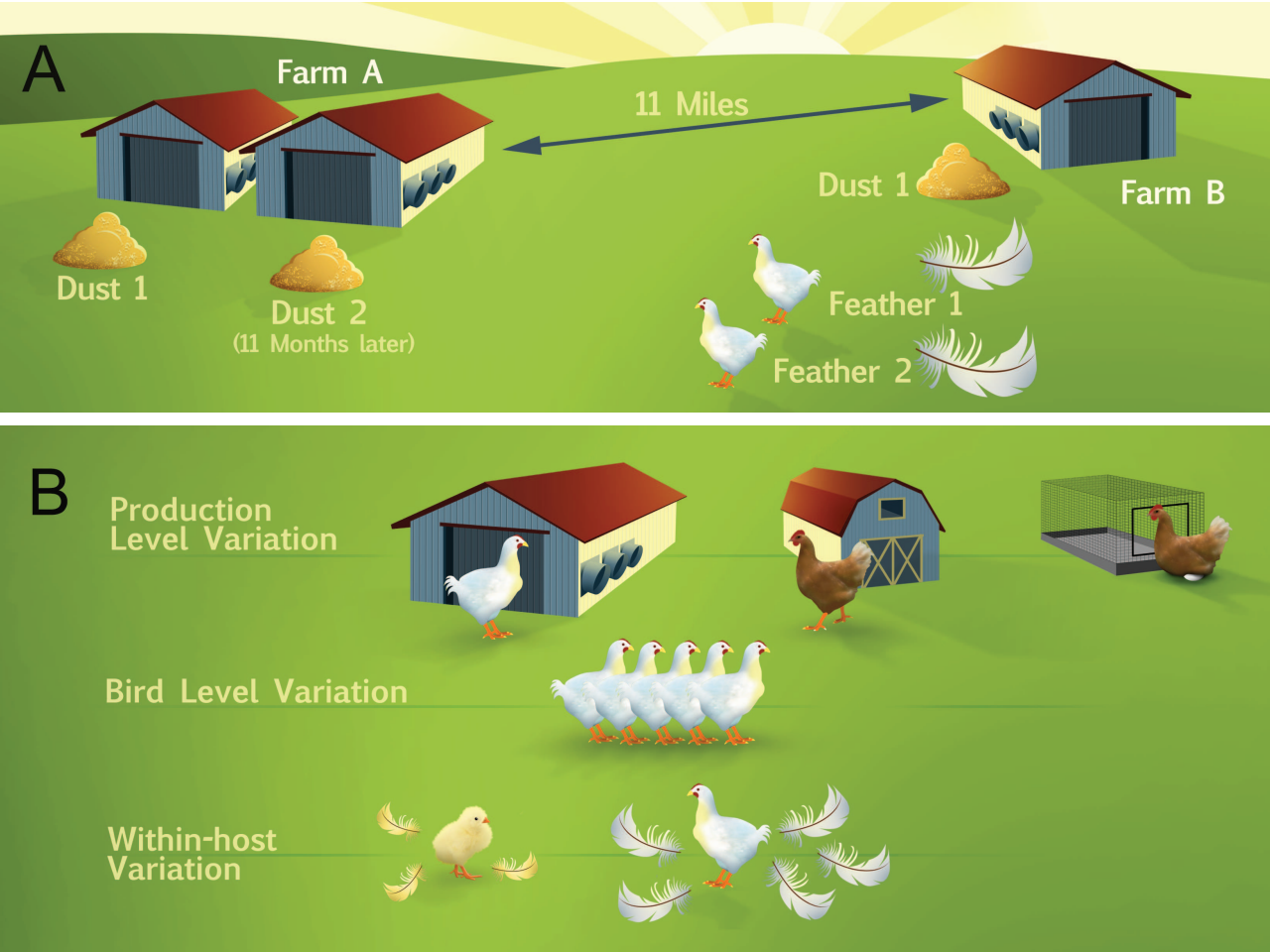
1062 **Supplemental Figure S3: Genome-wide distribution of polymorphisms within each**
1063 **consensus genome, using high-stringency criteria.** Polymorphic base calls from each MDV
1064 genome were grouped by position in bins of 5 kb and the sum of polymorphisms in each bin was
1065 plotted. Stricter parameters of polymorphism detection (see **Methods** for details) revealed a
1066 similar distribution to those in **Figure 3**. No polymorphisms were detected in feather-derived
1067 genomes using high-stringency criteria, due to their lower coverage depth (see **Table** 1). Note
1068 that the upper and lower segments of the y-axis have different scales; the number of polymorphic
1069 bases per segment for the split column on the right are thus labeled on the graph.
1070

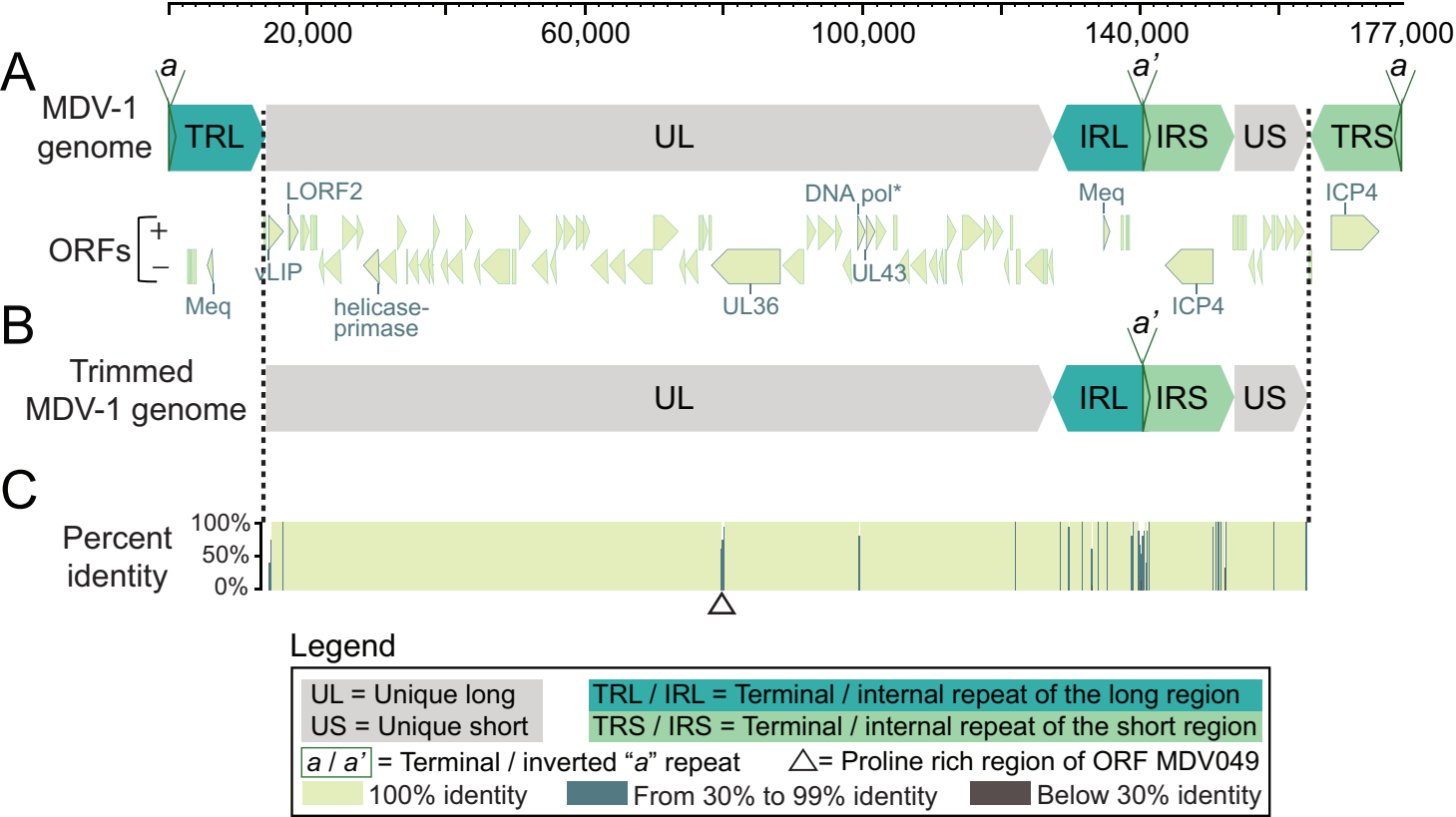
1071 **Supplemental Figure S4. Taxonomic diversity in dust and chicken feathers from Farm B.**
1072 We used an iterative BLASTN workflow to generate taxonomic profiles for all samples from
1073 Farm B (see **Methods** for details). Major categories are shown here, with a full list of taxa (to
1074 family level) in **Supplemental Table S5**. Farm B-feather 1 and Farm B-feather 2 show less
1075 overall diversity, as would be expected from direct host-sampling, vs. the environmental mixture
1076 of the dust samples. Since the viral DNA enrichment procedures remove variable amounts of
1077 host and environmental contaminants, the proportion of taxa present is representative but not
1078 fully descriptive of those present initially. The asterisk indicates sequences that were unclassified
1079 or at low prevalence.
1080

1081 **Supplemental Table S1** [PDF]**:** Yield and percent MDV1+MDV2 and total nanograms of DNA
1082 in each sample for Farm A-dust 1, Farm A-dust 2, and Farm B-dust
1083

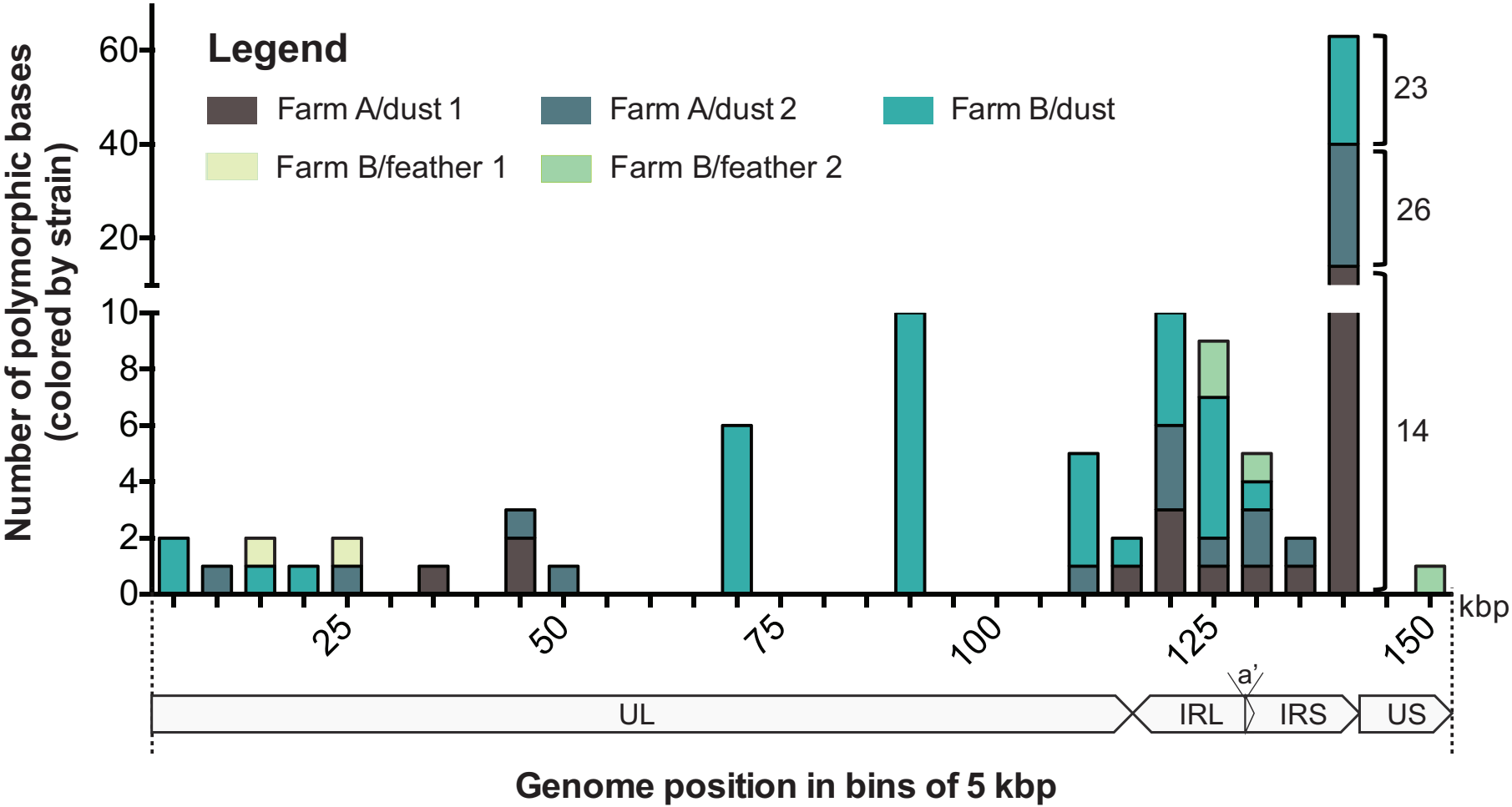
.

1084 **Supplemental Table S2** [PDF]**:** Yield and percent MDV1+MDV2 and total nanograms of DNA
1085 in each sample for Farm B-feathers
1086
1087 **Supplemental Table S3** [Excel]**:** Summary and annotation of all polymorphic loci detected in
1088 MDV-1 consensus genomes.
1089
1090 **Supplemental Table S4** [PDF]**:** Chi-squared values from pairwise comparisons of different
1091 categories of polymorphisms.
1092
1093 **Supplemental Table S5** [Excel]: Summary of taxonomic classification to the family level for all
1094 samples

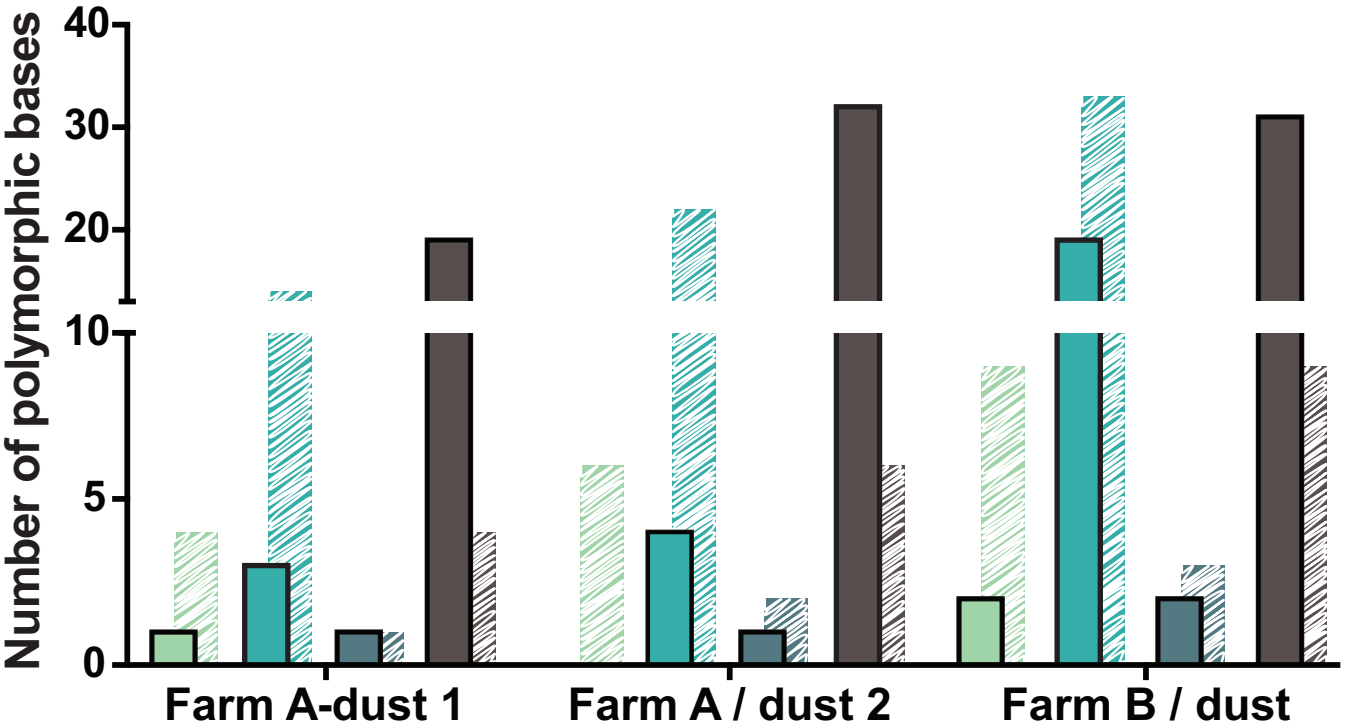

A
Farm A
11 Miles
Farm B
Dust 1
Dust 1
Dust 2
(11 Months later)
Feather 1
Feather 2
B
Production
Level Variation
Bird Level Variation
Within-host
Variation


20,000
60,000
100,000
140,000
177,000
A
MDV-1 genome
a
a'
a
TRL
UL
IRL
IRS
US
TRS
ORFs
+
−
LORF2
vLIP
Meq
helicase-primase
UL36
DNA pol*
UL43
Meq
ICP4
ICP4
B
Trimmed MDV-1 genome
a'
UL
IRL
IRS
US
C
Percent identity
100%
50%
0%
Legend
UL = Unique long
US = Unique short
TRL / IRL = Terminal / internal repeat of the long region
TRS / IRS = Terminal / internal repeat of the short region
a / a' = Terminal / inverted "a" repeat
= Proline rich region of ORF MDV049
100% identity
From 30% to 99% identity
Below 30% identity

Distribution of polymorphic loci in MDV genomes



Number of observed vs. expected polymorphisms in each genome



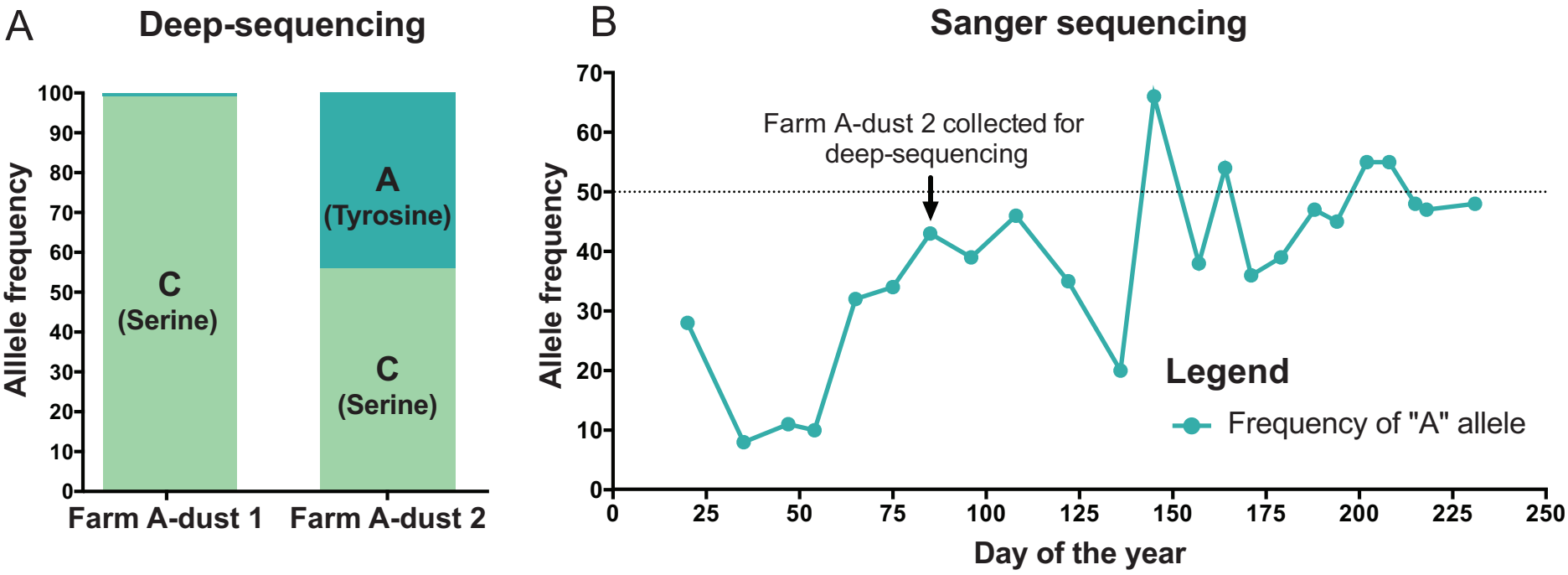
Legend

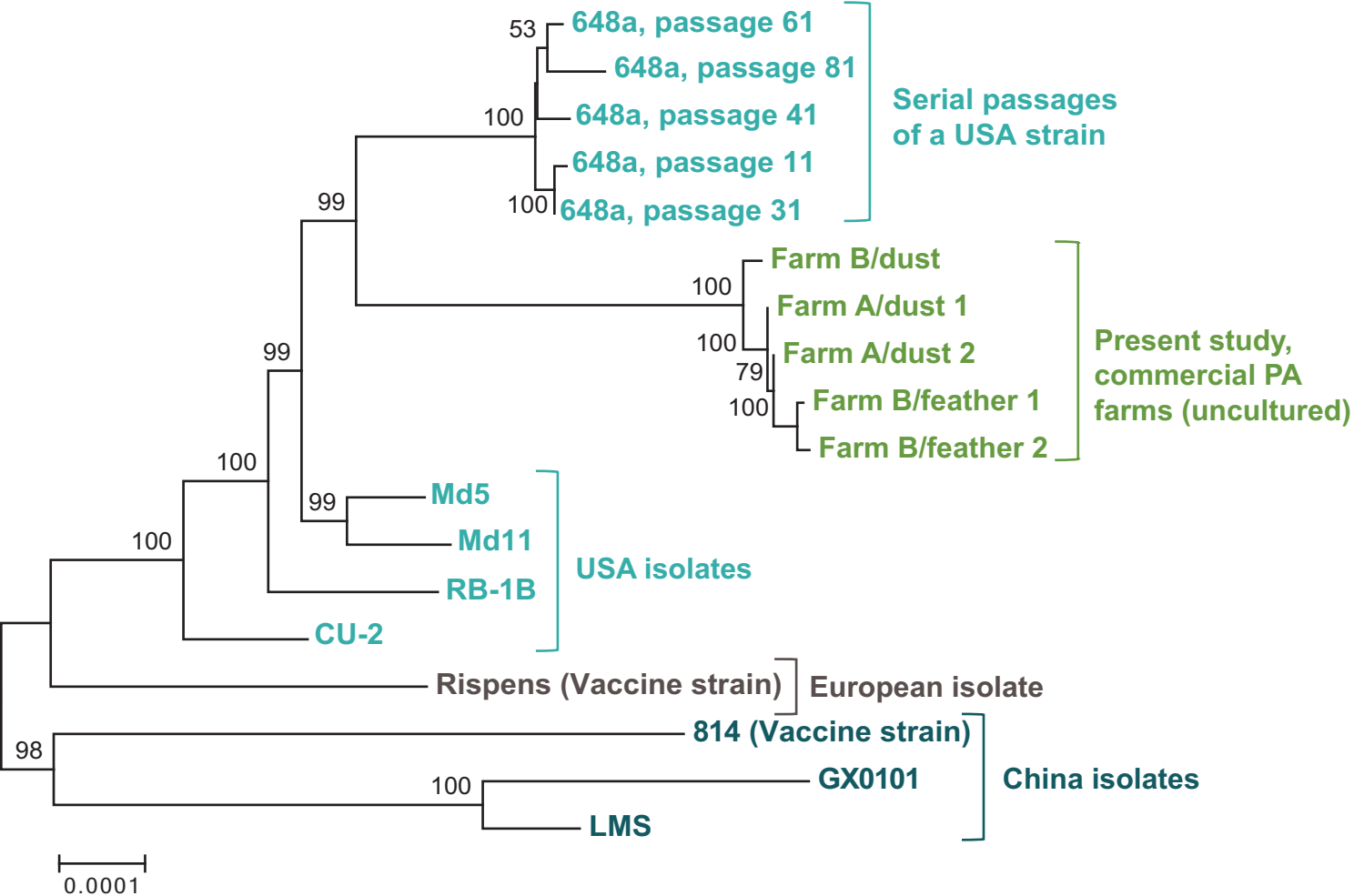
Observed / Expected **synonymous** polymorphisms

Observed / Expected **non-synonymous** polymorphisms

Observed / Expected **genic-untranslated** polymorphisms

Observed / Expected **intergenic** polymorphisms


A
Deep-sequencing
Allele frequency
100
90
80
70
60
50
40
30
20
10
0
C
(Serine)
A
(Tyrosine)
C
(Serine)
Farm A-dust 1
Farm A-dust 2
B
Sanger sequencing
Allele frequency
70
60
50
40
30
20
10
0
Farm A-dust 2 collected for
deep-sequencing
Legend
Frequency of "A" allele
0
25
50
75
100
125
150
175
200
225
250
Day of the year


53
648a, passage 61
648a, passage 81
100
648a, passage 41
648a, passage 11
100
648a, passage 31
Serial passages of a USA strain
99
100
Farm B/dust
Farm A/dust 1
100
Farm A/dust 2
79
100
Farm B/feather 1
Farm B/feather 2
Present study, commercial PA farms (uncultured)
99
100
99
Md5
Md11
100
RB-1B
CU-2
USA isolates
Rispens (Vaccine strain)
European isolate
98
814 (Vaccine strain)
100
GX0101
LMS
China isolates
0.0001

Supplemental Table S1: Yield and percent MDV-1+MDV-2 and total nanograms of DNA in each sample for Farm A-dust 1, Farm A-dust2, and Farm B-dust

Samples	Washes on 0.1 µm filter[a]	%MDV-1	%MDV-2	% MDV1 + MDV2	DNA (ng)
Farm A-dust1					
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	3.3	4.96	6.97
6	**1**	**9.13**	**13.99**	**23.12**	**2.69**
7	**1**	**9.29**	**15.7**	**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	2.9	4.7	17.35
11	0	2.69	5.33	8.02	8.96
12	0	4.49	7.8	12.29	4.14
13	0	1.16	2.49	3.65	20
14	0	1.36	2.83	4.19	19.47
Farm A-dust2					
1	0	1.5	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**	**8.9**	**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
Farm B-dust					
1	0	0.84	6.68	7.52	14.1
2	0	0.46	5.2	5.66	26.64
3	0	0.65	4.85	5.5	19.43
4	0	0.75	5.91	6.66	16.84
5	0	0.23	3.67	3.9	25.9
6	0	0.53	4.65	5.18	23.5
7	**1**	**1.1**	**14.5**	**15.6**	**4.59**
8	**1**	**0.95**	**15.77**	**16.72**	**4.29**
9	**1**	**0.95**	**14.4**	**15.35**	**4.81**
10	**1**	**1.02**	**10.69**	**11.71**	**3.59**

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

Supplemental Table S2: 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

Supplementary Table S3: Summary and annotation of all polymorphic loci detected in MDV-1 consensus genomes

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	Function
Farm A/dust 1	115376	C	A	8.16%	93	42	6	6	50%	50.00%	Intergenic	N/A	N/A
Farm A/dust 1	115377	C	A	29.20%	51	29	21	12	64%	36.36%	Intergenic	N/A	N/A
Farm A/dust 1	137099	A	C	34.74%	42	20	18	15	55%	45.45%	Intergenic	N/A	N/A
Farm A/dust 1	137101	A	C	5.93%	75	36	4	3	57%	42.86%	Intergenic	N/A	N/A
Farm A/dust 1	137264	A	G	5.22%	87	40	5	2	71%	28.57%	Intergenic	N/A	N/A
Farm A/dust 1	138209	C	A	8.27%	88	34	8	3	73%	27.27%	Intergenic	N/A	N/A
Farm A/dust 1	138281	A	C	12.82%	47	21	6	4	60%	40.00%	Intergenic	N/A	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	Gene
Farm A/dust 2	40519	G	A	19.88%	51	78	14	18	44%	56%	Non-synonymous variant	MDV034	gH, glycoprotein H; UL22 homolog; heterodimer with gL; part of fusion/entry complex
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	28%	Non-synonymous variant	MDV040	gB, glycoprotein B; UL27 homolog; part of fusion/entry complex
Farm A/dust 2	116937	G	A	12.67%	86	176	11	27	29%	71%	Intergenic	N/A	N/A
Farm A/dust 2	121872	T	C	36.76%	52	108	35	58	38%	62%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm A/dust 2	130968	G	T	43.48%	77	105	52	88	37%	63%	Non-synonymous variant	MDV084	ICP4 (RS1) homolog; transactivator of gene expression; immediate-early protein
Farm A/dust 2	137156	C	A	5.17%	81	139	4	8	33%	67%	Intergenic	N/A	N/A
Farm A/dust 2	138433	C	A	6.54%	215	85	13	8	62%	38%	Intergenic	N/A	N/A
Farm A/dust 2	138436	C	G	7.12%	221	92	18	6	75%	25%	Intergenic	N/A	N/A
Farm A/dust 2	138437	T	A	8.19%	218	96	20	8	71%	29%	Intergenic	N/A	N/A
Farm A/dust 2	138505	C	A	28.91%	73	136	52	33	61%	39%	Intergenic	N/A	N/A
Farm A/dust 2	138506	A	C	9.15%	117	161	15	13	54%	46%	Intergenic	N/A	N/A
Farm A/dust 2	138593	C	G	6.19%	81	222	10	10	50%	50%	Intergenic	N/A	N/A
Farm A/dust 2	138594	G	A	12.81%	76	203	13	28	32%	68%	Intergenic	N/A	N/A
Farm A/dust 2	138596	T	C	5.35%	81	220	9	8	53%	47%	Intergenic	N/A	N/A
Farm A/dust 2	138599	A	C	5.40%	87	211	5	12	29%	71%	Intergenic	N/A	N/A
Farm A/dust 2	138748	A	G	19.15%	12	64	5	13	28%	72%	Intergenic	N/A	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	Gene
Farm B/dust	2072	T	G	43.64%	90	43	66	37	64%	36%	Non-synonymous variant	MDV010	vLIP; lipase homolog; role in virulence *in vivo; no HSV homolog*
Farm B/dust	15775	C	T	45.76%	94	53	78	46	63%	37%	Synonymous variant	MDV020	DNA helicase-primase subunit; UL8 homolog; role in DNA replication
Farm B/dust	65843	A	G	11.30%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	86626	T	C	40.19%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056	UL43 homolog; probably membrane protein; non-essential *in vitro*
Farm B/dust	108743	T	C	41.74%	173	95	119	73	62%	38%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm B/dust	115231	A	C	21.80%	127	221	36	61	37%	63%	Intergenic	N/A	N/A
Farm B/dust	115232	A	C	14.99%	161	270	19	57	25%	75%	Intergenic	N/A	N/A
Farm B/dust	121656	C	T	37.22%	76	118	43	72	37%	63%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/dust	124841	T	C	41.75%	188	151	130	113	53%	47%	Intergenic	N/A	N/A
Farm B/dust	137449	A	C	45.05%	149	62	103	70	60%	40%	Intergenic	N/A	N/A
Farm B/dust	138199	T	A	5.64%	393	142	24	8	75%	25%	Intergenic	N/A	N/A
Farm B/dust	138267	C	A	37.23%	145	231	81	142	36%	64%	Intergenic	N/A	N/A
Farm B/dust	138268	A	C	7.74%	188	396	25	24	51%	49%	Intergenic	N/A	N/A
Farm B/dust	138355	C	G	5.22%	83	407	9	18	33%	67%	Intergenic	N/A	N/A
Farm B/dust	138356	G	A	12.99%	74	368	15	51	23%	77%	Intergenic	N/A	N/A
Farm B/dust	138361	A	C	5.24%	89	381	6	20	23%	77%	Intergenic	N/A	N/A
Farm B/dust	138510	A	G	27.94%	25	73	10	28	26%	74%	Intergenic	N/A	N/A

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	Gene

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	Function
Farm A/dust 1	30905	G	A	7.61%	53	32	2	5	29%	71%	Synonymous variant	MDV030	capsid protein VP23; UL18 homolog; DNA packaging terminase subunit 1; DNA encapsidation
Farm A/dust 1	40519	G	A	17.78%	56	18	12	4	75%	25%	Non-synonymous variant	MDV034	gH; glycoprotein H; UL22 homolog; heterodimer with gL; part of fusion/entry complex
Farm A/dust 1	43053	G	A	17.24%	48	24	12	3	80%	20%	Genic_UTR	MDV035	UL24 homolog; nuclear protein
Farm A/dust 1	113439	G	T	7.61%	66	19	6	1	86%	14%	Non-synonymous variant	MDV073	pp38; 38 kDa phosphoprotein; role in pathogenesis; necessary for infection of B cells and latency in T cells; no HSV homolog
Farm A/dust 1	115376	C	A	8.16%	93	42	6	6	50%	50%	Intergenic	N/A	N/A
Farm A/dust 1	115377	C	A	29.20%	51	29	21	12	64%	36%	Intergenic	N/A	N/A
Farm A/dust 1	117527	G	T	4.55%	132	57	6	3	67%	33%	Intergenic	N/A	N/A
Farm A/dust 1	121803	C	T	26.11%	92	41	39	8	83%	17%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm A/dust 1	126256	T	G	12.96%	43	4	4	3	57%	43%	Intergenic	N/A	N/A
Farm A/dust 1	132086	T	C	11.97%	108	17	14	3	82%	18%	Non-synonymous variant	MDV084	ICP4 (RS1) homolog; transactivator of gene expression; immediate-early protein
Farm A/dust 1	137099	A	C	34.74%	42	20	18	15	55%	45%	Intergenic	N/A	N/A
Farm A/dust 1	137100	A	C	8.41%	58	40	8	1	89%	11%	Intergenic	N/A	N/A
Farm A/dust 1	137101	A	C	5.93%	75	36	4	3	57%	43%	Intergenic	N/A	N/A
Farm A/dust 1	137264	A	G	5.22%	87	40	5	2	71%	29%	Intergenic	N/A	N/A
Farm A/dust 1	138209	C	A	8.27%	88	34	8	3	73%	27%	Intergenic	N/A	N/A
Farm A/dust 1	138212	C	G	6.29%	93	41	8	1	89%	11%	Intergenic	N/A	N/A
Farm A/dust 1	138213	T	A	7.87%	77	40	9	1	90%	10%	Intergenic	N/A	N/A
Farm A/dust 1	138281	A	C	12.82%	47	21	6	4	60%	40%	Intergenic	N/A	N/A
Farm A/dust 1	138377	G	C	8.33%	60	50	1	9	10%	90%	Intergenic	N/A	N/A
Farm A/dust 1	138379	A	C	9.71%	39	54	1	9	10%	90%	Intergenic	N/A	N/A
Farm A/dust 1	138381	A	T	10.78%	44	47	2	9	18%	82%	Intergenic	N/A	N/A
Farm A/dust 1	138490	A	T	12.64%	56	20	3	8	27%	73%	Intergenic	N/A	N/A
Farm A/dust 1	138492	C	T	10.84%	52	22	2	7	22%	78%	Intergenic	N/A	N/A
Farm A/dust 1	138523	A	G	35.48%	31	9	12	10	55%	45%	Intergenic	N/A	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	Gene
Farm A/dust 2	7251	G	A	5.14%	63	103	1	8	11%	88.89%	Intergenic	N/A	N/A
Farm A/dust 2	22829	A	G	4.35%	105	71	6	2	75%	25.00%	Non-synonymous variant	MDV025	serine/threonine kinase; UL13 homolog
Farm A/dust 2	40519	G	A	19.75%	51	79	14	18	44%	56.25%	Non-synonymous variant	MDV034	gH, glycoprotein H; UL22 homolog; heterodimer with gL; part of fusion/entry complex
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	27.78%	Non-synonymous variant	MDV040	gB, glycoprotein B; UL27 homolog; part of fusion/entry complex
Farm A/dust 2	109048	G	A	5.57%	186	102	2	15	12%	88.24%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm A/dust 2	115422	C	T	2.47%	178	134	1	7	13%	87.50%	Intergenic	N/A	N/A
Farm A/dust 2	115441	A	C	2.20%	129	225	6	2	75%	25.00%	Intergenic	N/A	N/A
Farm A/dust 2	116937	G	A	12.67%	86	176	11	27	29%	71.05%	Intergenic	N/A	N/A
Farm A/dust 2	121872	T	C	36.76%	52	108	35	58	38%	62.37%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm A/dust 2	126692	C	G	26.35%	108	1	33	6	85%	15.38%	Intergenic	N/A	N/A
Farm A/dust 2	126693	C	T	27.21%	105	1	33	7	83%	17.50%	Intergenic	N/A	N/A
Farm A/dust 2	130968	G	T	43.48%	77	105	52	88	37%	62.86%	Non-synonymous variant	MDV084	ICP4 (RS1) homolog; transactivator of gene expression; immediate-early protein
Farm A/dust 2	137156	C	A	5.17%	81	139	4	8	33%	66.67%	Intergenic	N/A	N/A
Farm A/dust 2	137320	T	A	9.94%	120	41	3	15	17%	83.33%	Intergenic	N/A	N/A
Farm A/dust 2	138433	C	A	6.50%	215	85	13	8	62%	38.10%	Intergenic	N/A	N/A
Farm A/dust 2	138434	T	A	4.43%	202	89	7	7	50%	50.00%	Intergenic	N/A	N/A
Farm A/dust 2	138436	C	G	7.06%	221	93	18	6	75%	25.00%	Intergenic	N/A	N/A
Farm A/dust 2	138437	T	A	8.12%	219	96	20	8	71%	28.57%	Intergenic	N/A	N/A
Farm A/dust 2	138451	G	A	2.42%	244	110	4	5	44%	55.56%	Intergenic	N/A	N/A
Farm A/dust 2	138505	C	A	28.26%	75	139	58	33	64%	36.26%	Intergenic	N/A	N/A
Farm A/dust 2	138506	A	C	8.97%	122	162	15	13	54%	46.43%	Intergenic	N/A	N/A
Farm A/dust 2	138593	C	G	6.17%	81	222	10	10	50%	50.00%	Intergenic	N/A	N/A
Farm A/dust 2	138594	G	A	12.81%	76	203	13	28	32%	68.29%	Intergenic	N/A	N/A
Farm A/dust 2	138595	G	C	5.02%	84	219	7	9	44%	56.25%	Intergenic	N/A	N/A
Farm A/dust 2	138596	T	C	5.61%	81	220	10	8	56%	44.44%	Intergenic	N/A	N/A
Farm A/dust 2	138599	A	C	5.38%	87	211	5	12	29%	70.59%	Intergenic	N/A	N/A
Farm A/dust 2	138600	G	C	2.59%	90	210	1	7	13%	87.50%	Intergenic	N/A	N/A
Farm A/dust 2	138600	G	C	2.59%	90	210	1	7	13%	87.50%	Intergenic	N/A	N/A
Farm A/dust 2	138601	G	T	2.97%	81	213	6	3	67%	33.33%	Intergenic	N/A	N/A
Farm A/dust 2	138601	G	T	2.97%	81	213	6	3	67%	33.33%	Intergenic	N/A	N/A
Farm A/dust 2	138602	G	C	4.15%	77	197	3	9	25%	75.00%	Intergenic	N/A	N/A
Farm A/dust 2	138602	G	C	4.15%	77	197	3	9	25%	75.00%	Intergenic	N/A	N/A
Farm A/dust 2	138604	A	C	4.96%	77	191	5	9	36%	64.29%	Intergenic	N/A	N/A
Farm A/dust 2	138604	A	C	4.96%	77	191	5	9	36%	64.29%	Intergenic	N/A	N/A

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	Function
Farm A/dust 2	138606	A	T	4.63%	75	192	3	10	23%	76.92%	Intergenic	N/A	N/A
Farm A/dust 2	138606	A	T	4.63%	75	192	3	10	23%	76.92%	Intergenic	N/A	N/A
Farm A/dust 2	138748	A	G	17.82%	12	65	5	13	28%	72.22%	Intergenic	N/A	N/A
Farm A/dust 2	138748	A	G	17.82%	12	65	5	13	28%	72.22%	Intergenic	N/A	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	Gene
Farm B/dust	2072	T	G	44%	90	43	66	37	64%	36%	Non-synonymous variant	MDV010	vLIP; lipase homolog; role in virulence *in vivo; no HSV homolog*
Farm B/dust	4411	G	T	44%	18	80	14	64	18%	82%	Non-synonymous variant	MDV012	LORF2; TAP transporter blocker; reduces MHCI presentation; no direct HSV homolog
Farm B/dust	13809	G	C	3%	173	133	4	4	50%	50%	Non-synonymous variant	MDV019	virion morphogenesis & egress; UL7 homolog; tegument protein
Farm B/dust	15775	C	T	46%	94	53	78	46	63%	37%	Synonymous variant	MDV020	DNA helicase-primase subunit; UL8 homolog; role in DNA replication
Farm B/dust	65764	G	A	10%	66	93	3	14	18%	82%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	65773	G	T	16%	62	95	4	26	13%	87%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	65796	A	G	35%	53	85	9	66	12%	88%	Synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	65804	A	T	34%	61	89	11	67	14%	86%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	65821	G	T	34%	53	83	8	63	11%	89%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	65843	A	G	11%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (UL36) homolog; ubiquitin specific protease; complexed w/ UL37 tegument protein
Farm B/dust	85939	A	C	95%	0	2	15	21	42%	58%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85954	C	T	29%	14	18	11	2	85%	15%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85959	C	A	36%	22	17	16	6	73%	27%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85961	C	A	39%	28	16	20	8	71%	29%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85962	G	A	40%	25	10	19	4	83%	17%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85963	T	C	56%	7	12	18	6	75%	25%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85966	G	A	9%	46	22	3	4	43%	57%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85971	G	C	11%	47	18	4	4	50%	50%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	85974	G	C	9%	49	21	2	5	29%	71%	Non-synonymous variant	MDV055	DNA polymerase processivity subunit; UL42 homolog; dsDNA binding protein
Farm B/dust	86626	T	C	40%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056	UL43 homolog; probably membrane protein; non-essential *in vitro*
Farm B/dust	108743	T	C	42%	173	95	119	73	62%	38%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm B/dust	108856	G	A	10%	254	164	6	41	13%	87%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm B/dust	108899	A	G	11%	227	192	7	45	13%	87%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm B/dust	109012	T	C	5%	184	176	2	16	11%	89%	Genic_UTR	MDV072	LORF5; function unknown; no HSV homolog
Farm B/dust	114927	T	G	2%	331	276	13	2	87%	13%	Intergenic	N/A	N/A
Farm B/dust	115231	A	C	22%	127	221	36	61	37%	63%	Intergenic	N/A	N/A
Farm B/dust	115232	A	C	15%	161	270	19	57	25%	75%	Intergenic	N/A	N/A
Farm B/dust	115241	C	A	4%	155	360	17	6	74%	26%	Intergenic	N/A	N/A
Farm B/dust	116288	T	A	2%	309	248	3	9	25%	75%	Intergenic	N/A	N/A

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	Function
Farm B/dust	120327	A	G	29%	4	64	3	25	11%	89%	Intergenic	N/A	N/A
Farm B/dust	121181	A	C	2%	203	194	4	6	40%	60%	Non-synonymous variant	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/dust	121656	C	T	37%	76	118	43	72	37%	63%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/dust	122052	A	T	3%	381	231	14	3	82%	18%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/dust	124841	T	C	42%	188	151	130	113	53%	47%	Intergenic	N/A	N/A
Farm B/dust	127347	G	T	42%	35	178	22	130	14%	86%	Intergenic	N/A	N/A
Farm B/dust	137081	A	C	4%	97	66	1	6	14%	86%	Intergenic	N/A	N/A
Farm B/dust	137101	A	G	2%	137	198	7	1	88%	13%	Intergenic	N/A	N/A
Farm B/dust	137102	A	G	4%	119	199	10	2	83%	17%	Intergenic	N/A	N/A
Farm B/dust	137249	A	G	3%	112	193	7	1	88%	13%	Intergenic	N/A	N/A
Farm B/dust	137449	A	C	45%	149	62	103	70	60%	40%	Intergenic	N/A	N/A
Farm B/dust	138195	C	A	4%	402	137	15	5	75%	25%	Intergenic	N/A	N/A
Farm B/dust	138196	T	A	3%	383	138	10	7	59%	41%	Intergenic	N/A	N/A
Farm B/dust	138198	C	G	5%	405	142	18	8	69%	31%	Intergenic	N/A	N/A
Farm B/dust	138199	T	A	6%	393	142	24	8	75%	25%	Intergenic	N/A	N/A
Farm B/dust	138266	C	A	4%	266	386	19	9	68%	32%	Intergenic	N/A	N/A
Farm B/dust	138267	C	A	37%	145	231	81	142	36%	64%	Intergenic	N/A	N/A
Farm B/dust	138268	A	C	8%	188	396	25	24	51%	49%	Intergenic	N/A	N/A
Farm B/dust	138355	C	G	5%	83	407	9	18	33%	67%	Intergenic	N/A	N/A
Farm B/dust	138356	G	A	13%	74	368	15	51	23%	77%	Intergenic	N/A	N/A
Farm B/dust	138357	G	C	4%	83	397	5	14	26%	74%	Intergenic	N/A	N/A
Farm B/dust	138358	T	C	4%	82	399	8	12	40%	60%	Intergenic	N/A	N/A
Farm B/dust	138361	A	C	5%	89	381	6	20	23%	77%	Intergenic	N/A	N/A
Farm B/dust	138362	G	C	3%	95	379	2	11	15%	85%	Intergenic	N/A	N/A
Farm B/dust	138364	G	C	4%	80	355	5	15	25%	75%	Intergenic	N/A	N/A
Farm B/dust	138366	A	C	5%	79	343	8	14	36%	64%	Intergenic	N/A	N/A
Farm B/dust	138368	A	T	3%	79	345	2	13	13%	87%	Intergenic	N/A	N/A
Farm B/dust	138476	A	C	4%	111	79	2	5	29%	71%	Intergenic	N/A	N/A
Farm B/dust	138510	A	G	28%	25	73	10	28	26%	74%	Intergenic	N/A	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	Gene
Farm B/feather 1	12176	C	A	15.56%	17	20	6	1	86%	14%	Non-synonymous variant	MDV018	capsid portal protein; UL6 homolog; DNA encapsidation
Farm B/feather 1	23126	G	T	12.28%	28	21	2	5	29%	71%	Non-synonymous variant	MDV025	serine/threonine kinase; UL13 homolog

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	Gene
Farm B/feather 2	122174	C	T	10.94%	32	25	1	6	14%	85.71%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/feather 2	122204	C	T	15.79%	21	27	2	7	22%	77.78%	Genic_UTR	MDV076	Meq; oncogene; role in tumor formation; no HSV homolog
Farm B/feather 2	128489	G	T	10.13%	38	33	2	6	25%	75.00%	Intergenic	N/A	N/A
Farm B/feather 2	144479	C	A	14.00%	33	10	5	2	71%	28.57%	Non-synonymous variant	MDV092	serine/threonine kinase; US3 homolog

Supplemental Table S4: Chi-squared values from pairwise comparisons of different categories of polymorphisms.

Sample[a]	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)

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

Supplemental table S5: Summary of classification to the family level for all samples

Family Name	Farm B/dust	Farm B/feather 1	Farm B/feather 2	Farm A/dust 1	Farm A/dust 2
Herpesviridae	1,372,838	103,939	165,216	370,393	512,531
Chicken	19,119,306	212,314	121,060	11,783,342	21,215,016
<100	570,555	50,009	27,895	794,467	533,986
Gramineae	48,512	6,145	13,189	17,909	59,709
Meleagrididae	47,292	9,570	6,352	16,493	76,814
Methylobacteriaceae	35,567	2,660	5,205	356	58,844
other sequences	27,430	1,678	3,012	10,400	14,568
Propionibacteriaceae	15,405	0	148	6,197	12,459
Bovidae	1,131,177	0	0	8,055	39,873
Bradyrhizobiaceae	986,082	0	0	682	99,710
Dermabacteraceae	605,631	0	0	75,916	375,290
Staphylococcaceae	337,057	0	0	178,747	240,507
Corynebacteriaceae	229,208	0	0	52,055	129,345
unclassified	196,911	0	0	147,502	109,899
Lactobacillaceae	196,157	0	0	147,351	195,604
Babesiidae	141,039	0	0	3,930	10,966
Bacillaceae	104,791	0	0	77,052	97,930
Vira	102,266	0	0	7,700	31,685
Sphingobacteriaceae	77,220	0	0	36,775	68,077
Bacteroidaceae	72,040	0	0	32,391	64,454
Streptococcaceae	63,088	0	0	23,633	42,208
Actinoplanaceae	55,273	0	0	19,379	52,946
Lachnospiraceae	54,841	0	0	35,767	100,136
Siphoviridae	54,286	0	0	28,164	74,766
Clostridiaceae	53,824	0	0	18,827	43,915
biota	53,379	0	0	20,321	64,013
Peptostreptococcaceae	49,713	0	0	4,585	29,781
Ruminococcaceae	36,850	0	0	22,980	63,335
Enterobacteraceae	35,496	0	0	11,304	72,340
Micrococcaceae	30,584	0	0	13,502	28,318
Nocardiaceae	26,664	0	0	6,869	20,515
Myoviridae	25,744	0	0	26,447	56,563
Actinosynnemataceae	24,644	0	0	7,354	21,047
Enterococcaceae	24,224	0	0	13,055	23,442
Mycobacteriaceae	23,366	0	0	6,151	18,818
Pseudomonadaceae	21,552	0	0	2,478	11,341
Burkholderiaceae	21,347	0	0	0	5,960
Rhizobiaceae	16,069	0	0	103	4,745
Nocardiopsaceae	14,995	0	0	11,804	22,330
Sphingomonadaceae	14,945	0	0	407	15,444
Campylobacter group	13,811	0	0	6,182	5,466
Porphyromonadaceae	13,008	0	0	6,759	25,153
Comamonadaceae	11,483	0	0	623	6,288
Rikenellaceae	11,272	0	0	13,944	61,763
Cellulomonadaceae	10,982	0	0	3,894	9,435
Xanthobacteraceae	10,979	0	0	389	2,367
Rhodospirillaceae	10,498	0	0	499	3,578
Microbacteriaceae	9,978	0	0	3,217	8,655
Cervidae	9,471	0	0	0	0
Lysobacteraceae	9,311	0	0	1,026	5,336
Phyllobacteriaceae	9,070	0	0	217	2,283
Promicromonosporaceae	8,756	0	0	2,975	7,802
Dermacoccaceae	8,713	0	0	2,667	7,232
Bifidobacteriaceae	8,667	0	0	5,241	52,059
Coriobacteriaceae	7,818	0	0	6,884	12,377
Geodermatophilaceae	7,635	0	0	2,153	6,594
Paenibacillaceae	6,921	0	0	4,357	7,627
Rhodobacteraceae	6,764	0	0	331	3,802
Caulobacteraceae	6,531	0	0	685	2,499
Listeriaceae	6,379	0	0	6,555	8,296
Nocardioidaceae	6,376	0	0	2,592	5,286
Gordoniaceae	6,318	0	0	1,779	4,944
Pasteurellaceae	6,071	0	0	2,431	1,853
Eubacteriaceae	5,647	0	0	3,392	8,423
Flavobacteriaceae	5,492	0	0	3,837	4,432
Erysipelothrix group	5,484	0	0	3,451	8,925
Alcaligenaceae	5,223	0	0	561	2,550
Frankiaceae	5,149	0	0	2,071	4,817

Supplemental table S5: Summary of classification to the family level for all samples

Peptococcaceae	4,819	0	0	3,381	6,319
Beutenbergiaceae	4,723	0	0	1,786	4,091
Sanguibacteraceae	4,681	0	0	1,619	3,950
Carnobacteriaceae	4,362	0	0	3,804	4,445
Kineosporiaceae	4,278	0	0	1,351	4,031
Catenulisporaceae	4,007	0	0	973	2,611
Megasphaera group	3,758	0	0	2,066	4,526
Rhodocyclaceae	3,754	0	0	484	2,155
Caryophanaceae	3,378	0	0	2,287	3,492
Oscillospiraceae	3,268	0	0	2,227	5,981
Delphinidae	3,118	0	0	0	0
Intrasporangiaceae	3,105	0	0	822	2,230
Verrucomicrobia subdivision 1	2,960	0	0	134	2,274
Aspergillaceae	2,918	0	0	0	0
Thermoanaerobacterales Family III. Ir	2,695	0	0	1,907	2,999
Acetobacteraceae	2,675	0	0	0	1,164
Leuconostoc group	2,675	0	0	2,382	3,363
Oxalobacteraceae	2,523	0	0	154	1,090
Ancylobacter group	2,446	0	0	0	977
Microsphaeraceae	2,427	0	0	773	2,196
Desulfovibrionaceae	2,370	0	0	605	3,042
Actinomycetaceae	2,328	0	0	1,052	1,947
Tsukamurellaceae	2,252	0	0	585	1,746
Fusobacteriaceae	2,232	0	0	872	1,807
Acinetobacteraceae	2,138	0	0	10,446	4,512
Borrelomycetaceae	2,006	0	0	546	1,011
Cytophaga-Flexibacter group	1,999	0	0	490	1,636
Alcanivorax/Fundibacter group	1,945	0	0	484	1,729
Sapromycetaceae	1,943	0	0	1,103	1,606
Spirochaetaceae	1,866	0	0	1,069	3,796
Muridae	1,847	0	0	1,525	3,953
Ectothiorhodospira group	1,784	0	0	121	1,441
Deinococcaceae	1,768	0	0	733	1,637
Nymphalidae	1,764	0	0	0	112
Glycomycetaceae	1,732	0	0	575	1,515
Thermoanaerobacteraceae	1,659	0	0	850	2,119
Prevotellaceae	1,573	0	0	244	2,229
Aeromonadaceae	1,534	0	0	273	1,471
Anaeromyxobacteraceae	1,496	0	0	326	1,337
Myxococcaceae	1,458	0	0	443	1,417
Microviridae	1,430	0	0	0	758
Geobacteraceae	1,371	0	0	383	1,653
Peptoniphilaceae	1,357	0	0	820	1,317
Chromobacteriaceae	1,305	0	0	398	943
Leptotrichiaceae	1,302	0	0	674	1,496
Sorangiaceae	1,280	0	0	145	943
Chromatiaceae	1,264	0	0	127	963
Rhodobiaceae	1,239	0	0	0	245
Spiroplasmataceae	1,234	0	0	824	1,207
Beijerinckiaceae	1,198	0	0	0	338
Halanaerobiaceae	1,153	0	0	605	936
Suidae	1,106	0	0	294	1,088
Thermaceae	1,089	0	0	220	1,568
Haloarchaeaceae	1,084	0	0	0	215
Sporolactobacillaceae	1,084	0	0	885	1,221
Conexibacteraceae	1,070	0	0	235	832
Brachyspiraceae	1,065	0	0	572	1,003
Acidobacteriaceae	1,001	0	0	0	887
Acidimicrobiaceae	980	0	0	236	478
Cercopithecidae	882	0	0	600	1,146
Hominidae	849	0	0	39,199	3,817
Nolanaceae	847	0	0	4,110	926
Acidaminococcaceae	818	0	0	515	1,206
Thermotogaceae	794	0	0	456	624
Chlorobiacea	774	0	0	0	954
Clostridiales Family XVIII. Incertae Se	771	0	0	375	1,055
Dietziaceae	755	0	0	129	338
Methanobacteriaceae	750	0	0	180	114
Methylocystaceae	739	0	0	0	202
Aerococcaceae	716	0	0	531	631

Supplemental table S5: Summary of classification to the family level for all samples

Jonesiaceae	675	0	0	377	574
Helicobacteraceae	671	0	0	0	393
Clostridiales Family XVII. Incertae Se	664	0	0	313	949
Cyprinidae	650	0	0	108	174
Methanomassiliicoccaceae	632	0	0	435	690
Haliangiaceae	628	0	0	146	435
Clostridiales Family XV. Incertae Sedi	626	0	0	244	812
Eremotheciaceae	622	0	0	146	312
Brucellaceae	614	0	0	105	185
Podoviridae	570	0	0	262	2,117
Giraffidae	569	0	0	0	0
Alteromonadaceae	568	0	0	113	469
Mustelidae	568	0	0	0	0
Rubrobacteraceae	548	0	0	118	556
Leporidae	506	0	0	103	129
Mimiviridae	500	0	0	0	309
Alicyclobacillaceae	498	0	0	122	564
Equidae	494	0	0	0	0
Opitutaceae	492	0	0	109	524
Euphorbiaceae	483	0	0	0	0
Hyphomonadaceae	472	0	0	0	153
Alcanivoracaceae	471	0	0	132	289
Halobacteroidaceae	465	0	0	408	648
Archangiaceae	458	0	0	149	372
Sphaerobacteraceae	458	0	0	0	330
Candidatus Brocadiaceae	457	0	0	249	236
Segniliparaceae	423	0	0	132	332
Schistosomatidae	416	0	0	296	524
Desulfobulbaceae	415	0	0	126	406
Erythrobacteraceae	400	0	0	0	382
Piscirickettsia group	399	0	0	102	248
Pelobacteraceae	396	0	0	249	588
Solibacteraceae	390	0	0	0	286
Dasyuridae	372	0	0	400	358
Heliobacteriaceae	371	0	0	198	502
Acaridae	341	0	0	111	217
Rhodothermaceae	334	0	0	0	722
Iridoviridae	333	0	0	0	0
Cyclobacteriaceae	328	0	0	184	264
Desulfarculaceae	317	0	0	149	440
Trueperaceae	309	0	0	104	200
Dictyoglomaceae	298	0	0	147	166
Phycisphaeraceae	296	0	0	135	370
Desulfobacteraceae	282	0	0	0	343
Vibrionaceae	277	0	0	0	400
Gallionella group	275	0	0	0	311
Hydrogenothermaceae	275	0	0	118	156
Ignavibacteriaceae	268	0	0	0	155
Methylococcaceae	256	0	0	0	296
Shewanellaceae	252	0	0	102	161
Nostocaceae	240	0	0	0	113
Nitrosomonadaceae	229	0	0	104	271
Hydrogenophilaceae	227	0	0	0	209
Orbaceae	223	0	0	158	189
Acidithermaceae	218	0	0	0	188
Caviidae	218	0	0	0	0
Parvularculaceae	218	0	0	0	137
Planctomycetaceae	217	0	0	111	206
Tetrahymenidae	214	0	0	0	0
Natranaerobiaceae	212	0	0	132	209
Nitrospiraceae	211	0	0	0	136
African mole-rats	207	0	0	104	158
Caldilineaceae	207	0	0	100	224
Deferribacteraceae	207	0	0	279	268
Syntrophaceae	205	0	0	115	216
Brevibacteriaceae	200	0	0	0	0
Gemmantimonadaceae	199	0	0	0	139
Desulfomicrobiaceae	194	0	0	103	219
Strongylocentrotidae	186	0	0	0	0
Thermoanaerobacterales Family IV. I	186	0	0	109	222

Supplemental table S5: Summary of classification to the family level for all samples

Fabaceae	184	0	0	0	389
Chitinophagaceae	178	0	0	0	239
Thermodesulfobacteriaceae	170	0	0	154	203
Desulfurobacteriaceae	168	0	0	0	104
Sulfuricellaceae	163	0	0	0	0
Legionellaceae	160	0	0	0	0
Colwelliaceae	159	0	0	0	144
Entomoplasma group	159	0	0	0	152
Syntrophomonadaceae	158	0	0	0	123
Camelidae	154	0	0	0	209
Debaryomycetaceae	152	0	0	174	0
Cryomorphaceae	151	0	0	113	273
Bdellovibrionaceae	145	0	0	0	142
Ferrimonadaceae	145	0	0	0	105
Saprospiraceae	144	0	0	0	166
Roseiflexaceae	140	0	0	0	324
Chrysiogenaceae	139	0	0	0	196
Tetraodontidae	139	0	0	0	0
Chloroflexaceae	133	0	0	0	0
Draconibacteriaceae	130	0	0	0	117
Cneoraceae	127	0	0	200	300
Culicidae	124	0	0	0	131
Flammeovirgaceae	117	0	0	0	0
Hypocreaceae	113	0	0	0	0
Leeaceae	112	0	0	216	218
Rivulariaceae	111	0	0	0	0
Arthrodermataceae	110	0	0	0	0
Desulfurellaceae	107	0	0	0	117
Hahellaceae	107	0	0	0	107
Acidithiobacillaceae	103	0	0	0	247
Noelaerhabdaceae	103	0	0	512	379
Pinaceae	103	0	0	0	0
Sphaeriaceae	103	0	0	0	0
Alligatoridae	0	0	0	0	253
Balaenopteridae	0	0	0	0	754
Chaetomiaceae	0	0	0	241	274
Costariaceae	0	0	0	130	0
Dipodascaceae	0	0	0	220	0
Drosophilidae	0	0	0	0	178
Fibrobacteraceae	0	0	0	0	140
Francisella group	0	0	0	0	146
Hydridae	0	0	0	0	181
Lagriidae	0	0	0	0	109
Loliginidae	0	0	0	0	171
Magnetococcaceae	0	0	0	0	102
Mamiellaceae	0	0	0	0	161
Mycosphaerellaceae	0	0	0	349	1,122
Mycosyringaceae	0	0	0	0	228
Onchocercidae	0	0	0	0	119
Pseudoalteromonadaceae	0	0	0	127	231
Retroviridae	0	0	0	296	174
Rhabditidae	0	0	0	126	0
Sarcocystidae	0	0	0	261	140
Syntrophobacteraceae	0	0	0	0	119
Thermodesulfobiaceae	0	0	0	0	136
	26,519,857	386,315	342,077	14,241,186	25,157,407