METAGENOMICS FOR PATHOGEN DETECTION DURING A MASS MORTALITY EVENT IN SONGBIRDS

Authors: Mwakibete, Lusajo, Greening, Sabrina S., Kalantar, Katrina, Ahyong, Vida, Anis, Eman, et al.

Source: Journal of Wildlife Diseases, 60(2) : 362-374

Published By: Wildlife Disease Association

URL: https://doi.org/10.7589/JWD-D-23-00109

DOI: 10.7589/JWD-D-23-00109

Journal of Wildlife Diseases, 60(2), 2024, pp. 362–374

Metagenomics for Pathogen Detection During a Mass Mortality Event in Songbirds

Lusajo Mwakibete,[1] Sabrina S. Greening,[2] Katrina Kalantar,[3] Vida Ahyong,[1] Eman Anis,[2,4] Erica A. Miller,[2] David B. Needle,[5] Michael Oglesbee,[6] W. Kelley Thomas,[7] Joseph L. Sevigny,[7] Lawrence M. Gordon,[7] Nicole M. Nemeth,[8,9] C. Brandon Ogbunugafor,[10] Andrea J. Ayala,[10] Seth A. Faith,[6] Norma Neff,[1] Angela M. Detweiler,[1] Tessa Baillargeon,[5] Stacy Tanguay,[5] Stephen D. Simpson,[7] Lisa A. Murphy,[2,4] Julie C. Ellis,[2] Cristina M. Tato,[1] and Roderick B. Gagne[2,11]

[1] Chan Zuckerberg Biohub, San Francisco, California 94158, USA
[2] Department of Pathobiology, Wildlife Futures Program, University of Pennsylvania School of Veterinary Medicine, New Bolton Center, Kennett Square, Pennsylvania 19348, USA
[3] Chan Zuckerberg Initiative, Redwood City, California 94063, USA
[4] Department of Pathobiology, PADLS New Bolton Center, University of Pennsylvania School of Veterinary Medicine, New Bolton Center, Kennett Square, Pennsylvania 19348, USA
[5] New Hampshire Veterinary Diagnostic Lab, University of New Hampshire, Durham, New Hampshire 03824, USA
[6] Infectious Diseases Institute, The Ohio State University, Columbus, Ohio 43210, USA
[7] Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824, USA
[8] Southeastern Cooperative Wildlife Disease Study and Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, Georgia 30602, USA
[9] Department of Pathology, College of Veterinary Medicine, University of Georgia, Georgia 30602, USA
[10] Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511, USA
[11] Corresponding author (email: rgagne@vet.upenn.edu)

ABSTRACT: Mass mortality events in wildlife can be indications of an emerging infectious disease. During the spring and summer of 2021, hundreds of dead passerines were reported across the eastern US. Birds exhibited a range of clinical signs including swollen conjunctiva, ocular discharge, ataxia, and nystagmus. As part of the diagnostic investigation, high-throughput metagenomic next-generation sequencing was performed across three molecular laboratories on samples from affected birds. Many potentially pathogenic microbes were detected, with bacteria forming the largest proportion; however, no singular agent was consistently identified, with many of the detected microbes also found in unaffected (control) birds and thus considered to be subclinical infections. Congruent results across laboratories have helped drive further investigation into alternative causes, including environmental contaminants and nutritional deficiencies. This work highlights the utility of metagenomic approaches in investigations of emerging diseases and provides a framework for future wildlife mortality events.

Key words: Avian, metagenomics, mortality event, rapid response, wildlife.

INTRODUCTION

Wildlife health and diversity are under increasing threats from many sources, with disease emergence in wildlife having the potential to affect the health of humans and domesticated species (Smith et al. 2009; Heard et al. 2013; Cunningham et al. 2017). The development of successful mitigation strategies for emerging infectious diseases in wildlife is often limited by the ability to identify the etiologic agent (Stallknecht 2007). For example, in May 2015, a mass mortality event was observed in central Kazakhstan in which more than half of all saiga antelopes (*Saiga tatarica*) were lost before the identification of the etiologic agent and before any mitigation measures could be implemented (Kock et al. 2018).

Rapid advances in high-throughput sequencing technologies have seen a rise in the number of genomic approaches being applied in disease investigations alongside more traditional techniques such as histopathology, bacterial culture, virus isolation, and PCR tests (Lipkin 2013; Blanchong et al. 2016). One such approach is metagenomic next-generation sequencing (mNGS), a culture-independent untargeted technique that can be used to analyze all nucleic acids (DNA or RNA) within a biological sample. Untargeted approaches, such as mNGS, are unbiased when it comes to capturing all the microbes within a clinical sample,

as the majority of microbes can be identified in the absence of a priori assumption (Simner et al. 2018). This ability is an advantage particularly when the etiologic agent is unknown, and untargeted approaches are increasingly being used to identify pathogenic agents in disease outbreaks affecting humans and livestock (Miller et al. 2013; Greninger et al. 2017; Bohl et al. 2022) although remaining relatively uncommon in wildlife (Zylberberg et al. 2016; Retallack et al. 2019; Fitak et al. 2019; Ko et al. 2022).

Here we highlight the recent use of mNGS to investigate a wildlife mortality event that began in late May 2021 when reports of sick and dead birds were received across the eastern US. Most reports involved nestling and juvenile passerine species, including the Common Grackle (*Quiscalus quiscula*), Blue Jay (*Cyanocitta cristata*), European Starling (*Sturnus vulgaris*), American Robin (*Turdus migratorius*), and Northern Cardinal (*Cardinalis cardinalis*), as well as limited reports of non-passeriform avian species that presented with similar clinical signs (e.g., swollen conjunctiva, crusty ocular discharge, head tilt, ataxia, hind limb paresis, and nystagmus). Several diagnostic laboratories launched investigations focused on identifying an etiologic agent using common diagnostic techniques from across multiple disciplines including pathology, virology, microbiology, parasitology, and toxicology (Richards et al. 2022). Findings from these investigations failed to identify a causative agent but were able to rule out common pathogens and toxicants previously associated with mass avian mortality, including *Salmonella* spp., *Chlamydia* spp., avian influenza viruses, West Nile virus, herpesvirus, *Trichomonas* spp., coccidiosis, and numerous pesticides (USGS 2021). Several *Mycoplasma* spp. were detected in diseased conjunctiva of some affected birds (E. Anis, pers. comm.), but detections were inconsistent, and these bacterial species are commonly detected in nondiseased birds (Sawicka et al. 2020).

To further investigate this event, three diagnostic laboratories, the University of New Hampshire (UNH) Veterinary Diagnostic Laboratory (Durham, New Hampshire, USA) in collaboration with Hubbard Center for Genome Studies (Durham, New Hampshire, USA), the University of Pennsylvania's Wildlife Futures Program (WFP, Kennett Square, Pennsylvania, USA) in collaboration with Chan Zuckerberg Biohub San Francisco (CZ Biohub SF, San Francisco, California, USA), and the Infectious Disease Institute at The Ohio State University (IDI, Columbus, Ohio, USA) undertook mNGS approaches to assist in the detection of a causative agent. Similar mNGS approaches have been previously used in response to mortality events in wild and captive avian species (Pankovics et al. 2018; Papineau 2019; Chang 2021). Here we describe the mNGS approaches undertaken by each laboratory and demonstrate how concurrent approaches can be helpful when investigating the primary cause of a mass mortality event in wildlife.

MATERIALS AND METHODS

Sample collection and processing

During the 2021 mass mortality event in passerines, three labs independently collected samples for mNGS (Fig. 1). In brief, WFP in collaboration with CZ Biohub SF collected whole eye (including conjunctiva) and brain samples from 94 birds including 86 suspected cases and eight controls, plus lung, cloacal bursa, and heart blood from 24 birds (16 cases and eight controls). All suspect cases were selected based on the presence of swollen conjunctiva, eye lesions, and/or crusty ocular discharge. For the WFP samples, suspect cases were all fledglings belonging to one of five species (American Robin, Blue Jay, Common Grackle, Northern Cardinal, and Northern Mockingbird [*Mimus polyglottos*]). Control birds of the same species were sourced from rehabilitation centers in the months following the mortality event (from September to November 2021) where they presented in good nutritional condition with no clinical signs of illness and had died or were euthanized due to acute traumatic injuries (i.e., vehicle collisions or window strikes). Control birds included a range of ages: five adults, five juveniles (1–6 mo old), one fledgling (10–21 d old), and 1 nestling (<10 d old).

In comparison, researchers at UNH collected conjunctiva and ear canal tissue from 103 birds (all suspected cases). These birds were a mixture of adults and fledglings submitted by a number of collaborators including Yale University and the U.S.

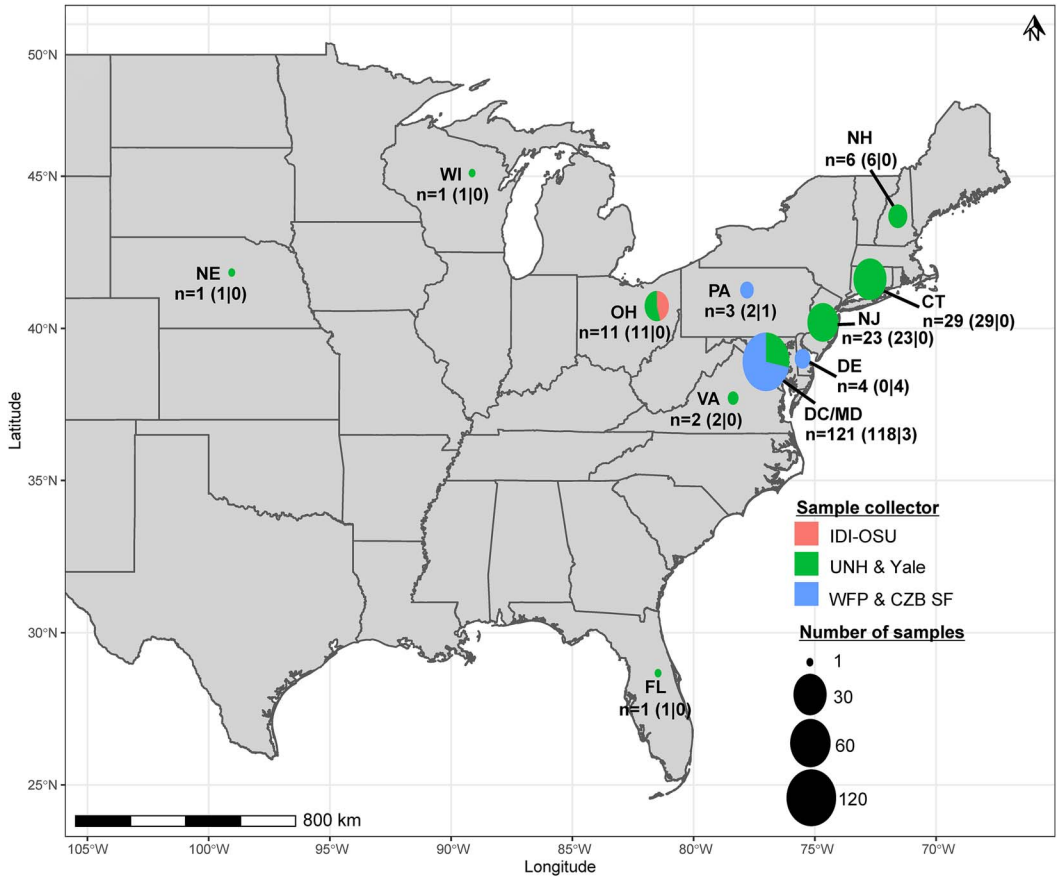

FIGURE 1. Map of the eastern US showing the distribution of sampled birds (both cases and controls) that were included in the mNGS analyses by state (n=number of birds (number of cases | number of controls)) from the 2021 mass mortality event in passerines in the eastern part of the US. The size of the pie chart is proportional to the number of samples, and the color of the pie chart indicates the laboratory to which the samples were sent. WFP: Wildlife Futures Program, University of Pennsylvania; CZ Biohub SF: Chan Zuckerberg Biohub San Francisco; OSU: The Ohio State University; UNH: University of New Hampshire; Yale: Yale University.

Geological Survey - National Wildlife Health Center, with 14 species being identified (American Robin, Blue Jay, Common Grackle, Northern Cardinal, Eastern Phoebe [*Sayornis phoebe*], Rose-breasted Grosbeak [*Pheucticus ludovicianus*], Rusty Blackbird [*Euphagus carolinus*], European Starling, House Finch [*Haemorhous mexicanus*], Tufted Titmouse [*Baeolophus bicolor*], Cooper's Hawk [*Accipiter cooperii*], Eastern Screech Owl [*Megascops asio*], Sharp-shinned Hawk [*Accipiter striatus*], and Mourning Dove [*Zenaida macroura*]) in addition to birds that were not identified to species level but were known to belong to one of the following broader taxonomic groups: finch (*Fringillidae*), pigeon or dove (*Columbidae*), thrush (*Turdidae*), sparrow (*Passeridae*), crow (*Corvidae*), or blackbird (*Icteridae*). The majority of these birds were found dead with only a small number having been euthanized at wildlife rehabilitation facilities. Birds outside the passeriform order were treated as suspect cases using the same criteria as suspect cases within the passeriform order (i.e., swollen conjunctiva, eye lesions, and/or crusty ocular discharge).

Lastly, IDI collected brain tissue from four birds (all suspected cases). These birds were a mixture of adults and fledglings belonging to one of four species (American Robin, Blue Jay, House Sparrow, and Mourning Dove). All birds were found alive with common clinical signs (swollen conjunctiva, ocular exudate, crusty eyes, and ataxia) and either died during transport or were euthanized at wildlife rehabilitation facilities. Additional details

regarding sample collection and processing by each laboratory are provided in the Supplementary Material.

Metagenomic next-generation sequencing

At each diagnostic laboratory, different metagenomic approaches were used for extraction, library preparation, and sequencing as summarized in Table 1 (for details see the Supplementary Material). Some of the major differences between the approaches included the use of different sample types and the extraction of RNA (by WFP-CZ Biohub SF), DNA (by UNH-Yale), or both (by IDI). Following sequencing, the mNGS bioinformatic analysis across the laboratories remained the same with each utilizing the CZ ID metagenomic pipeline: an open-source sequencing analysis platform for identifying microbial sequences within a metagenomic data set, CZ ID (ver. 6.8; Kalantar et al. 2020). The pipeline removes the avian and human host using STAR (Dobin et al. 2013) and Bowtie2 (Langmead and Salzberg 2012), trims adapters using Trimmomatic (Bolger et al. 2014), filters low-quality reads using PriceSeq (Ruby et al. 2013), filters low-complexity sequences using LZW, and identifies duplicate reads using czid-dedup (Morse and Kalantar 2022). The remaining reads are queried on the NCBI nucleotide (NT) and nonredundant protein (NR) databases using GSNAP-L (Wu and Nacu 2010) and RAPSearch2 (Ye et al. 2011), respectively, to determine the microbes (Kalantar et al. 2020).

To account for background contamination, 24 water controls were introduced from the extraction process and used for library preparation (when sequencing the WFP-CZ Biohub SF samples). These controls were selected on CZ ID to create a mass-normalized background model for processing their respective samples. Significant microbial hits were called from the normalized unique reads per million (rPM) that mapped to specific species and genera that passed the following threshold filters: z-score ≥ 1 (to denote significant presence in the sample compared to water background), NT rPM$\geq$10 ($\geq$10 nucleotide reads per million mapping to specific taxa), NR rPM$\geq$5 ($\geq$5 protein reads per million mapping to specific taxa) and average base pair nucleotide alignment$\geq$50 base pairs ($\geq$50 average nucleotide reads alignment mapping to specific taxa). Reads Per Million (rPM) is a scaled metric that allows the relative comparison of microbe abundance within and across samples. The metric is computed by dividing the number of taxon-specific NT or NR reads by the total reads, using the formula

$$rPM = \frac{Reads\ mapping\ to\ taxon\ (NT\ or\ NR\ database)}{(Total\ reads - ERCC\ reads) \times Subsampled\ fraction} \times 10^6.$$

To further increase the validity of the microbial hits, select samples were run on the CZ ID Consensus Genome pipeline (ver. 3.4.7), to assess the genomic coverage and ensure the number of reads was adequate to obtain consensus genomes. An example of this analysis looking at the validity of West Nile virus detected in a single bird is provided in the Supplementary Material.

Microbial composition analysis

To investigate the microbial composition, samples were first filtered to remove background taxa (i.e., those present in water controls) by eliminating taxa with a z-score$\leq$1. The proportion of microbes belonging to the taxonomic categories archaea, bacteria, eukaryotes, or viruses were reported per respective sample, in addition to the two most commonly detected taxa per sample. Further analyses were conducted across the WFP-CZ Biohub SF samples to evaluate differences between the cases and controls. For these analyses, sample reports containing taxonomic relative abundance data for all samples were downloaded from CZ ID and imported into R statistical software (ver. 4.2.1; R Core Team 2018). To investigate the difference in the abundance (NT rPM) of microbes in the eye and brain between the cases and the control group, a Wilcoxon rank-sum test was performed. The significantly differentially abundant microbe genera with P-values$<$0.01 were reported between the groups. Alpha (Simpsons) and beta (Bray Curtis) diversity measures were also calculated using the R package vegan (ver. 2.5; Oksanen et al. 2022) to further investigate the microbial diversity in the eye and brain samples both within and between the case and control groups. The statistical significance in the alpha and beta diversity metrics were evaluated using a Mann-Whitney U-test and Permutational Multivariate Analysis of Variance (PERMANOVA) analysis, respectively.

Data availability

The SRA files of nonhost reads for the WFP-CZ Biohub SF and UNH-Yale samples have been deposited with links to BioProject accession numbers PRJNA909835 and PRJNA961153, respectively, in the NCBI BioProject database (Barrett et al. 2012).

TABLE 1. Metagenomic next-generation sequencing (mNGS) approaches for detection of a potential pathogen taken across three diagnostic laboratories in response to the 2021 mass mortality event in passerines from the eastern US.

Institution	Samples (*n* states, *n* species)	Tissue type (*n* samples)	Extraction kit	Library preparation	Sequencing platform	Mean pair end read per sample
WFP[a] and CZB SF[b]	86 cases (2 states, 5 species)[e] 8 controls (3 states, 4 species)[e]	Eye/ brain (*n*=94) Bursa (*n*=24) Lung (*n*=24) Heart blood (*n*=24)	Quick RNA Pathogen MagBead kit (Zymo Research)	NEBNext Ultra II Library Prep Kit (New England Biolabs)	NovaSeq 6000 at paired 150 bp (Batch 1) and NextSeq 2000 at paired 150 bp (Batch 2)	65.7 million (batch 1) and 5.5 million (batch 2)
IDI-OSU[c]	4 cases (1 state, 4 species)	Brain (*n*=8)[f]	Maxwell RSC Tissue DNA kit (Promega) and Maxwell RSC Simpy RNA kit (Promega)	DNA prep kit (Illumina) and RNA Enrichment prep kit (Illumina)	NextSeq2000 at paired end 100 bp (DNA) and NextSeq 2000 at paired 150 bp (RNA)	58.3 million (RNA) and 16.3 million (DNA)
UNH[d] and Yale	103 cases (10 states, 18 species)	Conjunctiva and ear (*n*=103)	MagMax DNA Multi-Sample Ultra 2.0 kit (Applied Biosystems)	HyperPrep kit (Kapa Biosystems)	NovaSeq 6000 at paired 250 bp	7.9 million

[a] WFP: Wildlife Futures Program, University of Pennsylvania.
[b] CZB SF: Chan Zuckerberg Biohub San Francisco.
[c] IDI-OSU: Infectious Disease Institute at The Ohio State University.
[d] UNH: University of New Hampshire.
[e] Numbers representative for eye/ brain samples only. Lung, bursa, and heart blood were sampled from 16 cases and eight control birds.
[f] Brain samples were taken from the right and left hemispheres in four birds.

RESULTS

Investigating the microbial composition

No single pathogenic microbe was identified across all the cases, with the most commonly detected microbes varying across diagnostic and research laboratories (Table 2). The species-level distribution consisted mainly of bacterial microbes in both cases and controls, with the postfiltering species-level distribution ranging from 43.76–95.67% bacterial (mean 73.08%) across all the samples (Supplementary Material Figs. S1, S2, and S3). In addition to the most commonly detected microbes, the presence of other microbes known to be pathogenic to avian species also varied across laboratories. For instance, across the WFP-CZ Biohub SF eye and brain samples, *Avibacterium* spp. (including *Avibacterium. paragallinarum*, *Avibacterium endocarditidis*, *Avibacterium volantium*, and *Avibacterium avium*) and *Mycoplasma* spp. (including *Mycoplasma gallisepticum*, *Mycoplasma pneumoniae*, and *Mycoplasma mycoides*) were detected in a large proportion of the cases (72.1% and 57.0%, respectively) and controls (25.0% and 12.5%, respectively), whereas neither *Avibacterium* spp. nor *Mycoplasma* spp. were detected in any of the IDI cases, and they were found in only a small proportion from UNH-Yale (21.4% and 6.8%, respectively).

Plasmodium spp. were also detected in 20.4% of all samples collected (9.7% cases and 65.6% controls) by WFP-CZ Biohub SF but in only a limited number of samples across the other laboratories (Table 3). Conversely, other microbes were detected in the UNH-Yale and IDI samples that were not found in the WFP-CZ Biohub SF; for instance, canarypox virus was detected in 7.62% and 22.2% of the samples, respectively. Furthermore, a comparison of the microbes detected in DNA or RNA libraries at IDI revealed that *Burkholderia cenocepacia*, *Bifidobacterium* spp., and *Prevotella melaninogenica* were detected only in DNA libraries, whereas *Cutibacterium acnes* and *Escherichia coli* were detected in both DNA and RNA libraries.

Differences in the microbial taxa between case and control samples of the eye and brain were also detected in the WFP-CZ Biohub SF samples. Specifically, the microbial genera of *Mycoplasma* spp., *Campylobacter* spp., and *Avibacterium* spp. were detected at significantly higher levels ($P<0.01$) in cases (Fig. 2 and Table 4). In the diversity analyses, we observed marginally higher, though insignificant, differences in Simpson's alpha diversity ($P=0.09$) species richness in cases versus controls. Meanwhile, the Bray-Curtis beta diversity was significantly different between cases and controls ($P<0.001$), suggesting distinct microbial profiles across the two groups.

DISCUSSION

Here we describe the metagenomic approaches used to investigate the presence of potential etiologic agent(s) responsible for a mass mortality event in passerines in the eastern US. After concurrent investigations across multiple diagnostic labs using an array of both targeted (USGS 2021) and untargeted approaches, no singular pathogenic microbe was identified that would account for the observed morbidity and mortality in these birds, and to date, the results remain inconclusive.

Given the rapid onset and short time period of the event, it is likely that if a pathogen were the primary driver, it would have been detected in a larger percentage of samples. Variations in detection rates due to factors such as the disease stage at which the cases presented and the tissues collected for sampling are unlikely sufficient to explain the lack of detection of a consistent pathogen across samples. Because of the unpredictable nature of wildlife health events and the reliance on opportunistic sampling, particularly early in the event, some of these factors are easier to control than others. For example, in multispecies events, such as the one reported here, some species may be overrepresented if they thrive in urbanized areas close to humans who can observe the event. The lack of observations from remote regions often results in geographical biases in the samples (Wobeser 2006). Temporal biases are also likely, with samples

TABLE 2. Most frequently detected species taxa per respective sample type and status (i.e., case vs. control) following metagenomic next-generation sequencing (mNGS) bioinformatic analysis using the CZ ID metagenomic pipeline (ver. 6.8) for detection of a potential pathogen in samples from the 2021 mass mortality event in passerines from the eastern US.

Institution	Sample type	Group	Most detected taxa (n, % per respective group)
WFP[a] and CZB SF[b]	Eye and brain	Control (n=8)	*Delftia acidovorans* (n=3, 37.5%)
			Escherichia coli (n=3, 37.5%)
			Pasteurella multocida (n=3, 37.5%)
		Case (n=86)	*Escherichia coli* (n=67, 77.9%)
			Avibacterium paragallinarum (n=53, 61.6%)
			Pasteurrella multocida (n=51, 59.3%)
	Bursa	Control (n=8)	*Besnoitia besnoiti* (n=2, 25.0%)
			Clostridium perfringens (n=2, 25.0%)
			Neospora caninum (n=2, 25.0%)
			Rosa chinensis (n=2, 25.0%)
			Toxoplasma gondii (n=2, 25.0%)
		Case (n=16)	*Enterococcus faecalis* (n=14, 87.5%)
			Escherichia coli (n=13, 81.3%)
	Heart blood	Control (n=8)	*Cyclospora cayetanensis* (n=2, 25.0%)
			Plasmodium spp. (n=2, 25.0%)
			Delftia acidovorans (n=2, 25.0%)
		Case (n=16)	*Clostridium perfringens* (n=7, 43.8%)
			Escherichia spp. (n=7, 43.8%)
	Lung	Control (n=8)	*Cyclospora cayetanensis* (n=3, 37.5%)
			Pasteurella multocida (n=2, 25.0)
		Case (n=16)	*Escherichia coli* (n=13, 81.3%)
			Enterococcus faecalis (n=8, 50.0%)
			Clostridium perfringens (n=8, 50.0%)
UNH[c] and Yale	Conjunctiva and ear	Case (n=103)	*Escherichia coli* (n=49, 47.6%)
			Enterococcus faecalis (n=38, 35.0%)
IDI-OSU[d]	Eye and brain	Case (n=8)	*Cutibacterium acnes* (n=6, 75.0%)
			Bifidobacterium breve (n=4, 50.0%)
			Bifidobacterium longum (n=4, 50.0%)
			Burkholderia cenocepacia (n=4, 50.0%)
			Escherichia coli (n=4, 50.0%)
			Prevotella melaninogenica (n=4, 50.0%)

[a] WFP: Wildlife Futures Program, University of Pennsylvania.
[b] CZB SF: Chan Zuckerberg Biohub San Francisco.
[c] IDI-OSU: Infectious Disease Institute at The Ohio State University.
[d] UNH: University of New Hampshire.

coming from multiple sources and consisting of both morbid animals and animals that were dead for some time, often rendering them suboptimal for diagnostic purposes and raising concerns when comparing samples. Other factors are more easily controlled, such as the tissues selected for sampling, how those tissues are processed, and what diagnostic tests are performed. In our study, whole carcasses or tissues of birds suspected to be involved in the outbreak were evaluated at numerous veterinary and wildlife diagnostic laboratories with the utilization of cross-disciplinary diagnostic techniques, including pathology, virology, microbiology, parasitology, and toxicology (Richards et al. 2022). The tissues collected and protocols used by each group varied according to various factors, including resource limitations, time constraints,

TABLE 3. *Plasmodium spp*. and *Plasmodium relictum* detection across birds following metagenomic next-generation sequencing (mNGS) bioinformatic analysis using the CZ ID metagenomic pipeline (ver. 6.8) for detection of a potential pathogen in samples from the 2021 mass mortality event in passerines from the eastern US.

Institution	Sample type	Group	Microbe	
			Plasmodium spp.	*P. relictum*
WFP[a] and CZB SF[b]	Eye and brain	Control	4	0
		Case	7	1
	Bursa	Control	2	0
		Case	4	1
	Heart blood	Control	6	2
		Case	2	2
	Lung	Control	7	1
		Case	2	1
		Total samples	34	8
		Case birds	7	3
		Control birds	8	2
		Total birds	15	5
UNH[c] and Yale	Conjunctiva and ear	Case	1	0
		Total samples	1	0
		Total birds	1	0
IDI-OSU[d]	Eye and brain	Case	0	0
		Total samples	0	0
		Total birds	0	0

[a] WFP: Wildlife Futures Program, University of Pennsylvania.
[b] CZB SF: Chan Zuckerberg Biohub San Francisco.
[c] IDI-OSU: Infectious Disease Institute at The Ohio State University.
[d] UNH: University of New Hampshire.

and funding. This highlights a need for a minimum set of standards to help guide wildlife investigations and ensure some level of consistency across different working groups. A unified framework would also help facilitate collaboration in large multistate events.

Employing the use of unbiased mNGS for this type of investigation has previously revealed many pathogenic agents, including novel pathogens (Daszak et al. 2000; Michel et al. 2021; Fagre et al. 2022). Metagenomic findings in this study identified several bacterial pathogens significantly more often in cases compared to controls; however, these were deemed unlikely drivers of the mortality event. Importantly, none appeared in a large percentage of samples across groups. In addition, characteristics of the pathology of the specific bacteria identified were inconsistent across birds and with the observed clinical signs. Nevertheless, given the uniqueness of the event, care must be taken in the interpretation of the bacterial pathogens detected. For instance, *Mycoplasma* spp. were found at significantly higher levels in eye and brain samples in cases versus controls. This result might explain the conjunctivitis reported on examination (Fischer et al. 1997); however, *Mycoplasma* spp. are also a common commensal bacterium in many avian species (Sawicka et al. 2020). Thus, it is possible the detection of this microbe was due to opportunistic or subclinical infections rather than being a primary causative agent, particularly given that it was also found in a large percentage of control birds. In addition, *Mycoplasma* spp. were not consistently found across laboratories. Together, these findings suggest that *Mycoplasma* spp. were not primarily responsible for the mortality event.

Avibacterium spp. were also detected at significantly higher levels in the cases than in the controls; however, the clinical signs characteristic of *Avibacterium* spp. infections are more consistent with respiratory disease (i.e., mouth

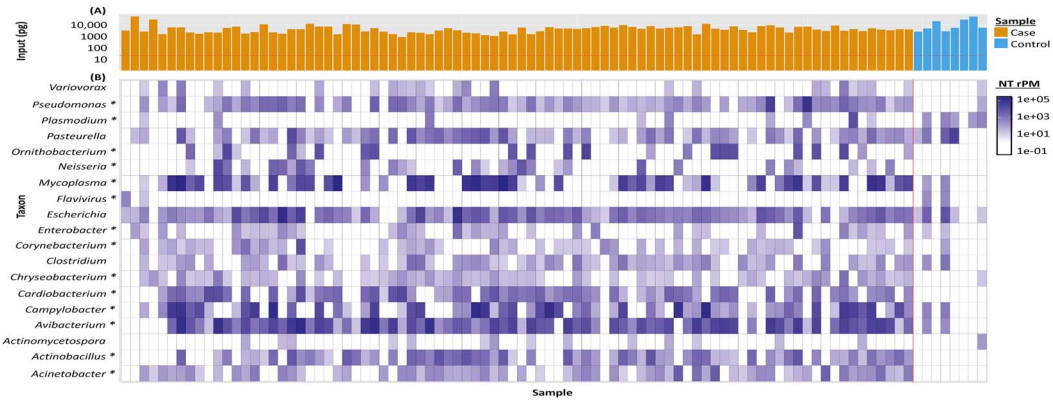

FIGURE 2. Microbes detected by mNGS in eye and brain samples (from the 2021 mass mortality event in passerines from the eastern US) from the the University of Pennsylvania's Wildlife Futures Program–Chan Zuckerberg Biohub San Francisco, sorted by genus. (A) Sample input (pg) in each sample; the color denotes if the sample was derived from a suspected case (orange) or control (blue). (B) Each microbe genus. The asterisk denotes significantly different ($P<0.01$) microbe presence between the cases (low + high) and controls detected in the sample postfiltering.

breathing, swollen sinuses, and nasal discharge; Blackall 1999; Paudel et al. 2017), which were not observed. However, these clinical signs associated with *Avibacterium* spp. infection typically have been described for chickens (*Gallus gallus domesticus*); infections in wild avian species are rarely reported and not well described, and therefore, we cannot be certain that wild avian species would show the same clinical signs. For example, a recent study reported severe periocular swelling, periocular skin crusting, fibrinous sinusitis, and conjunctivitis in wild turkeys infected with a novel clade of *Avibacterium* (Ellis et al. 2022). The presence of multiple pathogenic agents also makes it difficult to determine the contribution of different microbes. For instance, 45.3% of cases for which both eye and brain were assessed were coinfected with *Mycoplasma* and *Avibacterium* spp. This coinfection has also not been well described, and, thus, the resulting clinical signs are unknown; however, without more control cases or experimental infection trials, it is difficult to draw further conclusions. Despite the limited number of controls, the beta diversity analysis did reveal a significant difference in the microbial compositions between the clinical cases and controls, suggesting that the former contained a different microbial profile that may have contributed to the morbidity and mortality.

Although no causative agent for the mortality event was identified, findings highlight the potential of using nontargeted approaches, such as mNGS, to help describe the microbial community circulating in wild populations including both pathogenic and nonpathogenic microbes (Martin et al 2018; Jurburg et al. 2022). For example, the detection of *A. paragallinarum* in this study is important to document as it is known primarily as a respiratory disease affecting chickens and has rarely been described to cause disease in wild avian species (Nsengimana et al. 2022). The differences in background microbes detected between case and control birds may be attributed, in part, to the inherent diversity expected in wildlife stemming from various factors such as habitat, diet, and age. Irrespective of the diversity, we expect to detect similar microbes in the birds that exist in the same locations. Moreover, if the bird mortality was caused by a single pathogenic agent, we would expect to detect the pathogen in the case birds at a higher abundance than the control birds. Generating baseline data and understanding how it changes over time further enables downstream comparisons to be made between diseased and healthy individuals and supports diagnostic responses to future mortality events. For example, the finding of *Mycoplasma* spp. in the control birds highlights the importance of

TABLE 4. Wilcoxon rank test values between cases and controls using nucleotide reads per million (NT rPM) values from eye and brain samples collected by the Wildlife Futures Program, University of Pennsylvania, and Chan Zuckerberg Biohub San Francisco to determine microbial difference between cases and controls from the 2021 mass mortality event in passerines from the eastern US.

Microbe genus	*P* value
Acinetobacter[a]	9.38E-06
Actinobacillus[a]	5.78E-05
Actinomycetospora	1.14E-05
Avibacterium[a]	8.04E-05
Campylobacter[a]	4.15E-03
Cardiobacterium[a]	6.51E-05
Chryseobacterium[a]	9.78E-05
Clostridium	3.88E-02
Corynebacterium	1.55E-03
Enterobacter[a]	9.09E-03
Escherichia	2.16E-02
Flavivirus[a]	2.75E-06
Mycoplasma[a]	1.14E-03
Neisseria[a]	8.32E-04
Ornithobacterium[a]	1.09E-04
Pasteurella	5.83E-01
Plasmodium[a]	4.16E-04
Pseudomonas[a]	8.28E-05
Variovorax	1.06E-02

[a] *P* value<0.01.

using controls to help elucidate potential causes of disease in wildlife (Ryser-Degiorgis 2013; Van Hemert et al. 2014). Further, with the added health challenge posed by continual changes in the environments in which these birds live (e.g., landscape, climate, accumulation of potentially toxic substances), baseline microbes might be used as an indicator of the changing health status of an animal population (Sun et al. 2022).

Though mNGS is a powerful tool for detecting potential microbial pathogens, variability in experimental protocols, such as the sampling procedures, processing, and data analyses, may lead to artifacts that result in incorrect conclusions and false negative detections (Wooley and Ye 2009; Teeling and Glöckner 2012). Each of the three laboratories in this study used a different experimental design, which enabled us to compare results and make recommendations on the best practices for conducting an mNGS investigation in response to a wildlife mass mortality event. We suggest a framework (Fig. 3) to assist in addressing some of the limitations of this study for future explorations.

First, developing a strong case definition is pivotal to help guide sample selection. In the early stages of the investigation, a case definition may be primarily based on the range of clinical signs present in the affected population; however, as more cases are identified it is important to revisit the case definition and include information regarding clinical, laboratory, and pathologic characteristics, as well as information on the affected individuals (e.g., species, age, etc.) and any geographical or temporal characteristics. A strong case definition also helps in the selection of high-quality controls from the same population as the cases to provide background microbiota data from healthy specimens when performing downstream comparative analyses. Controls should be selected only if their cause of death is known and unrelated (e.g., they show no clinical or pathological signs of disease, or are from a different area or population). Water controls during sample extraction should also be implemented to remove background contamination during the library preparation, as well as water controls that may be used to spot contamination that may occur during sample collection and processing.

During sample collection, using a nucleic acid stabilizer is crucial to assist in retaining the total nucleic acid integrity and inactivating any infectious agents. We suggest starting with RNA libraries for the initial pass to reveal all actively replicating microbes within the samples that may be contributing to an active infection, while also allowing the capture of RNA viruses. However, if the RNA libraries yield no results, DNA libraries can be performed. For microbe detection in mNGS data, several steps can be taken to help increase the confidence in microbe detection further downstream in the analysis, including the implementation of threshold filters to help validate and increase confidence in the presence of detected pathogens, incorporating controls to compare microbial profiles and abundances between the case and control groups, and

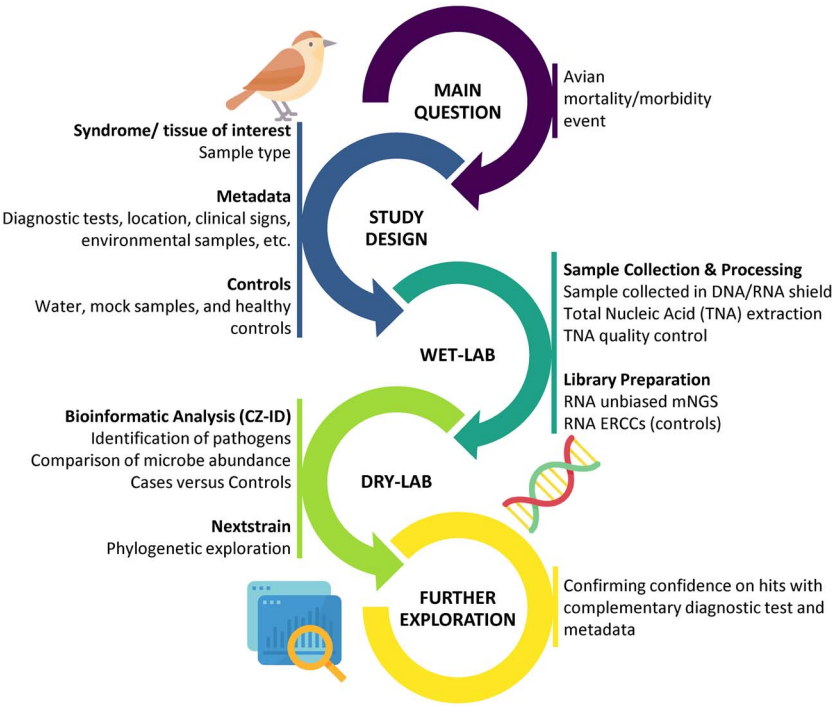

FIGURE 3. Potential framework for metagenomic next-generation sequencing (mNGS) in wild birds.

utilizing metadata to aid in the logical assessment of potentially pathogenic microbes that have been detected. It is also important to consider that any reference-based assembly might miss novel pathogens currently unavailable in the reference database, and further interrogation of the reads may be required (Nielsen et al. 2014; Yang et al. 2021).

Our suggested framework for mNGS approaches (Fig. 3) helps ensure that steps have been taken to minimize artifacts in the data. Our parallel analyses across three diagnostic laboratories revealed no single pathogen associated with the 2021 mass mortality event in passerines. Findings suggest that the underlying mortality is not due to pathogenic microorganisms and have guided the investigation to refocus time and resources on other potential factors, such as dietary deficiencies, to explain the mortality event.

ACKNOWLEDGMENTS

We are grateful to all wildlife rehabilitators, biologists, veterinarians and veterinary staff, volunteers, members of the public, and responding state wildlife agencies who assisted with sample collection. We thank the U.S. Geological Survey - National Wildlife Health Center (Madison, Wisconsin) for contributing samples used in this project. We also want to extend our thanks to the CZ Biohub SF Genomics team for the sequencing support, the Wildlife Futures team for sampling assistance, and the faculty and staff at the Pennsylvania Animal Diagnostic Laboratory System–New Bolton Center who contributed to the sample collection, diagnostics, and shipping. S.S.G. was supported by the Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation; A.J.A. was supported by an NSF Postdoctoral Fellowship in Biology (Award no. 2010904). UNH sequencing was funded through the USDA NIFA NH00694 awarded to W. K. Thomas.

SUPPLEMENTARY MATERIAL

Supplementary material for this article is online at http://dx.doi.org/10.7589/JWD-D-23-00109.

LITERATURE CITED

Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, et al. 2012. BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Res* 40:D57–63.

Blackall PJ. 1999. Infectious coryza: Overview of the disease and new diagnostic options. *Clin Microbiol Rev* 12:627–632.

Blanchong JA, Robinson SJ, Samuel MD, Foster JT. 2016. Application of genetics and genomics to wildlife epidemiology. *J Wildl Manage* 80:593–608.

Bohl JA, Lay S, Chea S, Ahyong V, Parker DM, Gallagher S, Fintzi J, Man S, Ponce A, et al. 2022. Discovering disease-causing pathogens in resource-scarce Southeast Asia using a global metagenomic pathogen monitoring system. *Proc Natl Acad Sci U S A* 119:e2115285119.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Chang WS. 2021. *Metagenomic applications in virus discovery, ecology, and the surveillance of Australian wildlife*. PhD Dissertation, University of Sydney, Sydney, Australia.

Cunningham AA, Daszak P, Wood JLN. 2017. One Health, emerging infectious diseases and wildlife: Two decades of progress? *Philos Trans R Soc Lond B Biol Sci* 372: e20160167.

Daszak P, Cunningham AA, Hyatt AD. 2000. Emerging infectious diseases of wildlife—Threats to biodiversity and human health. *Science* 287:443–449.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Ellis JS, MacGlover CA, Sondgeroth KS, Brown D, Daniels JB, Fox KA. 2022. Novel *Avibacterium* species associated with sinusitis and conjunctivitis in a Merriam's wild turkey (*Meleagris gallopavo merriami*) flock in Colorado, USA. *J Wildl Dis* 58:725–734.

Fagre AC, Cohen LE, Eskew EA, Farrell M, Glennon E, Joseph MB, Frank HK, Ryan SJ, Carlson CJ Albery GF. 2022. Assessing the risk of human-to-wildlife pathogen transmission for conservation and public health. *Ecol Lett* 25:1534–1549.

Fischer JR, Stallknecht DE, Luttrell P, Dhondt AA, Converse KA. 1997. Mycoplasmal conjunctivitis in wild songbirds: The spread of a new contagious disease in a mobile host population. *Emerg Infect Dis* 3:69–72.

Fitak RR, Antonides JD, Baitchman EJ, Bonaccorso E, Braun J, Kubiski S, Chiu E, Fagre AC, Gagne RB, et al. 2019. The expectations and challenges of wildlife disease research in the era of genomics: Forecasting with a horizon scan-like exercise. *J Hered* 110:261–274.

Greninger AL, Waghmare A, Adler A, Qin X, Crowley JL, Englund JA, Kuypers JM, Jerome KR, Zerr DM. 2017. Rule-out outbreak: 24-hour metagenomic next-generation sequencing for characterizing respiratory virus source for infection prevention. *J Pediat Infect Dis Soc* 6:168–172.

Heard MJ, Smith KF, Ripp K, Berger M, Chen J, Dittmeier J, Goter M, McGarvey ST, Ryan E. 2013. The threat of disease increases as species move towards extinction. *Conserv Biol* 27:1378–1388.

Jurburg SD, Buscot F, Chatzinotas A, Chaudhari NM, Clark AT, Garbowski M, Grenié M, Hom EFY, Karakoc C, et al. 2022. The community ecology perspective of omics data. *Microbiome* 10:225-238.

Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J, Holmes OB, Juan YF, et al. 2020. IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 9:giaa111.

Ko KKK, Chng KR, Nagarajan N. 2022. Metagenomics-enabled microbial surveillance. *Nat Microbiol* 7:486–496.

Kock RA, Orynbayev M, Robinson S, Zuther S, Singh NJ, Beauvais W, Morgan ER, Kerimbayev A, Khomenko S, et al. 2018. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv* 4:eaao2314.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.

Lipkin WI. 2013. The changing face of pathogen discovery and surveillance. *Nat Rev Microbiol* 11:133–141.

Martin TC, Visconti A, Spector TD, Falchi M. 2018. Conducting metagenomic studies in microbiology and clinical research. *Appl Microbiol Biotechnol* 102:8629–8646.

Michel AL, Van Heerden H, Crossley BM, Al Dahouk S, Prasse D, Rutten V. 2021. Pathogen detection and disease diagnosis in wildlife: Challenges and opportunities. *Rev Sci Tech* 40:105–118.

Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. 2013. Metagenomics for pathogen detection in public health. *Genome Med* 5:81.

Morse T, Kalantar K. 2022. *czid-dedup*. https://github.com/chanzuckerberg/czid-dedup. Accessed July 2023.

Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, et al. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 32:822–828.

Nsengimana O, Habarugira G, Ojok L, Ruhagazi D, Kayitare A, Shyaka A. 2022. Infectious coryza in a grey crowned crane (*Balearica regulorum*) recovered from captivity. *Vet Med Sci* 8:822–826.

Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, et al. 2022. *vegan: Community ecology package*. R package version 2.6-4. https://CRAN.R-project.org/package=vegan. Accessed June 2022.

Pankovics P, Boros Á, Phan TG, Delwart E, Reuter G. 2018. A novel passerivirus (family *Picornaviridae*) in an outbreak of enteritis with high mortality in estrildid finches (*Uraeginthus* sp.). *Arch Virol* 163:1063–1071.

Papineau A. 2019. *Targeted enrichment and viral metagenomics in the detection of livestock and wildlife viruses*. MSc Thesis, University of Manitoba, Winnipeg, Canada.

Paudel S, Ruhnau D, Wernsdorf P, Liebhart D, Hess M, Hess C. 2017. Presence of *Avibacterium paragallinarum* and histopathologic lesions corresponds with clinical signs in a co-infection model with *Gallibacterium anatis*. *Avian Dis* 61:335–340.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed June 2022.

Retallack H, Okihiro MS, Britton E, Van Sommeran S, DeRisi JL. 2019. Metagenomic next-generation sequencing reveals *Miamiensis avidus* (*Ciliophora: Scuticociliatida*) in the 2017 epizootic of leopard sharks (*Triakis semifasciata*) in San Francisco Bay, California, USA. *J Wildl Dis* 55:375–386.

Richards BJ, Nemeth NM, Lewis NJ, Needle DB, Sevigny JL, Oglesbee M, Faith SA, Greening SS, Ellis JC. 2022. The 2021 songbird mortality event—A summary of findings and discussion of ongoing investigations. In: *Proceedings of the 70th Annual International Wildlife Disease Association Conference*, 23–19 July 2022, Madison, Wisconsin, USA, p. 71.

Ruby JG, Bellare P, Derisi JL. 2013. PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)* 3:865–880.

Ryser-Degiorgis MP. 2013. Wildlife health investigations: Needs, challenges and recommendations. *BMC Vet Res* 9:223.

Sawicka A, Durkalec M, Tomczyk G, Kursa O. 2020. Occurrence of *Mycoplasma gallisepticum* in wild birds: A systematic review and meta-analysis. *PLoS One* 15:e0231545.

Simner PJ, Miller S, Carroll KC. 2018. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis* 66:778–788.

Smith KF, Acevedo-Whitehouse K, Pedersen AB. 2009. The role of infectious diseases in biological conservation. *Anim Conserv* 12:1–12.

Stallknecht DE. 2007. Impediments to wildlife disease surveillance, research, and diagnostics. *Curr Top Microbiol Immunol* 315:445-461.

Sun F, Chen J, Liu K, Tang M, Yang Y. 2022. The avian gut microbiota: Diversity, influencing factors, and future directions. *Front Microbiol* 13:e934272.

Teeling H, Glöckner FO. 2012. Current opportunities and challenges in microbial metagenome analysis—A bioinformatic perspective. *Brief Bioinform* 13:728–742.

USGS (US Geological Survey). 2021. *Update for the Association of Fish and Wildlife Agencies from the USGS National Wildlife Health Center*. USGS National Wildlife Health Center. Madison, Wisconsin. https://www.usgs.gov/media/files/update-association-fish-and-wildlife-agencies-sept-2021. Accessed June 2022.

Van Hemert C, Pearce JM, Handel CM. 2014. Wildlife health in a rapidly changing north: Focus on avian disease. *Front Ecol Environ* 12:548–556.

Wobeser GA. 2006. *Essentials of disease in wild animals*. Blackwell Publishing, Ames, Iowa, 390 pp.

Wooley JC, Ye Y. 2009. Metagenomics: Facts and artifacts, and computational challenges. *J Comput Sci Technol* 25:71–81.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881.

Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian Z, Zhang L. 2021. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput Struct Biotechnol J* 19:6301–6314.

Ye Y, Choi JH, Tang H. 2011. RAPSearch: A fast protein similarity search tool for short reads. *BMC Bioinformatics* 12:159.

Zylberberg M, Van Hemert C, Dumbacher JP, Handel CM, Tihan T, DeRisi JL. 2016. Novel picornavirus associated with avian keratin disorder in Alaskan birds. *mBio* 7:e00874-16.

Submitted for publication 12 July 2023.
Accepted 2 January 2024.