

DNA from Dust: Comparative Genomics of Large DNA Viruses in Field Surveillance Samples

Utsav Pandey,^a Andrew S. Bell,^b Daniel W. Renner,^a David A. Kennedy,^b Jacob T. Shreve,^a Chris L. Cairns,^b Matthew J. Jones,^b Patricia A. Dunn,^c Andrew F. Read,^b  Moriah L. Szpara^a

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, USA^a; Center for Infectious Disease Dynamics, Departments of Biology and Entomology, Pennsylvania State University, University Park, Pennsylvania, USA^b; Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, Pennsylvania, USA^c

ABSTRACT The intensification of the poultry industry over the last 60 years facilitated the evolution of increased virulence and vaccine breaks in Marek's disease virus (MDV-1). Full-genome sequences are essential for understanding why and how this evolution occurred, but what is known about genome-wide variation in MDV comes from laboratory culture. To rectify this, we developed methods for obtaining high-quality genome sequences directly from field samples without the need for sequence-based enrichment strategies prior to sequencing. We applied this to the first characterization of MDV-1 genomes from the field, without prior culture. These viruses were collected from vaccinated hosts that acquired naturally circulating field strains of MDV-1, in the absence of a disease outbreak. This reflects the current issue afflicting the poultry industry, where virulent field strains continue to circulate despite vaccination and can remain undetected due to the lack of overt disease symptoms. We found that viral genomes from adjacent field sites had high levels of overall DNA identity, and despite strong evidence of purifying selection, had coding variations in proteins associated with virulence and manipulation of host immunity. Our methods empower ecological field surveillance, make it possible to determine the basis of viral virulence and vaccine breaks, and can be used to obtain full genomes from clinical samples of other large DNA viruses, known and unknown.

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

KEYWORDS: Marek's disease virus, genomics, herpesviruses, polymorphism, virulence

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 unvaccinated hosts lacking

Received 14 May 2016 Accepted 25 August 2016 Published 5 October 2016

Citation Pandey U, Bell AS, Renner DW, Kennedy DA, Shreve JT, Cairns CL, Jones MJ, Dunn PA, Read AF, Szpara ML. 2016. DNA from dust: comparative genomics of large DNA viruses in field surveillance samples. *mSphere* 1(5):e00132-16. doi:10.1128/mSphere.00132-16.

Editor Gregory Allan Smith, Northwestern University Feinberg School of Medicine

Copyright © 2016 Pandey et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Moriah L. Szpara, moriah@psu.edu.

maternal antibodies in as little as 10 days (1–5). Today, mass immunizations with live attenuated vaccines help to control production losses, which are mainly associated with immunosuppression and losses due to condemnation of carcasses (4, 6). Almost 9 billion broiler chickens are vaccinated against MD each year in the United States 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 conditions favoring the evolutionary emergence of the hyperpathogenic strains that 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 (4, 9). 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 (10–13).

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

Recent high-throughput sequencing (HTS) applications have demonstrated that herpesvirus genomes can be captured from human clinical samples using genome amplification techniques such as oligonucleotide enrichment and PCR amplicon-based approaches (32–36). Here we present a method for the enrichment and isolation of viral genomes from dust and feather follicles, without the use of either of these solution-based enrichment methods. Chickens become infected with MDV by the inhalation of dust contaminated with virus shed from the feather follicles of infected birds. Although these vaccinated hosts were infected by and shedding wild MDV-1, there were no overt disease outbreaks. Deep sequencing of viral DNA from dust and feather follicles enabled us to observe, for the first time, the complete genome of MDV-1 directly from field samples of naturally infected hosts. This revealed variations in both new and 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/MDV100), 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 combination of HTS and targeted Sanger-based surveillance approaches can be combined to understand viral evolution in the field.

RESULTS

Sequencing, assembly, and annotation of new MDV-1 consensus genomes from the field. To assess the level of genomic diversity within and between field sites that are under real world selection, two commercial farms in central Pennsylvania (11 mi. apart) with a high prevalence of MDV-1 were chosen (Fig. 1A). These operations raise poultry for meat (also known as broilers) and house 25,000 to 30,000 individuals per house. The poultry were vaccinated with a bivalent vaccine composed of MDV-2 (strain SB-1) and herpesvirus of turkeys (HVT [strain FC126]). In contrast to the Rispen vaccine, which is an attenuated MDV-1 strain, MDV-2 and HVT can be readily distinguished from MDV-1 across the length of the genome, which allowed us to differentiate wild MDV-1

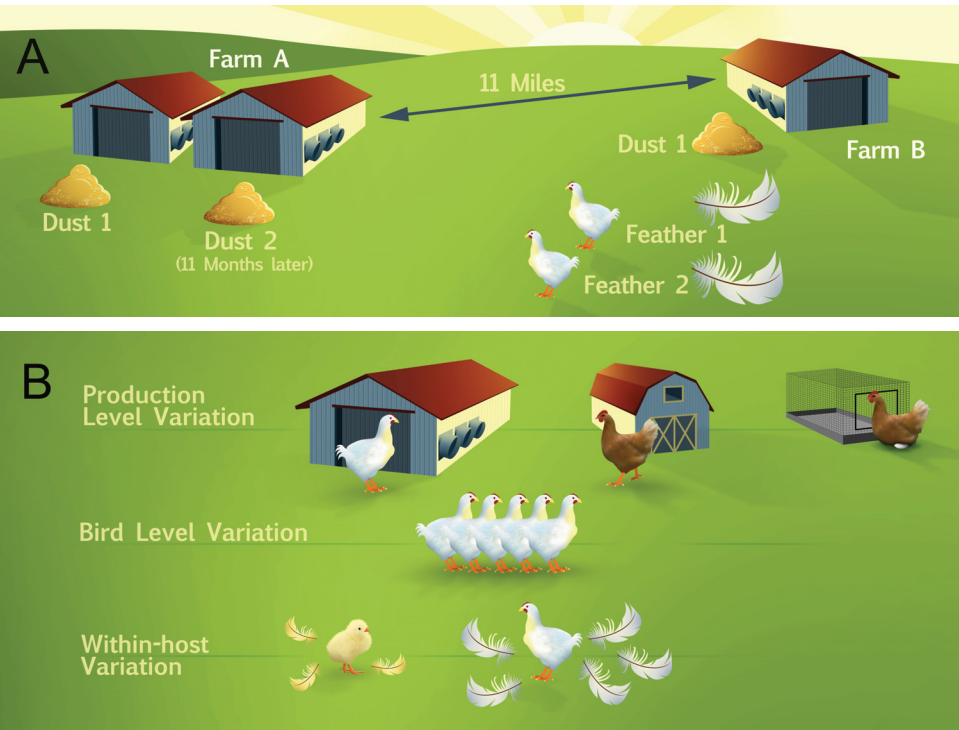


FIG 1 Diagram of samples collected for genome sequencing of field isolates of MDV. (A) Samples collected for genome sequencing were sourced from two Pennsylvania farms with large-scale operations that house approximately 25,000 to 30,000 individuals per building. These farms were separated by 11 mi. 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 (Table 1). (B) These methods can be used to explore additional aspects of variation in future studies. (Images are courtesy of Nick Sloff, Department of Entomology, Penn State University, reproduced with permission.)

from concomitant shedding of vaccine strains. These farms are part of a longitudinal study of MDV-1 epidemiology and evolution in modern agricultural settings (38).

To obtain material for genomic surveillance, we isolated MDV nucleocapsids from dust or epithelial tissues from the individual feather follicles of selected hosts (see Materials and Methods; in the supplemental material, see Fig. S1 and S2 for overviews and Tables S1 and S2 for DNA yields). A total of five uncultured wild-type samples of MDV were sequenced using an in-house Illumina MiSeq sequencer (Table 1, Illumina MiSeq output [see Materials and Methods for details]). The sequence read data derived from dust contained approximately 2 to 5% MDV-1 DNA, while the feather samples ranged from ~27% to 48% MDV-1 (Table 1, percentage of MDV-specific reads). 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-1 genomes from both types of source material.

TABLE 1 Field sample statistics and assembly of MDV-1 consensus genomes

Sample	Sample preparation			Illumina MiSeq output			New viral genomes		
	ng DNA	% MDV-1	% MDV-2	Total no. of reads ^a	No. of MDV-specific reads ^a	% MDV-specific reads	Avg depth (fold)	Genome length (bp)	NCBI accession no.
Farm A-dust 1	120	2.4	4.6	1.4 × 10 ⁷	3.7 × 10 ⁵	2.6	271	177,967	KU173116
Farm A-dust 2	127	1.3	2.7	2.5 × 10 ⁷	5.1 × 10 ⁵	2.0	333	178,049	KU173115
Farm B-dust	144	0.6	5.9	2.7 × 10 ⁷	1.4 × 10 ⁶	5.2	597	178,169	KU173119
Farm B-feather 1	12	40.6	0.1	3.9 × 10 ⁵	1.0 × 10 ⁵	26.9	44	178,327	KU173117
Farm B-feather 2	27	5.7	0	3.4 × 10 ⁵	1.7 × 10 ⁵	48.3	68	178,540	KU173118

^aThe sequence read counts shown are the sum of forward and reverse reads for each sample.

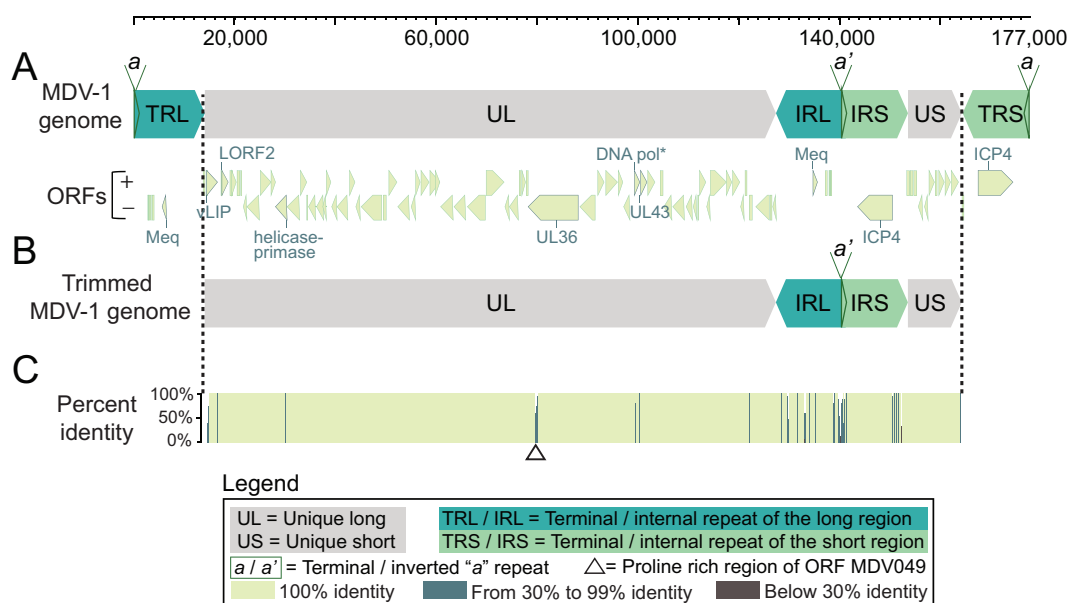


FIG 2 The complete MDV-1 genome includes two unique regions and two sets of large inverted repeats. (A) The full structure of the MDV-1 genome includes a unique long region (UL) and a unique short region (US), each of which is flanked by large repeats known as the terminal and internal repeats of the 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/MDV100), UL36 (MDV049), and Meq (MDV005/MDV076) (see Results for a complete list). (B) A trimmed-genome format without the terminal repeat regions was used for analyses in order to not overrepresent the repeat regions. (C) Percentage of identity from mean pairwise comparison of five consensus genomes, plotted spatially along the length of the genome. Darker colors indicate lower percentages of identity (see Legend).

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 (Fig. 2A) (39). 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, MDV-specific reads and average [fold] depth). The dust sample from Farm B had the highest coverage depth, at an average of almost 600× across the viral genome. Feather 1 from Farm B had the lowest coverage depth, averaging 44× genome-wide, which still exceeds that of most bacterial or eukaryotic genome assemblies. The genome length for all 5 samples was approximately 180 kb (Table 1), which is comparable to all previously sequenced MDV-1 isolates (14–21).

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 VZV (10, 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 viruses 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, examination of the polymorphic loci within each consensus genome assembly allowed us to observe the level of variation within the viral population at each point source (Fig. 3 to 4; see Fig. S3 and Table S3 in the supplemental material).

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

TABLE 2 Pairwise DNA identity and variant proteins between pairs of consensus genomes

Comparison	% DNA identity	Total no. of bp different	No. of intergenic:		No. of genic ^a :		
			Indels (events)	SNPs	Indels (events)	Synonymous SNPs	Nonsynonymous 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	1 in helicase-primase	3 (1 each in vLIP, LORF2, and UL43)
Farm B-dust vs Farm A-dust 2	99.87	195	49 (14)	76	66 (1) in DNA-pol	1 in helicase-primase	3 (1 each in vLIP, LORF2, and 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	1 in helicase-primase	3 (1 each in vLIP, LORF2, and UL43)
Farm B-dust vs Farm B-feather 2	99.52	687	572 (19)	45	66 (1) in DNA-pol	1 in helicase-primase	3 (1 each in vLIP, LORF2, and 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, 1 host vs another							
Farm B-feather 1 vs Farm B-feather 2	99.38	973	972 (9)	1	0	0	0

^aDNA-pol, DNA polymerase processivity subunit protein UL42 (MDV055); helicase-primase, helicase-primase subunit UL8 (MDV020); vLIP, lipase homolog (MDV010); LORF2, immune evasion protein (MDV012); UL43, probable membrane protein (MDV056).

percentage of homology ranging from 99.4% to 99.9% in pairwise comparisons (Fig. 2C; Table 2). These comparisons used a trimmed-genome format (Fig. 2B) in which 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 herpes simplex virus 1 (HSV-1) (39). Observed nucleotide differences were categorized as genic or intergenic and further subdivided

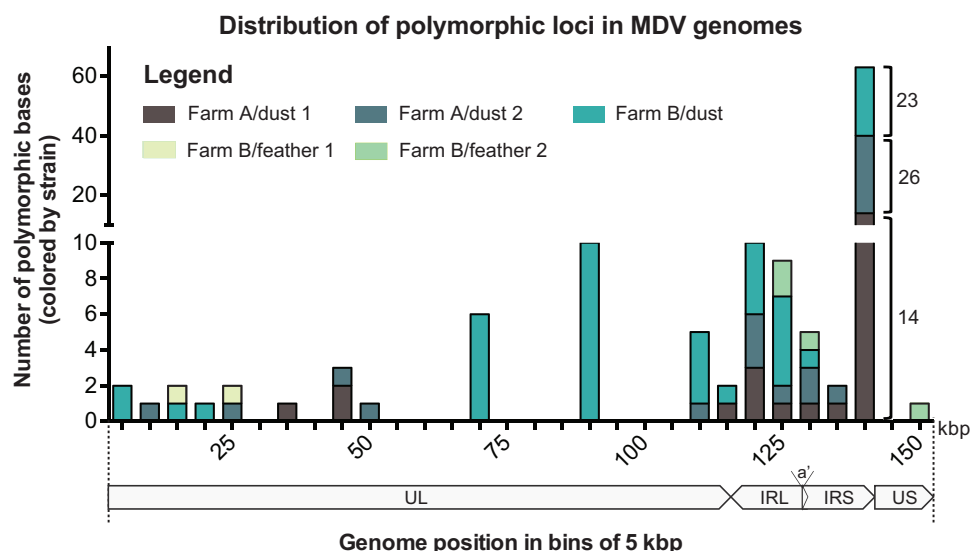


FIG 3 Genome-wide distribution of polymorphic bases within each consensus genome. Polymorphic base calls from each MDV genome were grouped in bins of 5 kb, and the sum of polymorphisms in each bin was plotted. Farm B-dust (aqua) contained the largest number of polymorphic bases, with the majority occurring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

Number of observed vs. expected polymorphisms in each genome

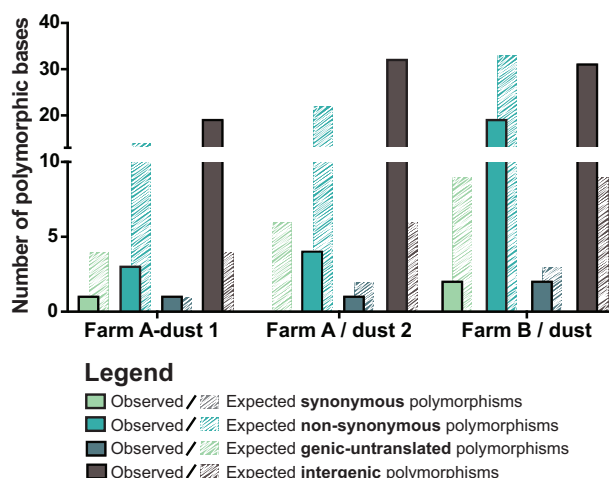


FIG 4 Observed versus expected polymorphism categories for each consensus genome. Each consensus genome was analyzed for the presence of polymorphic loci (see Materials and Methods for details). Observed polymorphic loci (solid bars) were categorized as causing synonymous (green) or nonsynonymous (aqua)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

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.

To understand the effect(s) of these nucleotide variations on protein coding and function, we next compared the amino acid sequences of all open reading frames (ORFs) for the five isolates. The consensus protein coding sequences of all five isolates were nearly identical, with just a few differences (Table 2). In comparison to the other four samples, Farm B-dust harbored amino acid substitutions in four proteins. A single nonsynonymous mutation was seen in each of the following: the virulence-associated lipase homolog vLIP (MDV010; Farm B-dust, S501A) (40), the major histocompatibility complex (MHC) class I immune evasion protein LORF2 (MDV012; Farm B-dust, L311W) (41), and the probable membrane protein UL43 (MDV056; Farm B-dust, S74L). A single synonymous mutation was observed in the DNA helicase-primase protein UL8 (MDV020; Farm B-dust, L253L). Finally, a 22-amino-acid (aa) insertion unique to Farm B-dust was observed in the DNA polymerase processivity subunit protein UL42 (MDV055; Farm B-dust, insertion at aa 277). We did not observe any coding differences between temporally separated dust isolates from Farm A or between feather isolates from different hosts in Farm B, although both of these comparisons (Table 2, bottom) revealed hundreds of noncoding differences.

Detection of polymorphic bases within each genome. Comparison of viral genomes found in different sites provides a macrolevel assessment of viral diversity. We next investigated the presence of polymorphic viral populations within each consensus genome to reveal how much diversity might exist within a field site (as reflected in dust-derived genomes) or within a single host (as reflected in feather genomes).

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 Materials and 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 to 58 polymorphic sites within each consensus genome (Fig. 3) (see Materials and Methods for details). The feather genomes had a lower number of polymorphisms than 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 (Fig. 3), with the majority concentrated in the repeat regions. Application of a more stringent set of parameters (see Materials and Methods for details) yielded a similar distribution of polymorphisms, albeit with no polymorphisms detected in feather samples due to their lower depth of coverage (see Fig. S3 in the supplemental material). 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, nonsynonymous, genic untranslated, or intergenic (see Table S3 in the supplemental material). The majority of all polymorphisms were located in intergenic regions (see Table S3). 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, nonsynonymous, 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 Materials and Methods), we determined that the distribution of variants differed from that expected by chance in each of our dust samples (Fig. 4, Farm A-dust 1, $\chi^2 = 68.16$, $df = 3$, $P < 0.001$; Farm A-dust 2, $\chi^2 = 128.57$, $df = 3$, $P < 0.001$; Farm B-dust 1, $\chi^2 = 63.42$, $df = 3$, $P < 0.001$). In addition, we found in pairwise tests that the number of observed intergenic polymorphisms was significantly higher than the observed values for other categories (see Table S4 in the supplemental material). 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/MDV100) gene (nucleotide position 5495) 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) (Fig. 5A). In contrast, this locus had been 99% A and only 1% C in Farm A-dust 1 (collected 11 months earlier in another house on the same farm), such that it was not counted as polymorphic in that sample by our parameters (see Materials and Methods for details). At this polymorphic locus, the nucleotide C encodes a serine, while nucleotide A encodes a tyrosine. The encoded amino acid lies in the C-terminal domain of ICP4 (aa 1832). ICP4 is an important immediate-early protein in all herpesviruses, where it serves as a major regulator of viral transcription (42–44). The role of ICP4 in MDV pathogenesis is also considered crucial because of its proximity to the latency-associated transcripts (LATs) and recently described miRNAs (44–46). In a previous study of MDV-1 attenuation through serial passage *in vitro*, mutations in ICP4 appeared to coincide with attenuation (31).

Given the very different allele frequencies at this ICP4 locus between two houses on the same farm 11 months apart, we examined dust samples from one of the houses over 9 months with targeted Sanger sequencing of this SNP (Fig. 5B). We found that this locus was highly polymorphic in time-separated dust samples. The A (tyrosine) allele

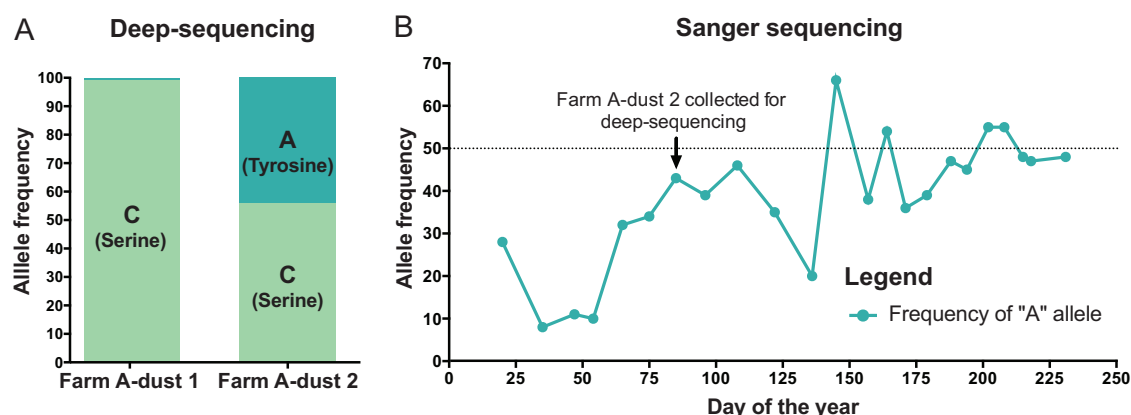


FIG 5 A new polymorphic locus in ICP4 and its shifting allele frequency over time. (A) HTS data revealed a new polymorphic locus in ICP4 (MDV084) at nucleotide position 5495. In the spatially and temporally separated dust samples from Farm A (see Fig. 1A and Materials and Methods for details), we observed different prevalences of C (encoding serine) and A (encoding tyrosine) alleles. (B) Using targeted Sanger sequencing of this locus, time-separated dust samples spanning 9 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.

rose to almost 50% frequency in the 9-month period. In four of the dust samples, the A (tyrosine) allele was dominant over the C (serine) allele. This reversible fluctuation in allele frequencies over a short period of time is unprecedented for alphaherpesviruses so far as we know. However, recent studies on human cytomegalovirus (HCMV) have shown that selection can cause viral populations to evolve in short periods of time (34, 35). While this is only one example of a polymorphic locus that shifts in frequency over time, similar approaches could be used at any of the hundreds of polymorphic loci detected here (see Table S3 in the supplemental material).

Comparison of field isolates of MDV-1 to previously sequenced isolates. To compare these new field-based MDV genomes to previously sequenced isolates of MDV, we created a multiple sequence alignment of all available MDV-1 genomes (14, 15, 17–21, 47, 48). The multiple sequence alignment was used to generate a dendrogram depicting genetic relatedness (see Materials and Methods). We observed that the five new isolates form a separate group compared to all previously sequenced isolates (Fig. 6). This may result from geographic differences as previously seen for HSV-1 and VZV (27, 49–52), from temporal differences in the time of sample isolations, or from the lack of cell culture adaptation in these new genomes.

We also noted a distinctive mutation in the genes encoding glycoprotein L (gL; also known as UL1 or MDV013). All of the field isolates had a 12-nucleotide deletion in gL that has been described previously in strains from the eastern United States. This deletion is found predominantly in very virulent or hypervirulent strains (vv and vv+ in the MDV-1 pathotyping nomenclature [1]) (53–56). This deletion falls in the putative cleavage site of gL, which is necessary for its posttranslational modification in the endoplasmic reticulum (54). Glycoprotein L forms a complex with another glycoprotein, gH. The gH/gL dimer is conserved across the *Herpesviridae* family and has been associated with virus entry (57, 58).

These field-isolated genomes also contain a number of previously characterized variations in the oncogenesis-associated Marek's EcoRI-Q-encoded protein (Meq; also known as MDV005, MDV076, and RLORF7). We observed three substitutions in the C-terminal (transactivation) domain of Meq (P153Q, P176A, and P217A) (59). The first two of these variations have been previously associated with MDV-1 strains of very virulent and hypervirulent pathotypes (vv and vv+) (53, 60, 61), while the third mutation has been shown to enhance transactivation (62). In contrast, the field isolates lacked the 59-aa insertion in the Meq proline repeats that is often associated with attenuation, as seen in the vaccine strain CVI988 and the mildly virulent strain CU-2 (47, 48, 63, 64). We also observed a C119R substitution in all five

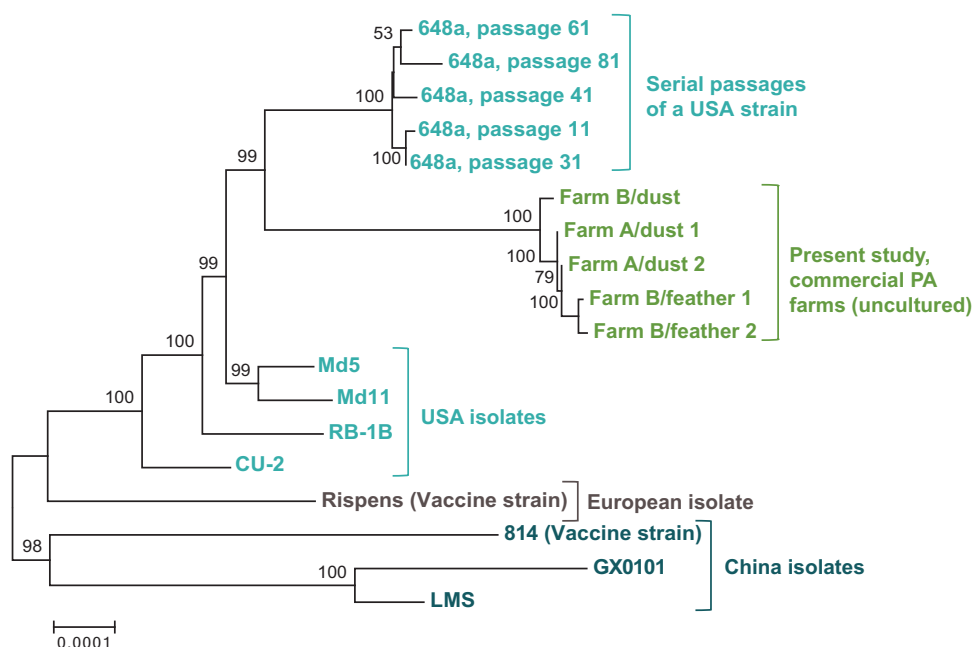


FIG 6 Dendrogram of genetic distances among all sequenced MDV-1 genomes. Using a multiple-genome alignment of all available complete MDV-1 genomes, we calculated the evolutionary distances between genomes using the Jukes-Cantor model. A dendrogram was then created using the neighbor-joining method in MEGA with 1,000 bootstraps. The five new field-sampled MDV-1 genomes (green) formed a separate group between the two clusters of United States isolates (blue). The European vaccine strain (Rispiens) formed a separate clade, as did the three Chinese MDV-1 genomes (aqua). GenBank accession numbers for all strains are as follows: new genomes, listed in 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](#); Rispiens (CVI988), [DQ530348](#); 814, [JF742597](#); GX0101, [JX844666](#); and LMS, [JQ314003](#).

field-derived genomes, which is absent from attenuated and mildly virulent isolates. This C119R mutation falls in the LXCXE motif of Meq, which normally binds to the tumor suppressor protein Rb to regulate cell cycle progression (53, 65). Although comprehensive *in vitro* and *in vivo* studies will be required to fully understand the biological implications of these variations, sequence comparisons of both gL and mEq from dust and feather genomes suggest that these closely resemble highly virulent (vv and vv+) variants of MDV-1 (47, 48). This is corroborated by the dendrogram (Fig. 6), where the dust- and feather-derived genomes cluster closely with 648a, which is a highly virulent (vv+) MDV isolate.

Assessment of taxonomic diversity in dust and 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 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, this provides useful insight into the overall complexity of each sample type. The results of the classification for the Farm B-dust, Farm B-feather 1, and Farm B-feather 2 samples are shown in Fig. S4 in the supplemental material. 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 Table S5 in the supplemental material. 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 to 5% were MDV specific (see also Table 1, percentage of MDV-specific reads). 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; see Fig. 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 oligonucleotide enrichment used for VZV or the PCR amplicon strategy used for HCMV (10, 33–35, 66). 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 nonsynonymous differences were detected between the field samples compared here, it is striking that several of these (vLIP, LORF2, and UL42) have been previously demonstrated to have roles in virulence and immune evasion. The N-glycosylated protein viral lipase (vLIP; MDV010) gene encodes a 120-kDa protein that is required for lytic virus replication in chickens (3, 40). The vLIP gene of MDV-1 is homologous to those of other viruses in the *Mardivirus* genus as well as to avian adenoviruses (67–69). The S501A mutation in the second exon of vLIP protein is not present in the conserved region that bears homology to other pancreatic lipases (40). The gene coding for the viral protein LORF2 (MDV012) is a viral immune evasion gene that suppresses MHC class I expression by inhibiting TAP transporter delivery of peptides to the endoplasmic reticulum (41). LORF2 is unique to the nonmammalian *Mardivirus* clade, but its function is analogous to that of the mammalian alphaherpesvirus product UL49.5 (41, 70, 71). Another study has shown that LORF2 is an essential phosphoprotein with a potential role as a nuclear/cytoplasmic shuttling protein (72). Interestingly, we also observed a 22-aa insertion in the DNA polymerase processivity subunit protein UL42 (MDV055). In herpesviruses, UL42 has been recognized as an integral part of the DNA polymerase complex, interacting directly with DNA and forming a heterodimer with the catalytic subunit of the polymerase (73–75). In HSV-1, the N-terminal two-thirds of UL42 has been shown to be sufficient for all known functions of UL42; the insertion in Farm B-dust falls at the edge of this N-terminal region (76). The nonsynonymous mutations and insertions detected here warrant further study to evaluate their impacts on protein function and viral fitness *in vivo*. The fact that any coding differences were observed in this small sampling of field-derived genomes suggests that the natural ecology of MDV-1 may include mutations and adaptations in protein function, in addition to genetic drift.

Drug resistance and vaccine failure have been attributed to the variation present in viral populations (10, 33, 77). Polymorphic populations allow viruses to adapt to diverse environments and withstand changing selective pressures, such as evading the host's immune system, adapting to different tissue compartments, and facilitating transmission between hosts (10, 26, 33–35, 66, 77, 78). Polymorphisms that were not fully penetrant in the consensus genomes, but that may be fodder for future selection, include residues in genes associated with virulence and immune evasion, such as ICP4 (Fig. 5), Meq, pp38, vLIP, LORF2, and others (see Table S3 in the supplemental material). The nonsynonymous polymorphism that we observed in Meq is a low-frequency variant present in the C-terminal domain (I201L) (Farm B-dust; see Table S3). However, a comparison of 88 different Meq sequences from GenBank and unpublished field isolates (A. S. Bell, D. A. Kennedy, P. A. Dunn, and A. F. Read, unpublished data) did not reveal any examples where leucine was the dominant allele; all sequenced isolates to date have isoleucine at position 201.

Previous studies have examined the accumulation of polymorphic loci in MDV-1 genomes after serial passage *in vitro* (30, 31). Overall, we found a similar quantity of polymorphisms in field-derived genomes to that found in these prior studies, but we

did not find any specific polymorphic loci that were identical between field-derived and *in vitro*-passaged genomes (30, 31). The ICP4 (MDV084/MDV100), LORF2 (MDV012), UL42, and MDV020 genes contain polymorphisms in both field and serially passaged isolates, albeit at different loci (30, 31). It is noteworthy that these coding variations are detected despite signs of clearance of polymorphisms from coding regions (Fig. 4), as indicated by the higher-than-expected ratios of intergenic to coding polymorphisms in these genomes. Together these findings suggest that MDV-1 exhibits genetic variation and undergoes rapid selection in the field, which may demonstrate the basis of its ability to overcome vaccine-induced host resistance to infection (3, 5, 79, 80).

For the viral transactivation protein ICP4, we explored the penetrance of a polymorphic locus (nucleotide position 5495) both in full-length genomes and also via targeted sequencing over time. Most of the work on this region in the MDV-1 genome has actually focused on the LATs that lie antisense to the ICP4 gene (44, 45). This polymorphic locus could thus impact either ICP4's coding sequence (aa 1832 serine versus tyrosine) or the sequence of the LATs. This variation in ICP4 lies in the C-terminal domain, which in HSV-1 has been implicated in the DNA synthesis, late gene expression, and intranuclear localization functions of ICP4 (81, 82). This combination of deep-sequencing genomic approaches to detect new polymorphic loci and fast gene-specific surveillance to track changes in SNP frequency over a larger number of samples illustrates the power of high-quality full genome sequences from field samples to provide powerful new markers for field ecology.

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 (Fig. 6). This pattern may result from geographic and temporal drift in these strains or from the wild, virulent nature of these strains versus 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 (27, 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 (14, 15, 17–21, 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 (25). Extended passage *in vitro* forms the basis of vaccine attenuation strategies, as for the successful vaccine strain (vOka) of the alphaherpesvirus VZV (83). 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 strains 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 (Fig. 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 (Fig. 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 (Fig. 1B). Evidence from tissue compartmentalization studies in HCMV and VZV suggests that viral genomes differ in distinct body niches (10, 35, 66). These new technique will enable us to ask similar questions about MDV-1 and to begin exploring the relative fitness levels of viruses found in different tissue compartments.

MATERIALS AND METHODS

Collection of dust and feathers. Samples were collected from two commercial-scale farms in central Pennsylvania, where each poultry building housed 25,000 to 30,000 individuals (Fig. 1A). The poultry on both farms were the same breed and strain of colored (“red”) commercial broiler chickens from the same hatchery and company. 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 airborne virus particles and feather dander. Sequential samples from Farm A (Table 1) were collected 11 months apart, from adjacent houses on the same farm (Fig. 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 to 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- to 1.0-cm proximal shaft or feather tip, which contains the feather pulp, was snipped into a sterile 1.5-ml microtube 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 of MDV-1 (see below for quantitative PCR [qPCR] details). The remaining feathers from the two animals with the 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 no. 46599).

Viral DNA isolation from dust. MDV nucleocapsids were isolated from dust as indicated in Fig. S1A in the supplemental material. Dust collected from poultry houses was stored in 50-ml polypropylene Falcon tubes (Corning) at 4°C until required. Five hundred milligrams of dust was suspended in 6.5 ml of $1\times$ phosphate-buffered saline (PBS). To distribute the dust particles into solution and help release cell-associated virus, the mixture was vortexed vigorously until homogenous and centrifuged at $2,000\times g$ for 10 min. This supernatant was further agitated on ice for 30 s using a Sonica ultrasonic processor Q125 (probe sonicator with 1/8-in. microtip) set to 20% amplitude. It was then vortexed before being centrifuged for a further 10 min at $2,000\times g$. To enrich viral capsids away from the remaining contaminants, the supernatant (approximately 5 ml in volume) was subjected to a series of filtration steps. First, we used a Corning surfactant-free cellulose acetate (SFCA) filter (0.8- μm pore) that had been soaked overnight in fetal bovine serum (FBS) to remove particles at the level of eukaryotic cells and bacteria. To remove smaller contaminants, the flowthrough was then passed through a Millipore Express Plus membrane vacuum filter (0.22- μm pore), and the membrane was subsequently washed twice with 2.5 ml of PBS. To remove contaminant DNA, the final filtrate (approximately 10 ml in volume) was treated with DNase (Sigma) at a concentration of 0.1 mg/ml for 30 min at room temperature. 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 pore). This filter membrane trapped the viral nucleocapsids, which are between 0.1 and 0.2 μm (84). 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 Table S1 in the supplemental material). 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 microtube containing 1.8 ml of lysis solution (ATL buffer and proteinase K from the Qiagen DNeasy blood and tissue kit). Digestion was allowed to proceed at 56°C for 1 h on an incubating microplate shaker (VWR) set to 1,100 rpm. The membrane was then removed, held vertically over a tilted sterile 5-cm-diameter plastic petri dish, and washed with a small volume of the lysis solution (from the 2-ml microtube). This wash was subsequently returned to the 2-ml microtube, and the tube was placed again on the heated shaker, where it was allowed to incubate overnight. The following day, the DNA was isolated as per the manufacturer’s instructions using the DNeasy blood and tissue kit (Qiagen). DNA was eluted in 200 μl DNase-free water. Ten to 14 aliquots of 500 ng each were used to obtain sufficient DNA for each dust sample (see Table S1). Quantitative PCR was used to assess the copy number of viral genomes in the resulting DNA. The total yield and percentage of MDV-1 versus MDV-2 DNA are listed in Table S1.

Isolation of viral DNA from feather follicles. The protocol for extraction of MDV DNA from feather follicles was optimized for the smaller input material and an expectation of higher purity (see Fig. S1B in the supplemental material). Sequential size filters were not used to filter out contaminants from

feather follicles, since these direct host samples have fewer impurities than the environmental samples of dust. However, the feather follicle cells were encased inside the keratinaceous shell of the feather tip, which required disruption to release the cells. Each tube containing a single feather tip and one sterile 5-mm-diameter steel bead was allowed to thaw, and then 200 μ l of PBS was added, and the sample was bead beaten for 30 s at 30 Hz using a TissueLyser (Qiagen) (see Fig. S1B). Vigorous bead beating achieved the desired destruction of the follicle tip. To dissociate the cells, 80 μ l of 2.5 mg/ml trypsin (Sigma) and 720 μ l of PBS were then added (final trypsin concentration, 0.8 mg/ml), and the solution was transferred to a new sterile 2-ml microtube and incubated for 2 h at 37°C on a heated microplate shaker (VWR) set to 700 rpm. To release cell-associated virus, the suspension was then sonicated on ice for 30 s using a Sonica ultrasonic processor Q125 (probe sonicator with 1/8-in. microtip) set to 50% amplitude. DNase I was added to a final concentration of 0.1 mg/ml and allowed to digest for 1 h at room temperature to remove nonencapsidated DNA. An equal volume of lysis solution (ATL buffer and proteinase K from the Qiagen DNeasy blood and tissue kit) was added, and the sample was incubated overnight at 56°C on an incubating microplate shaker (VWR) set to 1,100 rpm. The following day, the DNA was isolated as per the manufacturer's instructions using the DNeasy blood and tissue kit (Qiagen). While the overall amount of DNA obtained from feather follicles was lower than that obtained from pooled dust samples (see Table S2 in the supplemental material), it was of higher purity and was sufficient to generate libraries for sequencing (Table 1, sample preparation).

Measurement of total DNA and quantification of viral DNA. The total amount of DNA present in the samples was quantified by fluorescence analysis using a Qubit fluorescence assay (Invitrogen) following the manufacturer's recommended protocol. MDV genome copy numbers were determined using serotype-specific quantitative PCR (qPCR) primers and probes, targeting either the MDV-1 pp38 (MDV073; previously known as LORF14a) gene or MDV-2 (SB-1 strain) DNA polymerase (UL42, MDV055) gene. The MDV-1 assay was designed by Sue Baigent and used the forward primer Spp38for (5' GAGC TAACCGGAGAGGGAGA 3'), reverse primer Spp38rev (5' CGCATACCGACTTTCGTCAA 3'), and MDV-1 probe (FAM-CCCACTGTGACAGCC-BHQ1 [where FAM is 6-carboxyfluorescein and BHQ-1 is black hole quencher 1]) (S. Baigent, personal communication). The MDV-2 assay is that of Islam et al. (85), but with a shorter minor groove binder (MGB) probe (FAM-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 s followed by 40 cycles of denaturation at 95°C for 3 s and annealing and extension at 60°C for 30 s. Both assays included 4 μ l of DNA in a total PCR volume of 20 μ l with 1 \times 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 mixture incorporated 2 μ l bovine serum albumin (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 concentrations of the total DNA, MDV-1 DNA, and MDV-2 DNA present in the sample were known, we calculated the percentages of MDV-1 and MDV-2 genomic DNA in the total DNA pool (see Tables S1 and S2 in the supplemental material).

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. The genomic DNA inputs used for each sample are listed in Table 1. The DNA fragment size selected for library construction was 550 bp. All of the samples were sequenced on an in-house Illumina MiSeq using version 3 sequencing chemistry to obtain paired-end sequences of 300 by 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 (see Fig. S2 in the supplemental material) 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 (86) with a loose E value of $\leq 10^{-2}$ 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 versus vaccine virus (MDV-2 or HVT) strains. The overall DNA identity of MDV-1 and MDV-2 is just 61% (87). In a comparison of strains MDV-1 Md5 (NC_002229), and MDV-2 SB-1 (HQ840738), we found no spans of identical DNA greater than 50 bp (data not shown). This allowed us to accurately distinguish these 300- by 300-bp MiSeq sequence reads as being derived from either MDV-1 or MDV-2.

MDV genomes were assembled using the viral genome assembly VirGA (39) workflow, which combines quality control preprocessing of reads, *de novo* assembly, genome linearization and annotation, and postassembly 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 no. NC_002229.3). These new genomes were named according to recent recommendations, as outlined by Kuhn et al. (88). We use shortened forms of these names throughout the article (see Table 1 for short names). The full names for all five genomes are as follows: (i) MDV-1 *Gallus domesticus*-wt/Pennsylvania, United States/2015/Farm A-dust 1; (ii) MDV-1 *Gallus domesticus*-wt/Pennsylvania, United States/2015/Farm A-dust 2; (iii) MDV-1 *Gallus domesticus*-wt/Pennsylvania, United States/2015/Farm B-dust; (iv) MDV-1 *Gallus domesticus*-wt/Pennsylvania, United States/2015/Farm B-feather 1; and (v) MDV-1 *Gallus domesticus*-wt/Pennsylvania, United States/2015/Farm B-feather 2. GenBank accession numbers are listed below and in Table 1. Annotated copies of each genome, in a format compatible with genome and sequence browsers, are

available at the Pennsylvania State University ScholarSphere data repository: <https://scholarsphere.psu.edu/collections/1544bp14j>.

Between-sample consensus genome comparisons. Clustalw2 (37) was used to construct pairwise global nucleotide alignments between whole-genome sequences and pairwise global amino acid alignments between open reading frames. These alignments were utilized by downstream custom python scripts to calculate percentage of identity, protein differences, and variation between samples.

The proline-rich region of UL36 (also known as VP1/2 or MDV049), which contains an extended array of tandem repeats, was removed from all five consensus genomes prior to comparison. The amount of polymorphism seen in this region of UL36 is driven by fluctuations in the length of these tandem repeats, as has been seen in prior studies with other alphaherpesviruses such as HSV, VZV, and pseudorabies virus (PRV) (32, 48–50, 99, 100). Since the length of extended arrays of perfect repeats cannot be precisely determined by *de novo* assembly (22, 23, 26, 27), we excluded this region from pairwise comparisons of genome-wide variation. Genome alignments with and without the UL36 region removed are archived at the ScholarSphere site: <https://scholarsphere.psu.edu/collections/1544bp14j>.

Within-sample polymorphism detection within each consensus genome. VarScan v2.2.11 (89) was used to detect variants present within each consensus genome. To aid in differentiating true variants from potential sequencing errors (90), two separate variant-calling analyses were explored (10). Our main polymorphism detection parameters (used in Fig. 3 and 4; see Tables S3 and S4 in the supplemental material) were as follows: minimum variant allele frequency, ≥ 0.02 ; base call quality, ≥ 20 ; read depth at the position, ≥ 10 ; and number of independent reads supporting the minor allele, ≥ 2 . Directional strand bias of $\geq 90\%$ was excluded; a minimum of two reads in opposing directions was required. For comparison and added stringency, we also explored a second set of parameters (used in Fig. S3 in the supplemental material): minimum variant allele frequency, ≥ 0.05 ; base call quality, ≥ 20 ; read depth at the position, ≥ 100 ; and number of independent reads supporting the minor allele, ≥ 5 . Directional strand bias of $\geq 80\%$ was excluded. The variants obtained from VarScan were then mapped back to the genome to understand their distribution and mutational impact using SnpEff and SnpSift (91, 92). Polymorphisms in the proline-rich region of UL36 were excluded, as noted above.

Testing for signs of selection acting on polymorphic viral populations. For each of our five consensus genomes, which each represent a viral population, we classified the polymorphisms detected into categories of synonymous, nonsynonymous, genic untranslated, or intergenic, based on where each polymorphism was positioned in the genome. For these analyses (Fig. 4), we were only able to include polymorphisms detected in the three dust genomes, since the total number of polymorphisms obtained from feather genomes was too low for chi-square analysis. First, we calculated the total possible number of single nucleotide mutations that could be categorized as synonymous, nonsynonymous, genic untranslated, or intergenic. To remove ambiguity when mutations in overlapping genes could be classified as either synonymous or nonsynonymous, genes with alternative splice variants or overlapping reading frames were excluded from these analyses. This removed 25 open reading frames (approximately 21% of the genome). These tallies of potential mutational events were used to calculate the expected fraction of mutations in each category. We performed chi-square tests on each data set to assess whether the observed distribution of polymorphisms matched the expected distribution. We also performed a similar analysis in pairwise fashion (see Table S4 in the supplemental material), to assess whether the fraction of variants differed from what would be expected by random chance. Pairwise combinations included the following: synonymous versus nonsynonymous, synonymous versus intergenic, synonymous versus genic untranslated, nonsynonymous versus intergenic, nonsynonymous versus genic untranslated, and intergenic versus genic untranslated. Statistically significant outcomes would suggest that recent or historical selection differed between those categories of variants.

Sanger sequencing of polymorphic locus in ICP4. A potential locus of active selection within the ICP4 (MDV084/MDV100) gene was detected during deep sequencing of Farm B-dust. This locus was examined using Sanger sequencing. An approximately 400-bp region of the ICP4 gene was amplified using a Taq PCR core kit (Qiagen) and the following primers at 200 nM: forward primer ICP4seIF (5' AACACCTCTTGCCATGGTTC 3') and reverse primer ICP4seIR (5' GGACCAATCATCCTCTCTGG 3'). Cycling conditions included an initial denaturation of 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 1 min, with a terminal extension at 72°C for 10 min. The total reaction volume of 50 μ l included 10 μ l of DNA and 4 μ l bovine serum albumin (BSA [final concentration, 0.8 mg/ml]). Amplification products were visualized on a 1.5% agarose gel; the target amplicon was excised and then purified using the EZNA 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 (93).

Genetic distance and dendrogram. Multiple sequence alignments of complete MDV-1 (*Gallid herpesvirus 2*) genomes from GenBank and those assembled by our lab were generated using MAFFT (94). The evolutionary distances were computed using the Jukes-Cantor method (95), and the evolutionary history was inferred using the neighbor-joining method (96) in MEGA6 (97), with 1,000 bootstrap replicates (98). Positions containing gaps and missing data were excluded. The 18-strain genome alignment is archived at ScholarSphere (<https://scholarsphere.psu.edu/collections/1544bp14j>).

Taxonomic estimation of non-MDV sequences in dust and feathers. All sequence reads from each sample were submitted to a quality control preprocessing method to remove sequencing primers, artifacts, and areas of low confidence (39). Sequence annotation was performed using a massively iterative all-versus-all BLASTN (E value, $\leq 10^{-2}$) approach using the all-nucleotide database from NCBI.

Only a portion of the total sequence read pool could be identified with confidence using this method. We then used *de novo* assembly to extend the length of these unidentified sequences, therefore elongating them into contigs. These were iterated through BLASTN again, which revealed alignment to repetitive regions of the *Gallus domesticus* (chicken) genome. Since the viral DNA enrichment procedures include a level of stochasticity in removal of host and environmental contaminants, the proportion of taxa present is not a definitive outline of those present initially. The results of these classifications are shown in Fig. S4 and listed in Table S5 in the supplemental material.

Accession number(s). GenBank accession numbers 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; and Farm B-feather 2, KU173118. Additional files used in this article, such as multiple sequence alignments of these genomes, are archived and available at ScholarSphere (<https://scholarsphere.psu.edu/collections/1544bp14j>).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/mSphere.00132-16>.

Figure S1, PDF file, 1.1 MB.

Figure S2, PDF file, 0.9 MB.

Figure S3, PDF file, 0.9 MB.

Figure S4, PDF file, 0.9 MB.

Table S1, PDF file, 0.1 MB.

Table S2, PDF file, 0.1 MB.

Table S3, XLSX file, 0.1 MB.

Table S4, PDF file, 0.1 MB.

Table S5, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We thank Sue Baigent, Michael DeGiorgio, Peter Kerr, and members of the 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, as well as by startup funds (M.L.S.) from the Pennsylvania State University. This work was partly funded by the Institute of General Medical Sciences, National Institutes of Health (R01GM105244 [A.F.R.]) as part of the joint NSF-NIH-USDA Ecology & Evolution of Infectious Diseases Program.

FUNDING INFORMATION

This work, including the efforts of Andrew F. Read, was funded by HHS | National Institutes of Health (NIH) (NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program). This work, including the efforts of Andrew F. Read, was funded by HHS | NIH | National Institute of General Medical Sciences (NIGMS) (R01GM105244). This work, including the efforts of Moriah L. Szpara, was funded by Pennsylvania State University (PSU) (Startup funds).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Witter RL. 1997. Increased virulence of Marek's disease virus field isolates. *Avian Dis* **41**:149–163.
- Biggs PM. 2004. Marek's disease: long and difficult beginnings, p 8–16. In Davison F, Nair V (ed), *Marek's disease*. Academic Press, Oxford, United Kingdom.
- 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. <http://dx.doi.org/10.1038/nrmicro1382>.
- Gimeno IM. 2008. Marek's disease vaccines: a solution for today but a worry for tomorrow? *Vaccine* **26**(Suppl 3):C31–C41. <http://dx.doi.org/10.1016/j.vaccine.2008.04.009>.
- 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**:e1002198. <http://dx.doi.org/10.1371/journal.pbio.1002198>.
- 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. <http://dx.doi.org/10.1080/03079458408418510>.
- US Department of Agriculture Economics, Statistics and Market Information System. 2016. Poultry slaughter annual summary. US Department of Agriculture Economic, Statistics and Market Information System, US Department of Agriculture, Washington, DC.
- Fakhrul Islam AF, 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. <http://dx.doi.org/10.1080/03079450701802230>.
- Nair V. 2005. Evolution of Marek's disease—a paradigm for incessant

- race between the pathogen and the host. *Vet J* **170**:175–183. <http://dx.doi.org/10.1016/j.tvjl.2004.05.009>.
10. 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. <http://dx.doi.org/10.1093/molbev/mst210>.
11. 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. <http://dx.doi.org/10.1002/rmv.1789>.
12. Zerboni L, Sen N, Oliver SL, Arvin AM. 2014. Molecular mechanisms of varicella zoster virus pathogenesis. *Nat Rev Microbiol* **12**:197–210. <http://dx.doi.org/10.1038/nrmicro3215>.
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. <http://dx.doi.org/10.1093/molbev/msu406>.
14. 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. <http://dx.doi.org/10.1128/JVI.74.17.7980-7988.2000>.
15. 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. <http://dx.doi.org/10.1007/s00705-005-0633-7>.
16. Spatz SJ, Silva RF. 2007. Sequence determination of variable regions within the genomes of Gallid herpesvirus-2 pathotypes. *Arch Virol* **152**:1665–1678. <http://dx.doi.org/10.1007/s00705-007-0992-3>.
17. 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. <http://dx.doi.org/10.1007/s11262-007-0157-1>.
18. Zhang F, Liu C-J, Zhang Y-P, Li Z-J, Liu A-L, Yan F-H, Cong F, Cheng Y. 2012. Comparative full-length sequence analysis of Marek's disease virus vaccine strain 814. *Arch Virol* **157**:177–183. <http://dx.doi.org/10.1007/s00705-011-1131-8>.
19. Cheng Y, Cong F, Zhang YP, Li ZJ, Xu NN, Hou GY, Liu CJ. 2012. Genome sequence determination and analysis of a Chinese virulent strain, LMS, of Gallid herpesvirus type 2. *Virus Genes* **45**:56–62. <http://dx.doi.org/10.1007/s11262-012-0739-4>.
20. 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. <http://dx.doi.org/10.1007/s11262-012-0792-z>.
21. 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. <http://dx.doi.org/10.1128/JVI.02583-12>.
22. Peters GA, Tyler SD, Grose C, Severini A, Gray MJ, Upton C, Tipples GA. 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. <http://dx.doi.org/10.1128/JVI.00715-06>.
23. 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**:447–458. <http://dx.doi.org/10.1016/j.virol.2006.09.037>.
24. Bradley AJ, Lurain NS, Ghazal P, Trivedi U, Cunningham C, Baluchova K, Gatherer D, Wilkinson GW, Dargan DJ, Davison AJ. 2009. High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. *J Gen Virol* **90**:2375–2380. <http://dx.doi.org/10.1099/vir.0.013250-0>.
25. Dargan DJ, Douglas E, Cunningham C, Jamieson F, Stanton RJ, Baluchova K, McSharry BP, Tomassec P, Emery VC, Percivalle E, Sarasini A, Gerna G, Wilkinson GW, Davison AJ. 2010. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J Gen Virol* **91**:1535–1546. <http://dx.doi.org/10.1099/vir.0.018994-0>.
26. 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**:e1002282. <http://dx.doi.org/10.1371/journal.ppat.1002282>.
27. 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**:1209–1227. <http://dx.doi.org/10.1128/JVI.01987-13>.
28. 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 AA, 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* **16**:8219–8232. <http://dx.doi.org/10.1128/JVI.01303-15>.
29. 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.
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. <http://dx.doi.org/10.1016/j.virusres.2010.01.007>.
31. Hildebrandt E, Dunn JR, Perumbakkam S, Niikura M, Cheng HH. 2014. Characterizing the molecular basis of attenuation of Marek's disease virus via in vitro serial passage identifies de novo mutations in the helicase-primase subunit gene UL5 and other candidates associated with reduced virulence. *J Virol* **88**:6232–6242. <http://dx.doi.org/10.1128/JVI.03869-13>.
32. Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Derrick J, Thomson M, Griffiths PD, Wilkinson GW, Schulz TF, Dargan DJ, Davison AJ. 2010. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol* **91**:605–615. <http://dx.doi.org/10.1099/vir.0.015891-0>.
33. Depledge DP, Palser AL, Watson SJ, Lai IY, 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**:e27805. <http://dx.doi.org/10.1371/journal.pone.0027805>.
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. <http://dx.doi.org/10.1371/journal.ppat.1001344>.
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. <http://dx.doi.org/10.1371/journal.pgen.1003735>.
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. <http://dx.doi.org/10.1186/1471-2164-14-804>.
37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948. <http://dx.doi.org/10.1093/bioinformatics/btm404>.
38. Kennedy DA, Cairns CL, Jones MJ, Bell AS, Salathe RM, Baigent SJ, Nair VK, Dunn PA, Read AF. 2016. Industry-wide surveillance of Marek's disease virus on commercial poultry farms: underlying potential for virulence evolution and vaccine escape. *bioRxiv* <http://dx.doi.org/10.1101/075192>.
39. Parsons LR, Tafuri YR, Shreve JT, Bowen CD, Shipley MM, Enquist LW, Szpara ML. 2015. Rapid genome assembly and comparison decode intrastrain variation in human alphaherpesviruses. *mBio* **6**:e02213-14. <http://dx.doi.org/10.1128/mBio.02213-14>.
40. Kamil JP, Tischer BK, Trapp S, Nair VK, Osterrieder N, Kung H-J. 2005. vLIP, a viral lipase homologue, is a virulence factor of Marek's disease virus. *J Virol* **79**:6984–6996. <http://dx.doi.org/10.1128/JVI.79.11.6984-6996.2005>.
41. Hearn C, Preeyanon L, Hunt HD, York IA. 2015. An MHC class I immune evasion gene of Marek's disease virus. *Virology* **475**:88–95. <http://dx.doi.org/10.1016/j.virol.2014.11.008>.
42. Ahlers SE, Feldman LT. 1987. Immediate-early protein of pseudorabies virus is not continuously required to reinitiate transcription of induced genes. *J Virol* **61**:1258–1260.
43. Wu CL, Wilcox KW. 1991. The conserved DNA-binding domains encoded by the herpes simplex virus type 1 ICP4, pseudorabies virus IE180, and varicella-zoster virus ORF62 genes recognize similar sites in the corresponding promoters. *J Virol* **65**:1149–1159.
44. Xie Q, Anderson AS, Morgan RW. 1996. Marek's disease virus (MDV) ICP4, pp38, and Meq genes are involved in the maintenance of trans-

- formation of MDCC-MSB1 MDV-transformed lymphoblastoid cells. *J Virol* **70**:1125–1131.
45. Cantello JL, Parcells MS, Anderson AS, Morgan RW. 1997. Marek's disease virus latency-associated transcripts belong to a family of spliced RNAs that are antisense to the ICP4 homolog gene. *J Virol* **71**:1353–1361.
 46. Nair V. 2013. Latency and tumorigenesis in Marek's disease. *Avian Dis* **57**:360–365. <http://dx.doi.org/10.1637/10470-121712-Reg.1>.
 47. Spatz SJ, Rue CA. 2008. Sequence determination of a mildly virulent strain (CU-2) of Gallid herpesvirus type 2 using 454 pyrosequencing. *Virus Genes* **36**:479–489. <http://dx.doi.org/10.1007/s11262-008-0213-5>.
 48. Spatz SJ, Petherbridge L, Zhao Y, Nair V. 2007. Comparative full-length sequence analysis of oncogenic and vaccine (Rispen) strains of Marek's disease virus. *J Gen Virol* **88**:1080–1096. <http://dx.doi.org/10.1099/vir.0.82600-0>.
 49. Norberg P, Tyler S, Severini A, Whitley R, Liljeqvist JA, Bergström T. 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. *PLoS One* **6**:e22527. <http://dx.doi.org/10.1371/journal.pone.0022527>.
 50. Grose C. 2012. Pangaea and the out-of-Africa model of varicella-zoster virus evolution and phylogeography. *J Virol* **86**:9558–9565. <http://dx.doi.org/10.1128/JVI.00357-12>.
 51. Kolb AW, Ané C, Brandt CR. 2013. Using HSV-1 genome phylogenetics to track past human migrations. *PLoS One* **8**:e76267. <http://dx.doi.org/10.1371/journal.pone.0076267>.
 52. Chow VT, Tipples GA, Grose C. 2013. Bioinformatics of varicella-zoster virus: single nucleotide polymorphisms define clades and attenuated vaccine genotypes. *Infect Genet Evol* **18**:351–356. <http://dx.doi.org/10.1016/j.meegid.2012.11.008>.
 53. Shamblin CE, Greene N, Arumugaswami V, Dienglewicz RL, Parcells MS. 2004. Comparative analysis of Marek's disease virus (MDV) glycoprotein, lytic antigen pp38- and transformation antigen Meq-encoding genes: association of Meq mutations with MDVs of high virulence. *Vet Microbiol* **102**:147–167. <http://dx.doi.org/10.1016/j.vetmic.2004.06.007>.
 54. Santin ER, Shamblin CE, Prigge JT, Arumugaswami V, Dienglewicz RL, Parcells MS. 2006. Examination of the effect of a naturally occurring mutation in glycoprotein L on Marek's disease virus pathogenesis. *Avian Dis* **50**:96–103. <http://dx.doi.org/10.1637/7273-090704R1.1>.
 55. Tavlirides-Hontz P, Kumar PM, Amortegui JR, Osterrieder N, Parcells MS. 2009. A deletion within glycoprotein L of Marek's disease virus (MDV) field isolates correlates with a decrease in bivalent MDV vaccine efficacy in contact-exposed chickens. *Avian Dis* **53**:287–296. <http://dx.doi.org/10.1637/8558-121208-Reg.1>.
 56. Shaikh SA, Katneni UK, Dong H, Gaddamanugu S, Tavlirides-Hontz P, Jarosinski KW, Osterrieder N, Parcells MS. 2013. A deletion in the glycoprotein L (gL) gene of U.S. Marek's disease virus (MDV) field strains is insufficient to confer increased pathogenicity to the bacterial artificial chromosome (BAC)-based strain, RB-1B. *Avian Dis* **57**:509–518. <http://dx.doi.org/10.1637/10450-112012-Reg.1>.
 57. Gianni T, Massaro R, Campadelli-Fiume G. 2015. Dissociation of HSV gL from gH by $\alpha\text{v}\beta 6$ - or $\alpha\text{v}\beta 8$ -integrin promotes gH activation and virus entry. *Proc Natl Acad Sci U S A* **112**:E3901–E3910. <http://dx.doi.org/10.1073/pnas.1506846112>.
 58. Wu P, Reed WM, Lee LF. 2001. Glycoproteins H and L of Marek's disease virus form a hetero-oligomer essential for translocation and cell surface expression. *Arch Virol* **146**:983–992. <http://dx.doi.org/10.1007/s007050170130>.
 59. Qian Z, Brunovskis P, Rauscher F, Lee L, Kung HJ. 1995. Transactivation activity of Meq, a Marek's disease herpesvirus bZIP protein persistently expressed in latently infected transformed T cells. *J Virol* **69**:4037–4044.
 60. Nair V, Kung H-J. 2004. Marek's disease virus oncogenicity: molecular mechanisms, p 32–48. *In* Davison F, Nair V (ed), Marek's disease, vol. 4. Academic Press, Oxford, United Kingdom.
 61. Tian M, Zhao Y, Lin Y, Zou N, Liu C, Liu P, Cao S, Wen X, Huang Y. 2011. Comparative analysis of oncogenic genes revealed unique evolutionary features of field Marek's disease virus prevalent in recent years in China. *Virus J* **8**:121. <http://dx.doi.org/10.1186/1743-422X-8-121>.
 62. Murata S, Okada T, Kano R, Hayashi Y, Hashiguchi T, Onuma M, Konnai S, Ohashi K. 2011. Analysis of transcriptional activities of the Meq proteins present in highly virulent Marek's disease virus strains, RB1B and Md5. *Virus Genes* **43**:66–71. <http://dx.doi.org/10.1007/s11262-011-0612-x>.
 63. Spatz SJ, Silva RF. 2007. Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek's disease virus 1. *Virus Genes* **35**:41–53. <http://dx.doi.org/10.1007/s11262-006-0024-5>.
 64. Renz KG, Cooke J, Clarke N, Cheetham BF, Hussain Z, Fakhru Islam AF, Tannock GA, Walkden-Brown SW. 2012. Pathotyping of Australian isolates of Marek's disease virus and association of pathogenicity with meq gene polymorphism. *Avian Pathol* **41**:161–176. <http://dx.doi.org/10.1080/03079457.2012.656077>.
 65. Giacinti C, Giordano A. 2006. RB and cell cycle progression. *Oncogene* **25**:5220–5227. <http://dx.doi.org/10.1038/sj.onc.1209615>.
 66. 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 U S A* **112**:E4120–E4128. <http://dx.doi.org/10.1073/pnas.1501880112>.
 67. Afonso CL, Tulman ER, Lu Z, Zsak L, Rock DL, Kutish GF. 2001. The genome of turkey herpesvirus. *J Virol* **75**:971–978. <http://dx.doi.org/10.1128/JVI.75.2.971-978.2001>.
 68. Izumiya Y, Jang H-K, Ono M, Mikami T. 2001. A complete genomic DNA sequence of Marek's disease virus type 2, strain HPRS24, p 191–221. *In* Hirai PDK (ed), Marek's disease. Springer Verlag, Berlin, Germany.
 69. Ojick D, Nagy É. 2000. The complete nucleotide sequence of fowl adenovirus type 8. *J Gen Virol* **81**:1833–1837. <http://dx.doi.org/10.1099/0022-1317-81-7-1833>.
 70. Koppers-Lalic D, Verweij MC, Lipińska AD, Wang Y, Quinten E, Reits EA, Koch J, Loch S, Marcondes Rezende M, Daus F, Biełkowska-Szewczyk K, Osterrieder N, Mettenleiter TC, Heemskerck MH, Tampé R, Neefjes JJ, Chowdhury SI, Rensing ME, Rijsewijk FA, Ejh W. 2008. Varicellovirus UL49.5 proteins differentially affect the function of the transporter associated with antigen processing, TAP. *PLoS Pathog* **4**:e1000080. <http://dx.doi.org/10.1371/journal.ppat.1000080>.
 71. Verweij MC, Lipińska AD, Koppers-Lalic D, van Leeuwen WF, Cohen JI, Kinchington PR, Messaoudi I, Biełkowska-Szewczyk K, Rensing ME, Rijsewijk FA, Wiertz EJ. 2011. The capacity of UL49.5 proteins to inhibit TAP is widely distributed among members of the genus Varicellovirus. *J Virol* **85**:2351–2363. <http://dx.doi.org/10.1128/JVI.01621-10>.
 72. Schippers T, Jarosinski K, Osterrieder N. 2015. The ORF012 gene of Marek's disease virus type 1 produces a spliced transcript and encodes a novel nuclear phosphoprotein essential for virus growth. *J Virol* **89**:1348–1363. <http://dx.doi.org/10.1128/JVI.02687-14>.
 73. Zuccola HJ, Filman DJ, Coen DM, Hogle JM. 2000. The crystal structure of an unusual processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate polymerase. *Mol Cell* **5**:267–278. [http://dx.doi.org/10.1016/S1097-2765\(00\)80422-0](http://dx.doi.org/10.1016/S1097-2765(00)80422-0).
 74. Wang Y-P, Du W-J, Huang L-P, Wei Y-W, Wu H-L, Feng L, Liu C-M. 2016. The pseudorabies virus DNA polymerase accessory subunit UL42 directs nuclear transport of the holoenzyme. *Front Microbiol* **7**:124. <http://dx.doi.org/10.3389/fmicb.2016.00124>.
 75. Zhukovskaya NL, Guan H, Saw YL, Nuth M, Ricciardi RP. 2015. The processivity factor complex of yellow fever virus is a new drug target. *Antiviral Res* **115**:17–20. <http://dx.doi.org/10.1016/j.antiviral.2014.12.013>.
 76. Digard P, Chow CS, Pirrit L, Coen DM. 1993. Functional analysis of the herpes simplex virus UL42 protein. *J Virol* **67**:1159–1168.
 77. Domingo E, Martín V, Perales C, Grande-Pérez A, García-Arriaza J, Arias A. 2006. Viruses as quasispecies: biological implications, p 51–82. *In* Domingo E (ed), Quasispecies: concept and implications for virology. Springer Verlag, Berlin, Germany.
 78. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. 1982. Rapid evolution of RNA genomes. *Science* **215**:1577–1585. <http://dx.doi.org/10.1126/science.7041255>.
 79. Atkins KE, Read AF, Savill NJ, Renz KG, Islam AF, Walkden-Brown SW, Woolhouse ME. 2013. Vaccination and reduced cohort duration can drive virulence evolution: Marek's disease virus and industrialized agriculture. *Evolution* **67**:851–860. <http://dx.doi.org/10.1111/j.1558-5646.2012.01803.x>.
 80. Atkins KE, Read AF, Walkden-Brown SW, Savill NJ, Woolhouse ME. 2013. The effectiveness of mass vaccination on Marek's disease virus (MDV) outbreaks and detection within a broiler barn: a modeling study. *Epidemics* **5**:208–217. <http://dx.doi.org/10.1016/j.epidem.2013.10.001>.
 81. DeLuca NA, Schaffer PA. 1988. Physical and functional domains of the

- herpes simplex virus transcriptional regulatory protein ICP4. *J Virol* **62**:732–743.
82. **Wagner LM, Lester JT, Sivrich FL, DeLuca NA.** 2012. The N terminus and C terminus of herpes simplex virus 1 ICP4 cooperate to activate viral gene expression. *J Virol* **86**:6862–6874. <http://dx.doi.org/10.1128/JVI.00651-12>.
83. **Arvin AM, Gershon AA.** 1996. Live attenuated varicella vaccine. *Annu Rev Microbiol* **50**:59–100. <http://dx.doi.org/10.1146/annurev.micro.50.1.59>.
84. **Pellett PE, Roizman B.** 2013. Herpesviridae, p 1802–1822. *In* Fields virology, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
85. **Islam A, Harrison B, Cheetham BF, Mahony TJ, Young PL, Walkden-Brown SW.** 2004. Differential amplification and quantitation of Marek's disease viruses using real-time polymerase chain reaction. *J Virol Methods* **119**:103–113. <http://dx.doi.org/10.1016/j.jviromet.2004.03.006>.
86. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
87. **Spatz SJ, Schat KA.** 2011. Comparative genomic sequence analysis of the Marek's disease vaccine strain SB-1. *Virus Genes* **42**:331–338. <http://dx.doi.org/10.1007/s11262-011-0573-0>.
88. **Kuhn JH, Bao Y, Bavari S, Becker S, Bradfute S, Brister JR, Bukreyev AA, Chandran K, Davey RA, Dolnik O, Dye JM, Enterlein S, Hensley LE, Honko AN, Jahrling PB, Johnson KM, Kobinger G, Leroy EM, Lever MS, Mühlberger E, Netesov SV, Olinger GG, Palacios G, Patterson JL, Paweska JT, Pitt L, Radoshitzky SR, Saphire EO, Smither SJ, Swanepoel R, Towner JS, van der Groen G, Volchkov VE, Wahl-Jensen V, Warren TK, Weidmann M, Nichol ST.** 2013. Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. *Arch Virol* **158**:301–311. <http://dx.doi.org/10.1007/s00705-012-1454-0>.
89. **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. <http://dx.doi.org/10.1101/gr.129684.111>.
90. **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**:e90. <http://dx.doi.org/10.1093/nar/gkq742>.
91. **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. <http://dx.doi.org/10.4161/fly.19695>.
92. **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**:35. <http://dx.doi.org/10.3389/fgene.2012.00035>.
93. **Roy S, Schreiber E.** 2014. Detecting and quantifying low level gene variants in Sanger sequencing traces using the ab1PeakReporter tool. *J Biomol Tech* **25**(Suppl):S13–S14.
94. **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. <http://dx.doi.org/10.1093/nar/gkf436>.
95. **Jukes TH, Cantor CR.** 1969. Evolution of protein molecules, p 21–132. *In* Munro HN (ed), *Mammalian protein metabolism*. Academic Press, New York, NY.
96. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406–425.
97. **Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S.** 2013. MEGA6: molecular evolutionary genetics analysis, version 6.0. *Mol Biol Evol* **30**:2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.
98. **Felsenstein J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791. <http://dx.doi.org/10.2307/2408678>.
99. **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**:5303–5313. <http://dx.doi.org/10.1128/JVI.00312-10>.
100. **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**:528–537. <http://dx.doi.org/10.1016/j.virol.2012.08.043>.