DNA from dust: comparative genomics of vaccine-break virus in the field

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: Comparative genomics of vaccine-break MDV-1

Word counts: Abstract, 151; Importance, 132; Intro/Results/Discussion, 3800; Methods, 2956.

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). 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 direct from field samples without the need for sequence-based enrichment strategies prior to sequencing. 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 repeated cycles of replication in the laboratory, raising concerns about the ability of cultured isolates to accurately reflect virus in the field. The ability to directly sequence and compare field isolates of this virus is critical to understanding the genetic basis of rising virulence in the wild. Our approaches remove the prior requirement for cell culture, and allow direct measurement of

viral genomic variation within and between hosts, over time, and during adaptation to changing conditions.

Introduction

Marek's disease virus (MDV), a large DNA alphaherpesvirus of poultry, became increasingly virulent over the second half of the 20th century, evolving from a virus that caused relatively mild disease to one that can kill all unvaccinated hosts in ten days (1–5). Today, mass immunizations with live-attenuated vaccines derived from related herpesviruses control production losses (3, 6), with almost 9 billion broiler chickens vaccinated against MD each year in the US alone (7). MD vaccines prevent host animals from developing disease symptoms, but do not prevent them from becoming infected, nor do they block transmission of the virus (6, 8). Perhaps because of that, those vaccines may have created the conditions favoring the evolutionary emergence of the hyperpathogenic strains which dominate the poultry industry today (5). Certainly, virus evolution undermined two generations of MD vaccines (1–4) However, the genetics underlying MDV-1 evolution into more virulent forms and vaccine breaks are not well understood (9, 10). Likewise, the nature of the vaccine-break lesions that can result from human immunization with live-attenuated varicella zoster virus (VZV) vaccine is an area of active study (11–14).

Remarkably, our understanding of MDV-1 (also known as gallid herpesvirus type 2; GaHV-2) genomics and genetic variation comes exclusively from the study of 10 different laboratory-grown strains (15–22). 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 (23–29). 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 (10, 26, 30, 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 (32–37). 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. 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 known candidates for virulence and modulation of host immunity. These variations were detected both within and between the virus populations at different field sites, and during sequential sampling. One of the new loci potentially associated with virulence, in the viral transactivator ICP4 (MDV084), was tracked using targeted gene surveillance of longitudinal field samples. These findings confirm the genetic flexibility of this large DNA virus in a field setting, and demonstrate how a new

103 combination of HTS and targeted Sanger-based surveillance approaches can be combined to
104 understand viral evolution in the field.
105

106 # Results

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

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

119 To obtain material for genomic surveillance, we isolated MDV nucleocapsids from dust or
120 epithelial tissues from individual feather follicles from selected hosts (see Methods and
121 **Supplemental Figure S2** for overview, and **Supplemental Tables S1-S4** for DNA yields). A
122 total of five uncultured wild-type samples of MDV were sequenced on an in-house Illumina
123 MiSeq sequencer (**Table 1**, lines 4-6; see **Methods** for details). The sequence read data derived
124 from dust contained approximately 2-5% MDV-1 DNA, while the feather samples ranged from

~27%-48% MDV-1 (**Table 1**, line 6). Since dust represents the infectious material that transmits MDV from host to host, and across generations of animals that pass through a farm or field site, we pursued analysis of wild MDV genomes from both types of source material.

Consensus genomes were created for each of the five samples in **Table 1**, using a recently described combination of *de novo* assembly and reference-guided alignment of large sequence blocks, or contigs (38). Like other alphaherpesviruses, MDV genomes have a class E organization (**Figure 2A**) that consists of unique long (UL) and unique short (US) regions (39, 40). Each unique region is flanked by inverted repeats, which are named according to their position (terminal, T, vs. internal, I), and whether they flank the unique long (L) or short (S) region (TRL, IRL, TRS, IRS; see **Figure 2A** for illustration). Nearly complete genomes were obtained for all five samples (**Table 1**). The coverage depth for each genome was directly proportional to the number of MDV-1 specific reads obtained from each sequencing library (**Table 1**, line 5,7). The dust sample from Farm B had the highest coverage depth, at an average of almost 600X across the viral genome. Feather 1 from Farm B had the lowest coverage depth, averaging 44X genome-wide, which still exceeds that of most bacterial or eukaryotic genome assemblies. The genome length for all 5 samples was approximately 180 kilobases (**Table 1**), which is comparable to all previously sequenced MDV-1 isolates (15, 17–22, 41).

For each field sample collected and analyzed here, we assembled a consensus viral genome. We anticipated that the viral DNA present in a single feather follicle might be homotypic, based on similar results found for individual vesicular lesions of the alphaherpesvirus varicella zoster virus (VZV) (11, 33). We further expected that the genomes assembled from a dust sample would

represent a mix of viral genomes, summed over time and space. Viral genomes assembled from dust represent the most common genome sequence, or alleles therein, from all of the circulating MDV-1 on a particular farm. The comparison of consensus genomes provided a view into the amount of sequence variation between Farm A and Farm B, or between two individuals on the same Farm (**Table 2**). In contrast, examining the polymorphic loci within each consensus genome assembly allowed us to observe the level of variation within the viral population at each point source (**Supplemental Table S5**, **Figures 3-4**).

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

We began our assessment of genetic diversity by determining the extent of DNA and amino acid variations between the five different consensus genomes. We found that the five genomes are highly similar to one another at the DNA level, with the percent homology ranging from 99.4% to 99.9% in pairwise comparisons (**Figure 2C**, **Table 2**). These comparisons used a trimmed genome format (**Figure 2B**) where the terminal repeat regions had been removed, so that these sequences were not over-represented in the analyses. The level of identity between samples is akin to that observed in closely related isolates of HSV-1 (38). Observed nucleotide differences were categorized as genic or intergenic, and further sub-divided based on whether the differences were insertions or deletions (INDELs) or single-nucleotide polymorphism (SNPs) **(Table 2)**. The number of nucleotide differences was higher in intergenic regions than in genic regions for all genomes. For the INDEL differences, we also calculated the minimum number of events that could have led to the observed differences, to provide context on the relative frequency of these loci in each genome. We anticipate that these variations include silent mutations, as well as potentially advantageous or deleterious coding differences.

171

172 To understand the effect(s) of these nucleotide variations on protein coding and function, we
173 next compared the amino acid (AA) sequences of all open reading frames (ORFs) for the five
174 isolates. The consensus protein coding sequences of all five isolates were nearly identical, with
175 just a few differences (**Table 2**). In comparison to the other four samples, Farm B-dust harbored
176 AA substitutions in four proteins. A single non-synonymous mutation was seen in each of the
177 following: the virulence-associated lipase homolog vLIP (MDV010; Farm B-dust, S501A) (42),
178 the MHC class I immune evasion protein LORF2 (MDV012; Farm B-dust, L311W) (43), and the
179 probable membrane protein UL43 (MDV056; Farm B-dust, S74L). A single synonymous
180 mutation was observed in the DNA helicase-primase protein UL8 (MDV020). Finally, a 66 AA
181 insertion unique to Farm B-dust was observed in the DNA polymerase processivity subunit
182 protein UL42 (MDV055). We did not observe any coding differences between temporally
183 separated dust isolates from Farm A or between feather isolates from different hosts in Farm B,
184 although both of these comparisons (**Table 2**, bottom) revealed hundreds of noncoding
185 differences. The fact that any coding differences were observed in this small sampling of field
186 viruses suggests that the natural ecology of MDV-1 may include drift and adaptation in protein
187 function, in addition to genetic drift.

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). Drug resistance

and vaccine failure have been attributed to the variation present in viral populations (11, 33, 44). 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 (11, 27, 33–35, 44, 45).

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

To address the potential effect(s) of these polymorphisms on MDV biology, we divided the observed polymorphisms into categories of synonymous, non-synonymous, genic-untranslated, or intergenic (**Supplemental Table S5**). The majority of all polymorphisms were located in intergenic regions (**Supplemental Table S5**). We next investigated whether evidence of selection could be detected from the distribution of polymorphisms in our samples. One way to assess this is to determine whether the relative frequencies of synonymous, non-synonymous, genic-untranslated, and intergenic polymorphisms can be explained by random chance. If the observed frequencies differ from those expected from a random distribution, it would suggest genetic selection. After calculating the expected distribution in each sample (as described in **Methods**), we determined that the distribution of variants differed from that expected by chance 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 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 found in pairwise tests that the number of observed intergenic polymorphisms was significantly higher than the observed values for other categories (**Table S6**). This suggests that the mutations that occurred in the intergenic regions were better tolerated and more likely to be maintained in the genome; *i.e.* that purifying selection was acting on coding regions.

Tracking shifts in polymorphic loci over time

In addition to observing polymorphic SNPs in each sample at a single moment in time, we explored whether any shifts in polymorphic allele frequency were detected in the two sequential dust samples from Farm A. We found one locus in the ICP4 (MDV084) gene (nucleotide position 5,495) that was polymorphic in the Farm A-dust 2 sample, with nearly equal proportions of sequence reads supporting the major allele (C) and the minor allele (A) (**Figure 5A**). In

239 contrast, this locus had been 99% A and only 1% C in Farm A-dust 1 (collected 11 months
240 earlier in another house on the same farm), such that it was not counted as polymorphic in that
241 sample by our parameters (see **Methods** for details). At this polymorphic locus, the nucleotide C
242 encodes a serine, while nucleotide A encodes a tyrosine. The encoded AA lies in the C-terminal
243 domain of ICP4 (AA position 1,832). ICP4 is an important immediate-early protein in all
244 herpesviruses, where it serves as a major regulator of viral transcription (46). The role of ICP4 in
245 MDV pathogenesis is also considered crucial because of its proximity to the latency associated
246 transcripts (LAT) and recently described miRNAs (9).
247
248 Given the very different allele frequencies at this ICP4 locus between two houses on the same
249 farm 11 months apart, we examined dust samples from one of the houses over 9 months with
250 targeted Sanger sequencing of this SNP (**Figure 5B**). We found that this locus was highly
251 polymorphic in time-separated dust samples. The A (Tyrosine) allele rose to almost 50%
252 frequency in the 9 month period. In four of the dust samples, the A (Tyrosine) allele was
253 dominant over the C (serine) allele. This reversible fluctuation in allele frequencies over a short
254 period of time is unprecedented for alphaherpesviruses so far as we know. However, recent
255 studies on human cytomegalovirus (HCMV) have shown that selection can cause viral
256 populations to evolve in short periods of time (34, 35). While this is only one example of a
257 polymorphic locus that shifts in frequency over time, similar approaches could be used at any of
258 the hundreds of polymorphic loci detected here (**Supplemental Table S5**). This combination of
259 deep-sequencing genomic approaches to detect new polymorphic loci, and fast gene-specific
260 surveillance to track changes in SNP frequency over a larger number of samples, illustrate the

power of high-quality full genome sequences from field samples to provide powerful markers for field ecology.

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 (15, 18–22, 41, 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 compared to all previously sequenced isolates (**Figure 6**). This may result from geographic differences as previously seen for HSV-1 and VZV (28, 49–52), or from temporal differences in the time of sample isolations, or from the lack of cell-culture adaptation in these new genomes.

Assessment of taxonomic diversity in dust and chicken feathers

As noted in **Table 1**, only a fraction of the reads obtained from each sequencing library were specific to MDV-1. We analyzed the remaining sequences to gain insight into the taxonomic diversity found in poultry dust and chicken feathers. Since our enrichment for viral capsids removed most host and environmental contaminants, the taxa observed here represent only a fraction of the material present initially. However it provides useful insight into the overall complexity of each sample type. The results of the classification for Farm B- dust, Farm B-feather 1, and Farm B-feather 2 are shown in **Supplemental Figure S4**. We divided the sequence reads by the different kingdoms they represent. Complete lists of taxonomic diversity for all samples to the family level are listed in **Supplemental Table S7**. As expected, the

taxonomic diversity of dust is greater than that of feather samples. The majority of sequences in the dust samples mapped to the chicken genome, and only about 2-5% were MDV specific (see also **Table 1**, line 6). We found that single feathers were a better source of MDV DNA, due to their reduced level of taxonomic diversity and higher percentage of MDV-specific reads (**Table 1**, line 6 and **Supplemental Figure S4**).

Discussion

This study presents the first description of MDV-1 genomes sequenced directly from a field setting. This work builds on recent efforts to sequence VZV and HCMV genomes directly from human clinical samples, but importantly the approaches presented here do not employ either the oligo-enrichment used for VZV or the PCR-amplicon strategy used for HCMV (11, 33, 35, 37, 53). This makes our technique widely accessible and reduces potential methodological bias. It is also more rapid to implement and is applicable to the isolation of unknown large DNA viruses, since it does not rely on sequence-specific enrichment strategies. These five genomes were interrogated at the level of comparing consensus genomes – between-host variation – as well as within each consensus genome – within-host variation. By following up with targeted PCR and Sanger sequencing, we demonstrate that HTS can rapidly empower molecular epidemiological field surveillance of loci undergoing genetic shifts.

Although a limited number of non-synonymous differences were detected between the field samples compared here, it is striking that several of these have been previously demonstrated to

have roles in virulence and immune evasion. The N-glycosylated protein viral lipase (vLIP; MDV010) encodes a 120 kDa protein that is required for lytic virus replication in chickens (4). 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 (43). LORF2 is unique to the non-mammalian *Mardivirus* clade, but its function is analogous to that of the mammalian alphaherpesvirus product UL49.5 (43, 54, 55). The non-synonymous mutations detected in Farm B dust lie in protein residues where their impact is unknown, but the presence of these variations suggests that these targets warrant further study. 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, Meq, pp38, vLIP, LORF2, and others (**Supplemental Table S5**). It is noteworthy that these coding variations are detected despite signs of clearance of polymorphisms from coding regions, as indicated by the higher-than-expected ratios of intergenic to coding polymorphisms in these genomes (Figure 4). Together these findings suggest that MDV exhibits genetic variation and undergoes rapid selection in the field, which may provide a basis for its ability to overcome vaccine-induced host resistance to infection (4, 56–58). Future work will be needed to deduce if the mechanisms and phenotypic outcomes in wild virus reflect those predicted by prior *in vitro* evolution studies (10, 30).

Our comparison of new field-isolated MDV-1 genomes revealed a distinct genetic clustering of these genomes, separate from other previously sequenced MDV-1 genomes (**Figure 6**). This pattern may results from geographic and temporal drift in these strains, or from the wild, virulent nature of these strains vs. the adaptation(s) to tissue culture in all prior MDV-1 genome

sequences. The impact of geography on the genetic relatedness of herpesvirus genomes has been previously shown for related alphaherpesviruses such as VZV and HSV-1 (28, 49–52). Phenomena such as recombination can also have an impact on the clustering pattern of MDV isolates. It is worth noting that the genetic distance dendrogram constructed here included genomes from isolates that were collected over a 40 year span, which introduces the potential for temporal drift (15, 18–22, 41, 47, 48). Agricultural and farming practices have evolved significantly during this time, and we presume that pathogens have kept pace. To truly understand the global diversity of MDV, future studies will need to include the impacts of recombination and polymorphisms within samples, in addition to the overall consensus-genome differences reflected by static genetic distance analyses.

Prior studies have shown that when MDV is passaged for multiple generations in cell-culture, the virus accumulates a series of mutations, including several that affect virulence (30). The same is true for the betaherpesvirus HCMV (26). Extended passage *in vitro* forms the basis of vaccine attenuation strategies, as for the successful vaccine strain (vOka) of the alphaherpesvirus VZV (59). Cultured viruses can undergo bottlenecks during initial adaptation to cell culture, and they may accumulate variations and loss of function mutations by genetic drift or positive selection. The variations and mutations thus accumulated may have little relationship to virulence and the balance of variation and selection in the field. We thus anticipate that these field-isolated viral genomes more accurately reflect the genomes of wild MDV-1 that are circulating in the field. The ability to access and compare virus from virulent infections in the field will enable future analyses of vaccine-break viruses.

Our data and approaches provide powerful new tools to measure viral diversity in field settings, and to track changes in large DNA virus populations over time in hosts and ecosystems. In the case of MDV-1, targeted surveillance based on an initial genomic survey could be used to track viral spread across a geographic area, or between multiple end-users associated with a single parent corporation (**Figure 1B**). Similar approaches could be implemented for public-or animal-health programs, for instance to guide management decisions on how to limit pathogen spread and contain airborne pathogens. The ability to sequence and compare large viral genomes directly from individual hosts and field sites will allow a new level of interrogation of host-virus fitness interactions, which form the basis of host resistance to infection (**Figure 1B**). Finally, the analysis of viral genomes from single feather follicles, as from single VZV vesicles, enables our first insights into naturally-occurring within-host variation during infection and transmission (**Figure 1B**). Evidence from tissue compartmentalization studies in HCMV suggests that viral genomes differ in distinct body niches (35, 37). If the same were true for MDV-1, it is would suggest that prior studies of spleen-derived virus may not fully reflect the virus that is shed from feather-follicle epithelia and transmitted to new hosts.

Materials and Methods

Collection of dust and feathers

Dust samples were collected into 1.5 ml tubes from fan louvers. This location contains less moisture and contaminants than floor-collected samples, and represents a mixture of air-borne virus particles and feather dander. Sequential samples from Farm A (**Table 1**) were collected 11 months apart, from adjacent houses on the same farm (**Figure 1A**). Samples from Farm B (**Table**

1) were collected from a single house, at a single point in time. Feathers were collected just before hosts were transported from the farms for sale, to maximize the potential for infection and high viral titer. At the time of collection the animals were 10-12 weeks old. Ten individuals were chosen randomly throughout the entirety of one house for feather collection. Two feathers from each animal were collected from the axillary track (breast feathers). The distal 0.5-1.0 cm proximal shaft or feather tip, which contains the feather pulp, was snipped into a sterile 1.5 ml micro-tube containing a single sterile 5 mm steel bead (Qiagen). On return to the laboratory, tubes were stored at -80°C until processing. One feather from each animal was tested for the presence and quantity and MDV-1 present (see below for quantitative PCR details). The remaining feather from the two animals with highest apparent MDV-1 titer were used for a more thorough DNA extraction (see below for details) and next-generation sequencing. Animal procedures were approved by the Institutional Animal Care and Use Committee of the Pennsylvania State University (IACUC Protocol#: 46599).

Viral DNA isolation from dust

MDV nucleocapsids were isolated from dust as indicated in **Supplemental Figure S2A**. Dust collected from poultry houses was stored in 50 ml polypropylene Falcon® tubes (Corning) at 4°C until required. 500 mg of dust was suspended in 6.5 ml of 1X phosphate buffered saline (PBS). The mixture was vortexed vigorously until homogenous and centrifuged at 2000 × g for 10 minutes. This supernatant was agitated on ice for 30 sec. using a Sonica Ultrasonic Processor Q125 (probe sonicator with 1/8th inch microtip) set to 20% amplitude. It was then vortexed before being centrifuged for a further 10 minutes at 2000 × g. These steps in the isolation of capsids from the dust were important for the release of cell-associated virus into the supernatant.

Vortexing distributed the dust into solution and aided in the release of virions from the cells. Centrifugation allowed large dust particles to sediment out, leaving a portion of virus in solution. Sonication of the resulting supernatant further aided in release of virus. A second round of centrifugation removed additional particulates from the supernatant. To enrich viral capsids away from the remaining contaminants, we then utilized a series of filtration techniques. The supernatant (approximately 5 ml in volume) was introduced to a 5 ml syringe and pushed through a Corning® surfactant-free cellulose acetate (SFCA) filter (0.8 μM) that had been soaked overnight in fetal bovine serum (FBS), which filtered out dust particles, host cells and bacteria. The flow-through was then passed through a Millipore Express® PLUS Membrane vacuum filter (0.22 μM) and the membrane subsequently washed twice with 2.5 ml of PBS, to remove additional bacteria and fungal spores. The resulting flow-through (approximately 10 ml in volume) was treated with DNase (Sigma) at a concentration of 0.1 mg/ml for 30 minutes at room temperature to remove contaminant DNA. In the absence of DNase treatment we observed a higher yield of viral DNA, but with much lower purity (data not shown). The MDV nucleocapsids present in the DNase-treated solution were captured on a polyethersulfone (PES) membrane (VWR) filter (0.1 μM). This filter membrane trapped the viral nucleocapsids, which are between 0.1-0.2 μm (60). An increased MDV purity, but ultimately reduced total nanograms of DNA yield, may be achieved by washing this membrane once with 2.5 ml PBS (see **Supplemental Tables S1-S3**). In the future, samples with a higher percentage of MDV DNA could be obtained by applying these wash steps to all components of the sample pool. The membrane was then carefully excised using a sterile needle and forceps, and laid – exit side downwards – in a sterile 5 cm diameter plastic petri-dish where it was folded twice lengthwise. The "rolled" membrane was then placed into a 2 ml micro-tube containing 1.8 ml of lysis

418 solution (ATL buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen).
419 Digestion was allowed to proceed at 56°C for 1 hour on an incubating microplate shaker (VWR)
420 set to 1100 rpm. The membrane was then removed, held vertically over a tilted sterile 5 cm
421 diameter plastic petri-dish and washed with a small volume of the lysis solution (from the 2 ml
422 micro-tube). This wash was subsequently returned to the 2 ml micro-tube and the tube replaced
423 on the heated shaker where it was allowed to incubate overnight. The following day, the DNA
424 was isolated as per manufacturer's instructions using the DNeasy® Blood and Tissue kit
425 (Qiagen). DNA was eluted in 200 µl DNase-free water. Ten to fourteen aliquots of 500 mg each
426 were used to obtain sufficient DNA for each dust sample (see **Supplemental Tables S1-S3**).
427 Quantitative PCR was used to assess the copy number of viral genomes in the resulting DNA.
428 Total yield and percent MDV-1 vs. MDV-2 DNA are listed in **Supplemental Tables S1-S3**.
429

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

431 The protocol for extraction of MDV DNA from feather follicles was optimized for the smaller
432 input material and an expectation of higher purity (**Supplemental Figure S2B**). Sequential size
433 filters were not used to filter out contaminants from feather follicles, since these direct host
434 samples have fewer impurities than the environmental samples of dust. However, the feather
435 follicle cells were encased inside the keratinaceous shell of the feather tip, which required
436 disruption to release the cells. Each tube containing a single feather tip and one sterile 5 mm
437 diameter steel bead was allowed to thaw, and then 200 µl of PBS was added and the sample
438 bead-beaten for 30 seconds at 30 Hz using a Tissuelyser (Qiagen) (**Supplemental Figure S2B**).
439 Vigorous bead-beating achieved the desired destruction of the follicle tip. A further 720 µl of
440 PBS and 80 µl of 2.5 mg/ml trypsin (Sigma) were then added (final trypsin concentration: 0.8

441 mg/ml), and the solution was transferred to a new sterile 2 ml micro-tube. Feather follicle cells
442 were digested with trypsin to remove the extracellular matrix and to dissociate cells. Digestion
443 was allowed to proceed for 2 hours at 37°C on an incubating microplate shaker (VWR) set to
444 700 rpm. To release cell-associated virus, the suspension was then sonicated on ice for 30
445 seconds using a Sonica Ultrasonic Processor Q125 (probe sonicator with 1/8th inch microtip) set
446 to 50% amplitude. DNase I was added to a final concentration of 0.1 mg/ml and allowed to
447 digest for 1 hour at room temperature to remove non-encapsidated DNA. An equal volume of
448 lysis solution (ATL buffer and Proteinase K from the DNeasy® Blood and Tissue kit, Qiagen)
449 was added and the sample was incubated over night at 56°C on an Incubating Microplate Shaker
450 (VWR) set to 1100 rpm. The following day, the DNA was isolated as per manufacturer's
451 instructions using the DNeasy® Blood and Tissue kit (Qiagen). While the overall amount of
452 DNA obtained from feather follicles was lower than that obtained from pooled dust samples
453 (**Supplemental Table S4)**, it was of higher purity and was sufficient to generate libraries for
454 sequencing (**Table 1**, lines 1-3).
455

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

457 The total amount of DNA present in the samples was quantified by fluorescence analysis using a
458 Qubit® fluorescence assay (Invitrogen) following the manufacturer's recommended protocol.
459 MDV genome copy numbers were determined using serotype-specific quantitative PCR (qPCR)
460 primers and probes, targeting either the MDV-1 pp38 (MDV073; previously known as
461 LORF14a) gene or MDV-2 (SB-1 strain) DNA polymerase (UL42, MDV055) gene. The MDV-1
462 assay was designed by Sue Baigent: forward primer (Spp38for) 5'-
463 GAGCTAACCGGAGAGGGAGA-3'; reverse primer (Spp38rev) 5'-

CGCATACCGACTTTCGTCAA-3'; probe (MDV-1) 6FAM-CCCACTGTGACAGCC-BHQ1 (S. Baigent, *pers. comm.*). The MDV-2 assay is that of Islam et al. (61), but with a shorter MGB probe (6FAM-GTAATGCACCCGTGAC-MGB) in place of their BHQ-2 probe. Real-time quantitative PCRs were performed on an ABI Prism 7500 Fast System with an initial denaturation of 95°C for 20 seconds followed by 40 cycles of denaturation at 95°C for 3 seconds and annealing and extension at 60°C for 30 seconds. Both assays included 4 µl of DNA in a total PCR reaction volume of 20 µl with 1X PerfeCTa™ qPCR FastMix™ (Quanta Biosciences), forward and reverse primers at 300 nM and TaqMan ® BHQ (MDV-1) or MGB (MDV-2) probes (Sigma and Life Sciences, respectively) at 100 nM and 200 nM, respectively. In addition each qPCR reaction incorporated 2 µl BSA (Sigma). Absolute quantification of genomes was based on a standard curve of serially diluted plasmids cloned from the respective target genes. The absolute quantification obtained was then converted to concentration. Once the concentration of the total DNA, MDV-1, and MDV-2 DNA present in the sample were known, we calculated the percentage of MDV-1 and MDV-2 genomic DNA in the total DNA pool (see **Supplemental Tables S1-S4**).

Illumina next-generation sequencing

Sequencing libraries for each of the isolates were prepared using the Illumina TruSeq Nano DNA Sample Prep Kit, according to the manufacturer's recommended protocol for sequencing of genomic DNA. Genomic DNA inputs used for each sample are listed in **Table 1**. The DNA fragment size selected for library construction was 550 base pairs (bp). All the samples were sequenced on an in-house Illumina MiSeq using version 3 sequencing chemistry to obtain

paired-end sequences of 300 × 300 bp. Base calling and image analysis was performed with the MiSeq Control Software (MCS) version 2.3.0.

Consensus genome assembly

As our samples contained DNA from many more organisms than just MDV, we developed a computational workflow (**Supplemental Figure S2**) to preprocess our data prior to assembly. A local BLAST database was created from every *Gallid herpesvirus* genome available in GenBank. All sequence reads for each sample were then compared to this database using BLASTN (62) with a loose e-value less than or equal to 10^{-2} in order to computationally enrich for sequences related to MDV. These "MDV-like" reads were then processed for downstream genome assembly. The use of bivalent vaccine made it possible for us to readily distinguish sequence 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% (63). 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 (38) 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 (64). We use shortened forms of these names

509 throughout the manuscript (see Table 1 for short names). The full names for all five genomes are
510 as follows: MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 1; MDV-1
511 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm A-dust 2; MDV-1 *Gallus domesticus*-
512 wt/Pennsylvania, USA/2015/Farm B-dust; MDV-1 *Gallus domesticus*-wt/Pennsylvania,
513 USA/2015/Farm B-feather 1; MDV-1 *Gallus domesticus*-wt/Pennsylvania, USA/2015/Farm B-
514 feather 2. GenBank Accessions are listed below and in Table 1. Annotated copies of each
515 genome, in a format compatible with genome- and sequence browsers, are available at the
516 Pennsylvania State University ScholarSphere data repository:
517 https://scholarsphere.psu.edu/collections/1544bp14j.
518

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

520 Clustalw2 (43) was used to construct pairwise global nucleotide alignments between whole
521 genome sequences, and pairwise global amino acid alignments between open reading frames.
522 These alignments were utilized by downstream custom Python scripts to calculate percent
523 identity, protein differences, and variation between samples.
524

525 The proline-rich region of UL36 (MDV049), which contains an extended array of tandem
526 repeats, was removed from all five consensus genomes prior to comparison. The amount of
527 polymorphism seen in this region of UL36 is driven by fluctuations in the length of these tandem
528 repeats, as has been seen in prior studies with other alphaherpesviruses such as HSV, VZV, and
529 pseudorabies virus (PRV) (32,48–50). Since the length of extended arrays of perfect repeats
530 cannot be precisely determined by *de novo* assembly (23, 24, 27, 28), we excluded this region
531 from pairwise comparisons of genome-wide variation. Genome alignments with and without the

532 UL36 region removed are archived at the ScholarSphere site:

533 https://scholarsphere.psu.edu/collections/1544bp14j.

534

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

536 VarScan v2.2.11 (67) was used to detect variants present within each consensus genome. To aid
537 in differentiating true variants from potential sequencing errors (68), two separate variant calling
538 analyses were explored. (11). Our main polymorphism-detection parameters (used in **Figures 3-**
539 **4** and **Supplemental Tables S5-S6**) were as follows: minimum variant allele frequency ≥ 0.02;
540 base call quality ≥ 20; read depth at the position ≥ 10; independent reads supporting minor allele
541 ≥ 2. Directional strand bias ≥ 90% was excluded; a minimum of two reads in opposing directions
542 was required. For comparison and added stringency, we also explored a second set of parameters
543 (used in **Supplemental Figure S3**): minimum variant allele frequency ≥ 0.05; base call quality ≥
544 20; read depth at the position ≥ 100; independent reads supporting minor allele ≥5. Directional
545 strand bias ≥ 80% was excluded. The variants obtained from VarScan were then mapped back to
546 the genome to understand their distribution and mutational impact using SnpEff and SnpSift (69,
547 70). Polymorphisms in the proline-rich region of UL36 were excluded, as noted above.

548

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

550 For each of our five consensus genomes, which each represent a viral population, we classified
551 the polymorphisms detected into categories of synonymous, non-synonymous, genic-
552 untranslated, or intergenic, based on where each polymorphism was positioned in the genome.
553 For these analyses (**Figure 4**), we were only able to include polymorphisms detected in the three

554 dust genomes, since the total number of polymorphisms obtained from feather genomes was too
555 low for chi-square analysis. First, we calculated the total possible number of single nucleotide
556 mutations that could be categorized as synonymous, non-synonymous, genic-untranslated or
557 intergenic. To remove ambiguity when mutations in overlapping genes could be classified as
558 either synonymous or non-synonymous, genes with alternative splice variants or overlapping
559 reading frames were excluded from these analyses. This removed 25 open reading frames
560 (approximately 21% of the genome). These tallies of potential mutational events were used to
561 calculate the expected fraction of mutations in each category. We preformed chi-squared tests on
562 each dataset to assess whether the observed distribution of polymorphisms matched the expected
563 distribution. We also performed a similar analysis in pairwise fashion (**Supplemental Table S6**),
564 to assess whether the fraction of variants differed from what would be expected by random
565 chance. Pairwise combinations included the following: synonymous vs. non-synonymous,
566 synonymous vs. intergenic, synonymous vs. genic-untranslated, non-synonymous vs. intergenic,
567 non-synonymous vs. genic-untranslated, and intergenic vs. genic-untranslated. Statistically
568 significant outcomes would suggest that recent or historical selection differed between those
569 categories of variants.
570

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

572 A potential locus of active selection within the ICP4 (MDV084) gene was detected during deep-
573 sequencing of Farm B-dust. This locus was examined using Sanger sequencing. An
574 approximately 400 bp region of the ICP4 gene was amplified using a Taq PCR Core Kit
575 (Qiagen) and the following primers at 200 nM: forward primer (ICP4selF)
576 5'AACACCTCTTGCCATGGTTC 3'; reverse primer (ICP4selR)

5'GGACCAATCATCCTCTCTGG 3'. Cycling conditions included an initial denaturation of 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 1 minute, with a terminal extension at 72°C for 10 minutes. The total reaction volume of 50 µl included 10 µl of DNA and 4 µl BSA (final concentration 0.8 mg/ml). Amplification products were visualized on a 1.5% agarose gel, the target amplicon excised and then purified using the E.Z.N.A. Gel Extraction Kit (Omega Bio-tek). Sanger sequencing was performed by the Penn State Genomics Core Facility utilizing the same primers as used for DNA amplification. The relative peak height of each base call at the polymorphic position was analyzed using the ab1PeakReporter tool (71).

Genetic distance and dendrogram

Multiple sequence alignments of complete MDV-1 (*Gallid herpesvirus 2*) genomes from GenBank (15, 18–22, 41, 47, 48) and those assembled by our lab were generated using MAFFT (72). The evolutionary distances were computed using the Jukes-Cantor method (73) and the evolutionary history was inferred using the neighbor-joining method (74) in MEGA6 (75), with 1,000 bootstrap replicates (76). 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 (38). Sequence annotation

was performed using a massively iterative all-vs.-all BLASTN (E-value ≤ 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 S7**.

GenBank accession numbers and availability of materials

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

Acknowledgements

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

621

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

26. **Dargan DJ**, **Douglas E**, **Cunningham C**, **Jamieson F**, **Stanton RJ**, **Baluchova K**, **McSharry BP**, **Tomasec P**, **Emery VC**, **Percivalle E**, **Sarasini A**, **Gerna G**, **Wilkinson**

GWG, **Davison AJ**. 2010. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. J Gen Virol **91**:1535–46.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

737 42. **Kamil JP**, **Tischer BK**, **Trapp S**, **Nair VK**, **Osterrieder N**, **Kung H-J**. 2005. vLIP, a
738 Viral Lipase Homologue, Is a Virulence Factor of Marek's Disease Virus. J Virol **79**:6984–
739 6996.

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

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

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

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

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

752 48. **Spatz SJ**, **Petherbridge L**, **Zhao Y**, **Nair V**. 2007. Comparative full-length sequence
753 analysis of oncogenic and vaccine (Rispens) strains of Marek's disease virus. J Gen Virol
754 **88**:1080–1096.

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

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

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

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

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

768 54. **Koppers-Lalic D**, **Verweij MC**, **Lipińska AD**, **Wang Y**, **Quinten E**, **Reits EA**, **Koch J**,
769 **Loch S**, **Rezende MM**, **Daus F**, **Bieńkowska-Szewczyk K**, **Osterrieder N**, **Mettenleiter**
770 **TC**, **Heemskerk MHM**, **Tampé R**, **Neefjes JJ**, **Chowdhury SI**, **Ressing ME**, **Rijsewijk**
771 **FAM**, **Wiertz EJHJ**. 2008. Varicellovirus UL49.5 Proteins Differentially Affect the
772 Function of the Transporter Associated with Antigen Processing, TAP. PLoS Pathog
773 **4**:e1000080.

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

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

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

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

787 59. **Arvin AM**, **Gershon AA**. 1996. Live Attenuated Varicella Vaccine. Annu Rev Microbiol
788 **50**:59–100.

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

791 61. **Islam A**, **Harrison B**, **Cheetham BF**, **Mahony TJ**, **Young PL**, **Walkden-Brown SW**.
792 2004. Differential amplification and quantitation of Marek's disease viruses using real-time
793 polymerase chain reaction. J Virol Methods **119**:103–113.

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

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

64. **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.

65. **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.

66. **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.

67. **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.

68. **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.

69. **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.

70. **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**.

71. **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.

72. **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.

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

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

75. **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.

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

836

837 **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

838 [a]Lines 1-3 refer to sample preparation, lines 4-6 to Illumina MiSeq output, and lines 7-9 to new
839 viral genomes.
840 [b] Sequence read counts in line 4 and 5 are the sum of forward and reverse reads for each sample.
841

842

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

844 **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	3 (one each in vLIP, LORF2, UL43)
Farm B-dust vs. Farm A-dust 2	99.87	195	49 (14)	76	66 (1) in DNA-pol[a]	1 in helicase -primase	3 (one each in vLIP, LORF2, UL43)
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	3 (one each in vLIP, LORF2, UL43)
Farm B-dust vs. Farm B-feather 2	99.52	687	572 (19)	45	66 (1) in DNA-pol[a]	1 in helicase -primase	3 (one each in vLIP, LORF2, UL43)
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

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

845

846

Figures

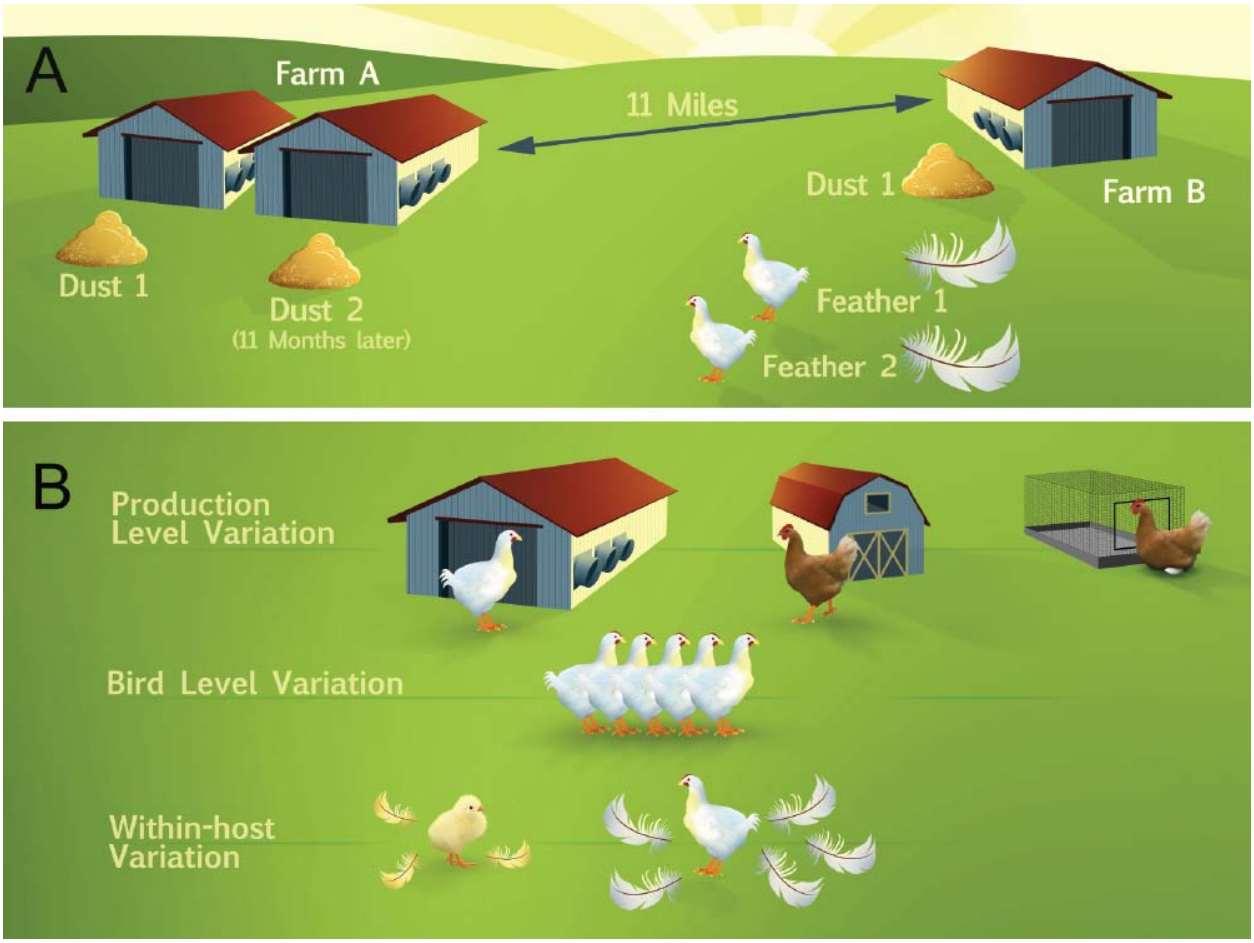


Figure 1. Diagram of samples collected for genome sequencing of field isolates of MDV. **(A)** Samples collected for genome sequencing were sourced from two Pennsylvania farms with large-scale operations that house approximately 25,000-30,000 individuals per building. These farms were separated by 11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

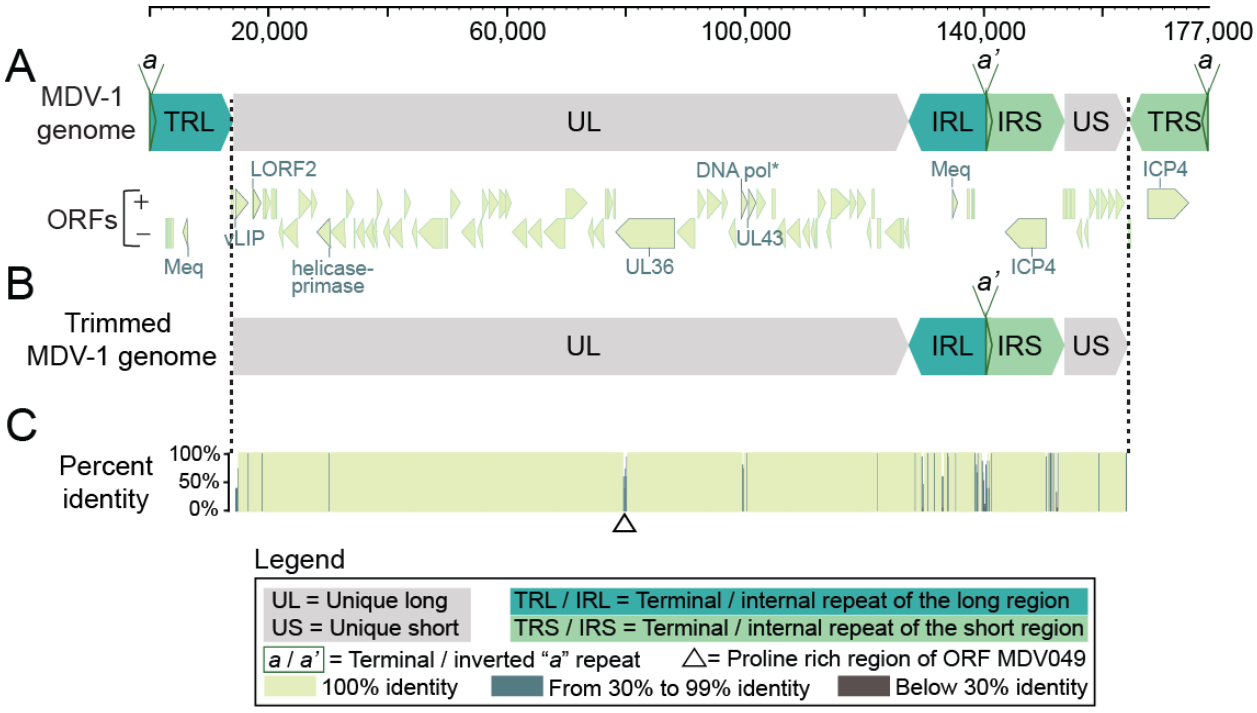
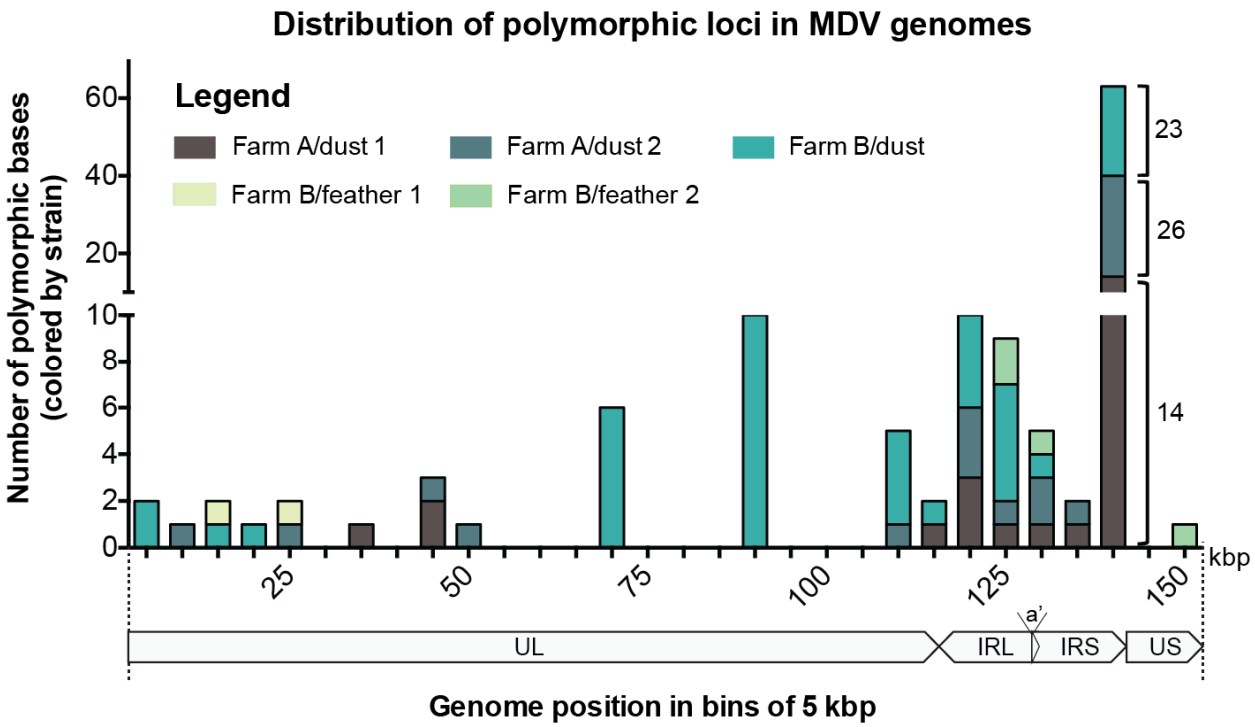


Figure 2. The complete MDV-1 genome includes two unique regions and two sets of large inverted repeats. (A) The full structure of the MDV-1 genome includes a unique long region (UL) and a unique short regions (US), each of which are flanked large repeats known as the terminal and internal repeats of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

874
875



876

877

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

887

888

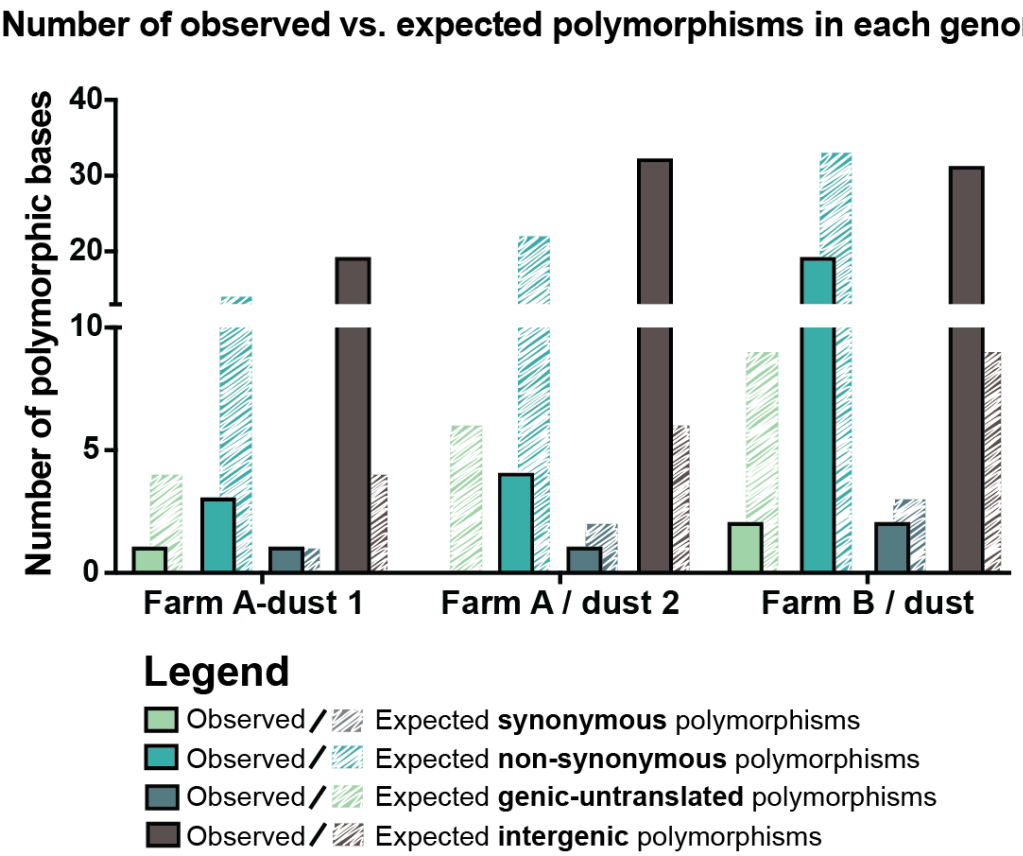


Figure 4. Observed vs. expected polymorphism categories for each consensus genome. Each consensus genome was analyzed for the presence of polymorphic loci (see Methods for details). Observed polymorphic loci (solid bars) were categorized as causing synonymous (green) or non-synonymous (aqua) mutations,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

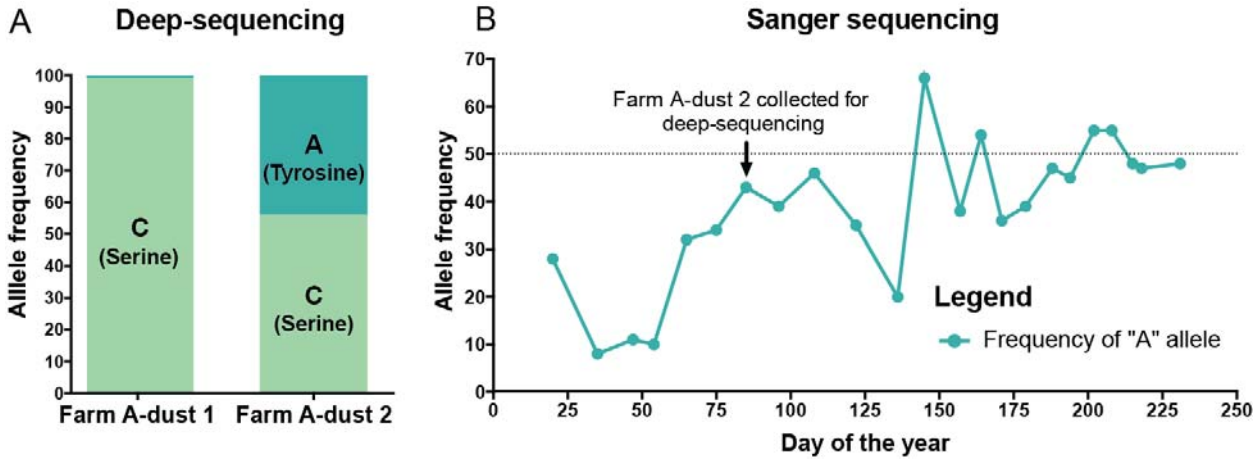


Figure 5: A new polymorphic locus in ICP4, and its shifting allele frequency over time. (A) HTS data revealed a new polymorphic locus in ICP4 (MDV084) at nucleotide position 5,495. In the spatially- and temporally-separated dust samples from Farm A (see **Figure 1A** and **Methods** for details), we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

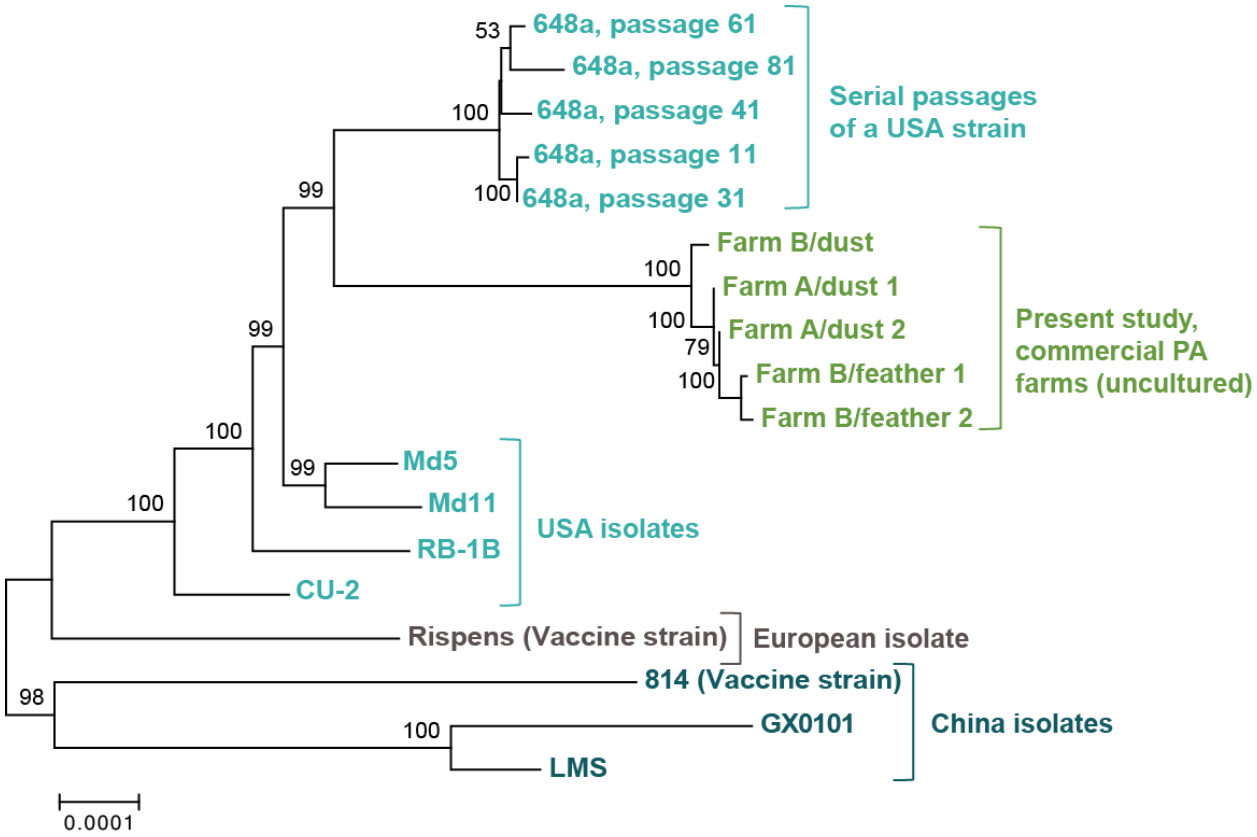


Figure 6: Dendrogram of genetic distances among all sequenced MDV-1 genomes. Using a multiple-genome alignment of all available complete MDV-1 genomes, we calculated the evolutionary distances between genomes using the Jukes-Cantor model. A dendrogram was then created using the neighbor-joining method in MEGA with 1000 bootstraps. The five new field-sampled MDV-1 genomes (green) formed a separate group between the two clusters of USA isolates (blue). The European vaccine strain (Rispens) formed a separate clade, as did the three Chinese MDV-1 genomes (aqua). GenBank Accessions for all strains: new genomes, Table 1; Passage 11-648a, JQ806361; Passage 31-648a, JQ806362; Passage 61-648a, JQ809692; Passage 41-648a, JQ809691; Passage 81-648a, JQ820250; CU-2, EU499381; RB-1B, EF523390; Md11, 170950; Md5, AF243438; Rispens (CVI988), DQ530348; 814, JF742597; GX0101, JX844666; LMS, JQ314003.

Supplemental Material Legends and Descriptions

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

Supplemental Figure S2. Workflow for computational enrichment for MDV sequences and subsequent viral genome assembly and taxonomic profiling. The VirGA workflow (38) requires an input of high-quality HTS data from the viral genome of interest. For this study we added an additional step that selected MDV-like sequence reads from the milieu of dust and feather samples. The sequence reads of interest were obtained by using BLAST to compare all reads against a custom MDV database with an E-value of 10^{-2}; these were then submitted to VirGA for assembly. Taxonomic profiling followed a similar path using NCBI's all-nucleotide database to identify the taxonomic kingdom for each sequence read. In this workflow diagram, parallelograms represent data outputs while rectangles represent computational actions.

949 **Supplemental Figure S3: Genome-wide distribution of polymorphisms within each**
950 **consensus genome, using high-stringency criteria.** Polymorphic base calls from each MDV
951 genome were grouped by position in bins of 5 kb and the sum of polymorphisms in each bin was
952 plotted. Stricter parameters of polymorphism detection (see **Methods** for details) revealed a
953 similar distribution to those in **Figure 3**. No polymorphisms were detected in feather-derived
954 genomes using high-stringency criteria, due to their lower coverage depth (see **Table** 1). Note
955 that the upper and lower segments of the y-axis have different scales; the number of polymorphic
956 bases per segment for the split column on the right are thus labeled on the graph.
957
958 **Supplemental Figure S4. Taxonomic diversity in dust and chicken feathers from Farm B.**
959 We used an iterative BLASTN workflow to generate taxonomic profiles for all samples from
960 Farm B (see **Methods** for details). Major categories are shown here, with a full list of taxa (to
961 family level) in **Supplemental Table S7**. Farm B-feather 1 and Farm B-feather 2 show less
962 overall diversity, as would be expected from direct host-sampling, vs. the environmental mixture
963 of the dust samples. Since the viral DNA enrichment procedures remove variable amounts of
964 host and environmental contaminants, the proportion of taxa present is representative but not
965 fully descriptive of those present initially.
966
967 **Supplemental Table S1:** Yield and percent MDV1+MDV2 and total nanograms of DNA in
968 each sample for Farm A-dust 1
969
970 **Supplemental Table S2:** Yield and percent MDV1+MDV2 and total nanograms of DNA in
971 each sample for Farm A-dust 2

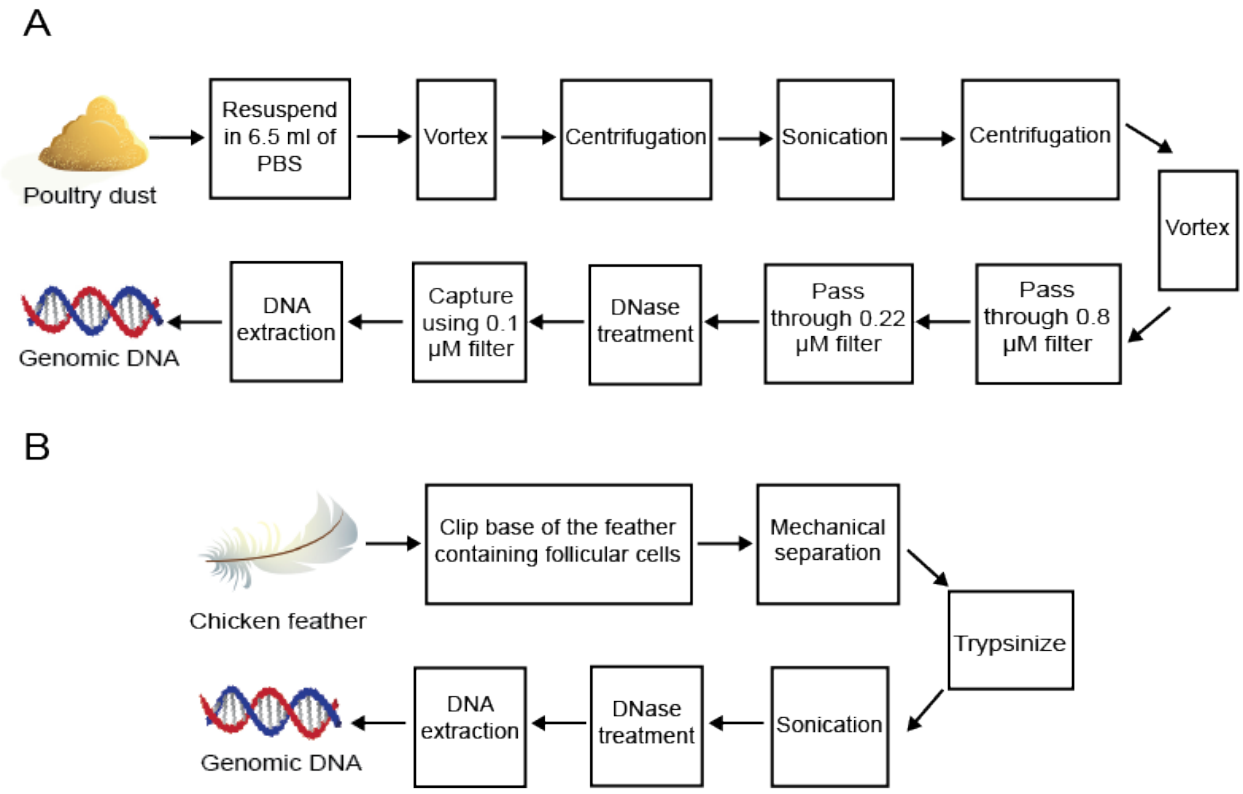
Supplemental Table S3: Yield and percent MDV1+MDV2 and total nanograms of DNA in each sample for Farm B-dust

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

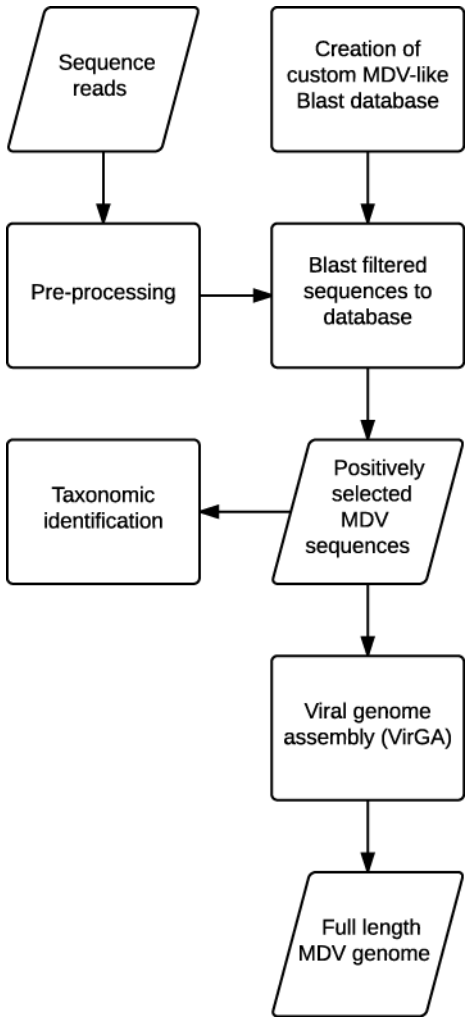
Supplemental Table S5: Summary and annotation of all polymorphic loci detected in MDV1 consensus genomes.

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

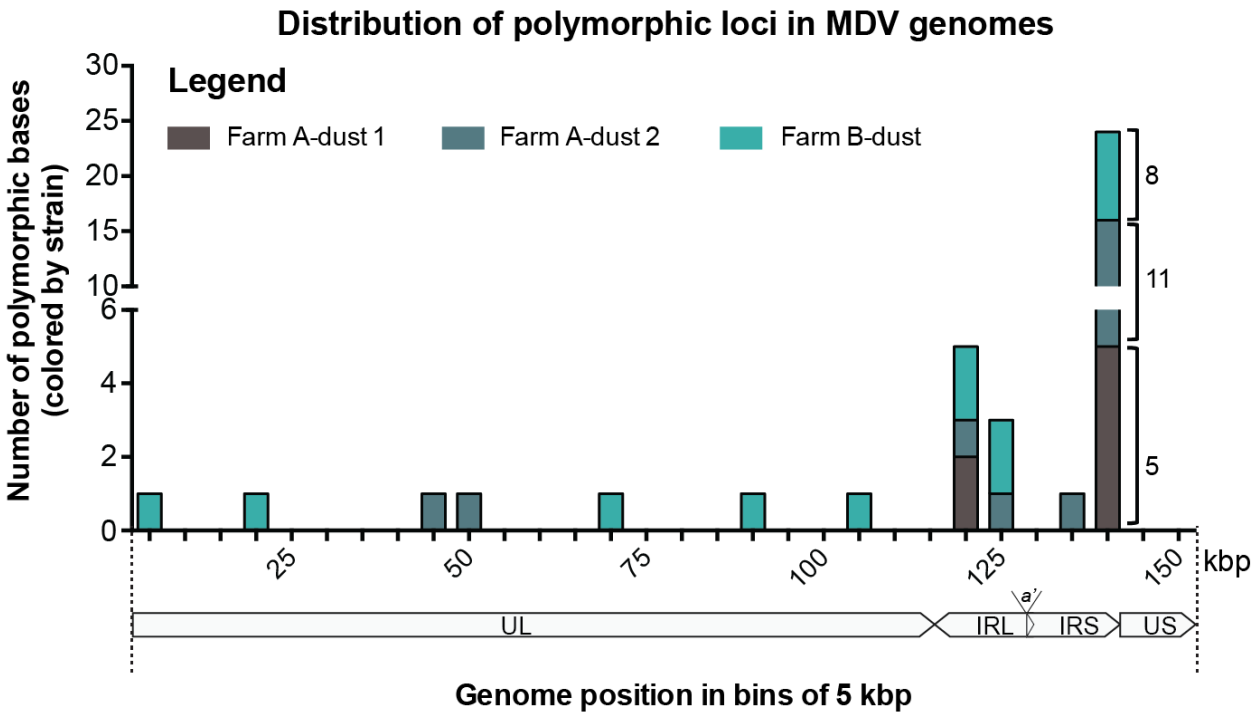
Supplemental Figures



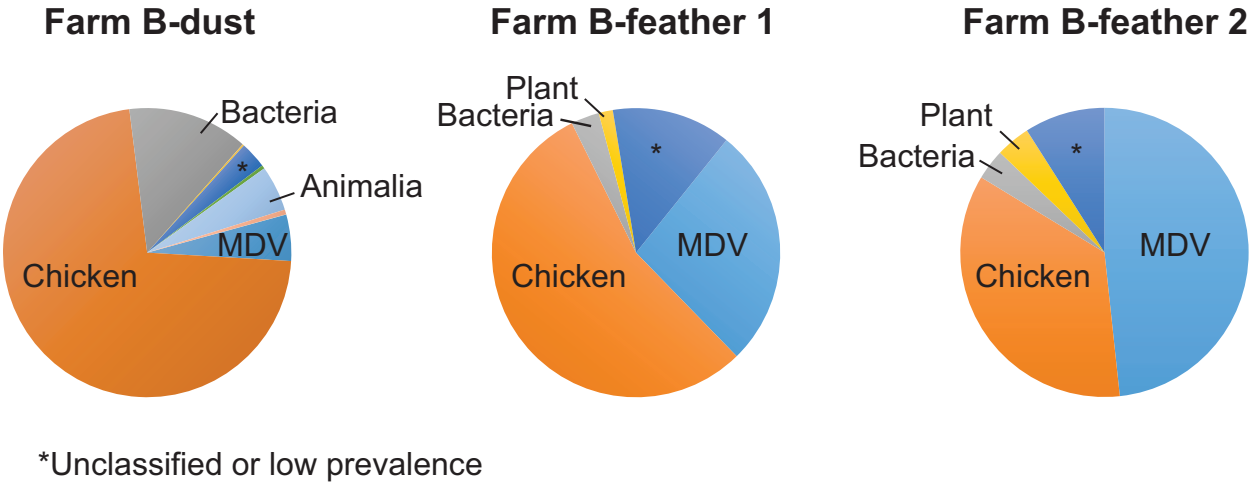
Supplemental Figure S1. Procedures for enrichment and isolation of MDV DNA from dust or individual feather follicles. (A) Procedure for enrichment of MDV DNA using dust as the source of viral DNA. Vortexing, centrifugation and sonication were essential to release cell-associated virus into the solution.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



Supplemental Figure S2. Workflow for computational enrichment for MDV sequences and subsequent viral genome assembly and taxonomic profiling. The VirGA workflow (38) requires an input of high-quality HTS data from the viral genome of interest. For this study we added an additional step that selected MDV-like sequence reads from the milieu of dust and feather samples. The sequence reads of interest were obtained by using BLAST to compare all reads against a custom MDV database with an E-value of 10^{-2}; these were then submitted to VirGA for assembly. Taxonomic profiling followed a similar path using NCBI's all-nucleotide database to identify the taxonomic kingdom for each sequence read. In this workflow diagram, parallelograms represent data outputs while rectangles represent computational actions.



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



Supplemental Figure S4. Taxonomic diversity in dust and chicken feathers from Farm B. We used an iterative BLASTN workflow to generate taxonomic profiles for all samples from Farm B (see **Methods** for details). Major categories are shown here, with a full list of taxa (to family level) in **Supplemental Table S7**. Farm B-feather 1 and Farm B-feather 2 show less overall diversity, as would be expected from direct host-sampling, vs. the environmental mixture of the dust samples. Since the viral DNA enrichment procedures remove variable amounts of host and environmental contaminants, the proportion of taxa present is representative but not fully descriptive of those present initially.

Supplemental Tables

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

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

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

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

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

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

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

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

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

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

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

Supplemental Table S5. Summary and annotation of all polymorphic loci detected in MDV1 consensus genomes.
DNA from dust: comparative genomics of large DNA viruses in field surveillance samples

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 homol
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	28%	Non-synonymous variant	MDV040	gB, glycoprotein B; UL27 homol
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 fo
Farm A/dust 2	130968	G	T	43.48%	77	105	52	88	37%	63%	Non-synonymous variant	MDV084	ICP4 (RS1) homolog; transactiv
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 vir
Farm B/dust	15775	C	T	45.76%	94	53	78	46	63%	37%	Synonymous variant	MDV020	DNA helicase-primase subunit;
Farm B/dust	65843	A	G	11.30%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (
Farm B/dust	86626	T	C	40.19%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056	UL43 homolog; probably membr
Farm B/dust	108743	T	C	41.74%	173	95	119	73	62%	38%	Genic_UTR	MDV072	LORF5; function unknown; no H
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 fo
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
Farm A/dust 1	30905	G	A	7.61%	53	32	2	5	29%	71%	Synonymous variant	MDV030	capsid protein VP23; UL18 hom
Farm A/dust 1	40519	G	A	17.78%	56	18	12	4	75%	25%	Non-synonymous variant	MDV034	gH; glycoprotein H; UL22 homol
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; r
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 fo
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; transactiv
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

Supplemental Table S5. Summary and annotation of all polymorphic loci detected in MDV1 consensus genomes.
DNA from dust: comparative genomics of large DNA viruses in field surveillance samples

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 h
Farm A/dust 2	40519	G	A	19.75%	51	79	14	18	44%	56.25%	Non-synonymous variant	MDV034	gH, glycoprotein H; UL22 homol
Farm A/dust 2	48554	C	T	9.68%	114	54	13	5	72%	27.78%	Non-synonymous variant	MDV040	gB, glycoprotein B; UL27 homol
Farm A/dust 2	109048	G	A	5.57%	186	102	2	15	12%	88.24%	Genic_UTR	MDV072	LORF5; function unknown; no H
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 fo
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; transactiv
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
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 vir
Farm B/dust	4411	G	T	44%	18	80	14	64	18%	82%	Non-synonymous variant	MDV012	LORF2; TAP transporter blocke
Farm B/dust	13809	G	C	3%	173	133	4	4	50%	50%	Non-synonymous variant	MDV019	virion morphogenesis & egress;
Farm B/dust	15775	C	T	46%	94	53	78	46	63%	37%	Synonymous variant	MDV020	DNA helicase-primase subunit;
Farm B/dust	65764	G	A	10%	66	93	3	14	18%	82%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	65773	G	T	16%	62	95	4	26	13%	87%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	65796	A	G	35%	53	85	9	66	12%	88%	Synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	65804	A	T	34%	61	89	11	67	14%	86%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	65821	G	T	34%	53	83	8	63	11%	89%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	65843	A	G	11%	39	118	5	15	25%	75%	Non-synonymous variant	MDV049	large tegument protein; VP1/2 (I
Farm B/dust	85939	A	C	95%	0	2	15	21	42%	58%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85954	C	T	29%	14	18	11	2	85%	15%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85959	C	A	36%	22	17	16	6	73%	27%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85961	C	A	39%	28	16	20	8	71%	29%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85962	G	A	40%	25	10	19	4	83%	17%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85963	T	C	56%	7	12	18	6	75%	25%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85966	G	A	9%	46	22	3	4	43%	57%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85971	G	C	11%	47	18	4	4	50%	50%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	85974	G	C	9%	49	21	2	5	29%	71%	Non-synonymous variant	MDV055	DNA polymerase processivity su
Farm B/dust	86626	T	C	40%	114	78	78	51	60%	40%	Non-synonymous variant	MDV056	UL43 homolog; probably membr
Farm B/dust	108743	T	C	42%	173	95	119	73	62%	38%	Genic_UTR	MDV072	LORF5; function unknown; no H
Farm B/dust	108856	G	A	10%	254	164	6	41	13%	87%	Genic_UTR	MDV072	LORF5; function unknown; no H
Farm B/dust	108899	A	G	11%	227	192	7	45	13%	87%	Genic_UTR	MDV072	LORF5; function unknown; no H
Farm B/dust	109012	T	C	5%	184	176	2	16	11%	89%	Genic_UTR	MDV072	LORF5; function unknown; no H
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
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 fo
Farm B/dust	121656	C	T	37%	76	118	43	72	37%	63%	Genic_UTR	MDV076	Meq; oncogene; role in tumor fo
Farm B/dust	122052	A	T	3%	381	231	14	3	82%	18%	Genic_UTR	MDV076	Meq; oncogene; role in tumor fo
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

Supplemental Table S5. Summary and annotation of all polymorphic loci detected in MDV1 consensus genomes.
DNA from dust: comparative genomics of large DNA viruses in field surveillance samples

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 homc
Farm B/feather 1	23126	G	T	12.28%	28	21	2	5	29%	71%	Non-synonymous variant	MDV025	serine/threonine kinase; UL13 h

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 fo
Farm B/feather 2	122204	C	T	15.79%	21	27	2	7	22%	77.78%	Genic_UTR	MDV076	Meq; oncogene; role in tumor fo
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 hc

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

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

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