

---

# Mass Die-Off of Saiga Antelopes, Kazakhstan, 2015

Sasan Fereidouni, Graham L. Freimanis, Mukhit Orynbayev, Paolo Ribeca, John Flannery, Donald P. King, Steffen Zuther, Martin Beer, Dirk Höper, Aidyn Kydrymanov, Kobey Karamendin, Richard Kock

In 2015, a mass die-off of ~200,000 saiga antelopes in central Kazakhstan was caused by hemorrhagic septicemia attributable to the bacterium *Pasteurella multocida* serotype B. Previous analyses have indicated that environmental triggers associated with weather conditions, specifically air moisture and temperature in the region of the saiga antelope calving during the 10-day period running up to the event, were critical to the proliferation of latent bacteria and were comparable to conditions accompanying historically similar die-offs in the same areas. We investigated whether additional viral or bacterial pathogens could be detected in samples from affected animals using 3 different high-throughput sequencing approaches. We did not identify pathogens associated with commensal bacterial opportunisms in blood, kidney, or lung samples and thus concluded that *P. multocida* serotype B was the primary cause of the disease.

The saiga antelope (*Saiga tatarica tatarica* and *S.t. mongolica*) is a critically endangered species (1) with populations located in Kazakhstan in addition to small remnants in Russia and Uzbekistan and a subspecies in Mongolia. Each year during the month of May, Saiga antelopes gather in Kazakhstan for calving. Mass die-offs in their populations have been reported previously and were attributed to viral and bacterial etiologies, including pasteurellosis (2). However, the diagnosis in most of these events has been unreliable because of insufficient fresh sampling and diagnostic work (2).

During a large outbreak in 2015, extensive diagnostics and environmental studies were undertaken, subject to restricting factors such as remoteness and limited cold chain

Author affiliations: University of Veterinary Medicine Vienna, Vienna, Austria (S. Fereidouni); The Pirbright Institute, Pirbright, UK (G.L. Freimanis, P. Ribeca, J. Flannery, D.P. King); Research Institute for Biological Safety Problems, Otar, Kazakhstan (M. Orynbayev); Association for the Conservation of Biodiversity of Kazakhstan, Astana, Kazakhstan (S. Zuther); Frankfurt Zoological Society, Frankfurt, Germany (S. Zuther); Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany (M. Beer, D. Höper); Institute of Microbiology and Virology, Almaty, Kazakhstan (A. Kydrymanov, K. Karamendin); Royal Veterinary College, London, UK (R. Kock)

DOI: <https://doi.org/10.3201/eid2506.180990>

resources. Annual disease monitoring in saiga antelopes had been established after die-offs occurred in western Kazakhstan in 2010, and an international multidisciplinary research team was on the ground at the time of the die-off, performing routine surveillance (3,4).

The mass die-off of saiga antelopes in Kazakhstan started around May 10, 2015, and caused ~200,000 deaths across several calving groups within 3 weeks. These subgroups of saiga antelopes were spread discretely across a landscape of several hundreds of thousands of square kilometers. The number of dead animals constituted more than two thirds of the global population of saiga antelope at the time. The outbreak wiped out 88% of the Betpak-Dala population in central Kazakhstan (5) and appeared to have a 100% case-fatality rate.

Laboratory results on the microbiologic, pathologic, and environmental conditions at the time of the 2015 outbreak suggested hemorrhagic septicemia caused by *Pasteurella multocida* serotype B and triggered by environmental conditions (3,6). However, whether a second unknown infectious agent had predisposed the animals to infection with *P. multocida* was unclear from the laboratory results. Given the opportunistic nature of *Pasteurella*, the objective of our study was to attempt to identify whether any additional unknown potential causative pathogens were present in samples (taken from 10 animals) that might have contributed to the die-off.

## Materials and Methods

### Field Assessment

The first dead animals were detected in the Amangeldy District (Kostanay region) of Kazakhstan on May 10, 2015, and additional die-offs were recorded in unconnected discrete locations in the Aktobe and Akmola regions (3). A primary diagnosis of hemorrhagic septicemia as the cause of death was proposed at the sites on the basis of clinical signs and gross pathology. We took FTA papers of whole blood spots from 8 freshly dead, female animals (Table 1) in a 2-km radius on the last 2 days of the operation and sent them to international reference laboratories for high-throughput sequencing (HTS) protocols. FTA cards were

**Table 1.** Details outlining the origins of the 8 FTA samples, including animals and GPS data, used in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015\*

Sample no.	Species	Age, y/sex	Comment	Sample type	GPS no.	Date
1	<i>Saiga tatarica</i>	3–4/F	Postmortem	FTA x2	427	2015 May 26
2	<i>Saiga tatarica</i>	3/F	Postmortem	FTA x2	426	2015 May 26
3	<i>Saiga tatarica</i>	1–2/F	Postmortem	FTA x2	456	2015 May 26
4	<i>Saiga tatarica</i>	1–2/F	Postmortem	FTA x2	452	2015 May 26
5	<i>Saiga tatarica</i>	5–6/F	Postmortem	FTA x2	457	2015 May 26
6	<i>Saiga tatarica</i>	>5/F	Postmortem	FTA x2	458	2015 May 26
7	<i>Saiga tatarica</i>	2/F	Postmortem	FTA x2	455	2015 May 26
8	<i>Saiga tatarica</i>	13/F	Postmortem	FTA x2	NA	2015 Jun 25

\*GPS, global positioning system; NA, no information available.

used as backup given the limited resources available and difficulties in maintaining cold chain and in transportation of fresh samples to local laboratories. Lung and kidney tissue from 2 dead saiga antelopes (lung tissue from animal X and kidney tissue from animal Y) from the Turgai River region were also processed for 16S metagenomics sequencing in the city of Almaty, Kazakhstan. Although these samples were from a region 175 km from the site where the FTA card samples were taken, they were considered part of the same saiga antelope population. Given the uniformity of the clinical syndrome and consistency of the pathogenesis, the sample of cases selected was small relative to the scale of the die-off, but each case was evaluated in considerable depth and considered representative of the affected population on the basis of the consistent pathology and disease characteristics observed in all the affected animals (3).

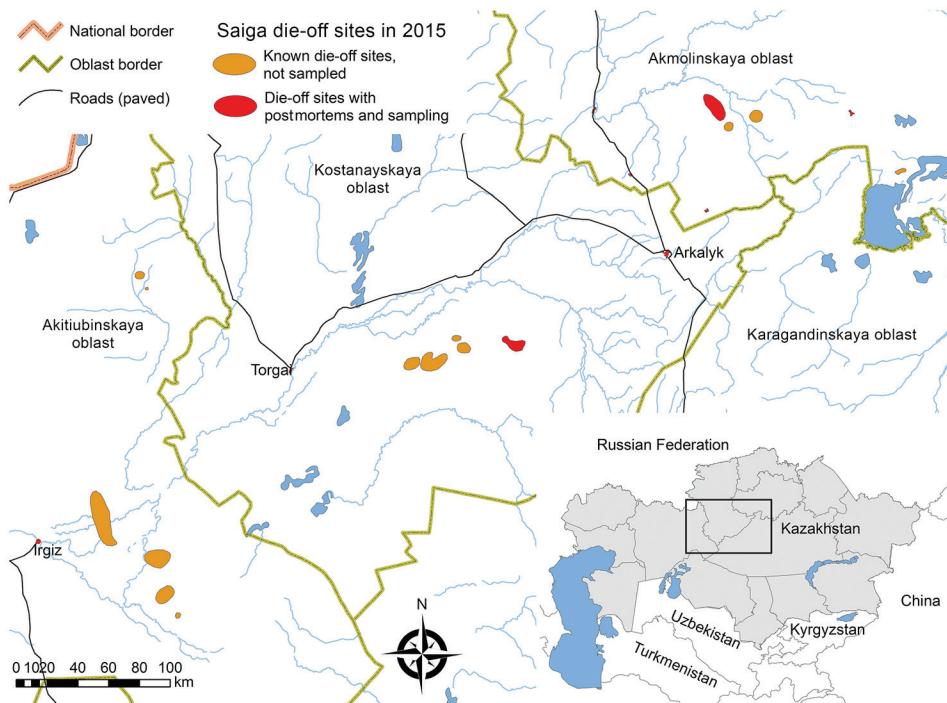
### Laboratory Assessment

We submitted samples of dried blood spots (2 cm in diameter) on FTA papers taken from 8 animals to 2 different

research institutions (the Pirbright Institute in the United Kingdom and the Friedrich-Loeffler-Institut [FLI] in Germany) for HTS analyses (Figure 1) under 2 different HTS protocols (random amplification–based sequencing at Pirbright and RNA sequencing at FLI). Six of the 8 FTA blood spot samples were processed for further testing by using HTS at Pirbright, and 4 of the 8 samples were processed for further testing at FLI. Two of the 8 samples were processed by both laboratories. Lung and kidney tissue from 2 dead saiga antelopes (Table 2) were tested for 16S bacterial diversity by using a 16S metagenomic sequencing protocol developed by the Institute of Microbiology and Virology in Almaty (Figure 2).

### Results

We analyzed reads from each of the parallel investigations by using established bioinformatics pipelines to identify microbial agents present within each sample. All raw datasets, de novo assemblies, and 16S sequencing metagenome data sets have been submitted to the European Nucleotide Archive



**Figure 1.** Geographic distribution of saiga antelope die-off events, Kazakhstan, 2015. Red and orange areas indicate known outbreak locations of the 3 saiga populations. Inset shows area in relation to the rest of Kazakhstan and neighboring countries of central Asia.

**Table 2.** Characteristics of fresh tissue samples transferred to Almaty for 16S ribosomal profiling used in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015\*

Animal	Date	GPS	Species	Age y/sex	Sample used for HTS
Animal X	2015 May 16	49°46.586N/ 65°26.369E	<i>Saiga tatarica</i>	2/F	Lung
Animal Y	2015 May 19	49°45.001N/065°27.536E	<i>S. tatarica</i>	3/F	Kidney

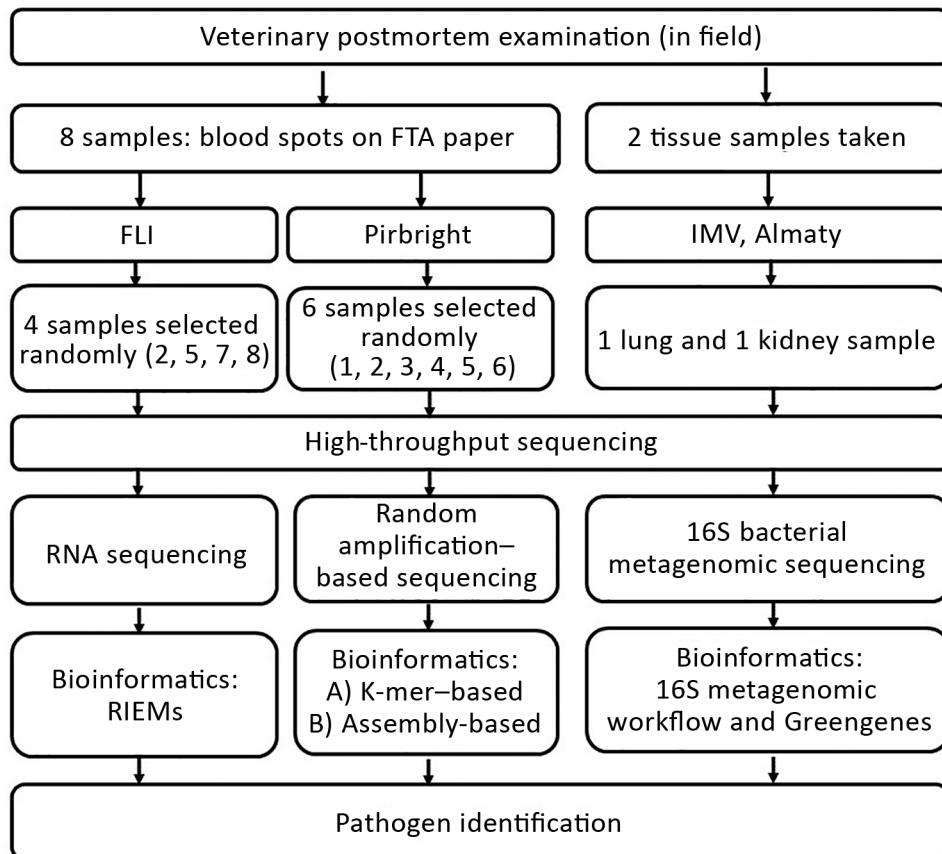
\*HTS, high-throughput sequencing.

and GenBank (accession nos. PRJEB28164, PRJEB28184, and PRJNA486600).

**Random Amplification–Based Sequencing Protocol**

The Pirbright protocol consisted of a random amplification workflow, with libraries sequenced using the MiSeq System (Illumina, <https://www.illumina.com>) to identify microbial nucleic acids present in dried peripheral blood spots. The classification of sequenced reads into different taxonomic groups was conducted by using 2 approaches; the first was a k-mer–based approach that assigned each read independently (Table 3; Appendix Table 1, <https://wwwnc.cdc.gov/EID/article/25/6/18-0990-App1.pdf>), and the second was a de novo approach that first assembled reads into contigs and then assigned contigs (Table 4; Appendix Tables 2–4). Approximately 72% of the original reads mapped to the assembly produced by the de novo approach. Neither of these 2 approaches conclusively identified a single virus as a causative agent in all samples. In all 6 samples

tested, 46.4% (geometric mean [GM]) of reads were unclassified against the Mini-Kraken Database (<https://ccb.jhu.edu/software/kraken>) and were possibly host-derived; these reads were labeled as unclassified (Table 3). This determination was further supported by the de novo analysis, which identified the largest contigs as being host-derived, having 35% of reads accounting for host material. In terms of microbial organisms, in 6 of 6 samples, the largest numbers of hits (GM 39%) were identified as *Pasteurella* spp. The specificity of this finding was increased for the *P. multocida* genome, which exhibited the greatest number of matches (GM 69,760 hits/sample [35.4%]). The microbial organism with the second highest number of hits in all samples was *Alteromonas macleodii* (GM 215 hits/sample [0.11%]). K-mers present in all 6 samples also aligned with the *Achromobacter xylosoxidans* (GM 102 hits/sample), *Haemophilus* spp. (GM 40 hits/sample), *Mannheimia haemolytica* (GM 13 hits/sample), *Klebsiella* spp. (GM 23 hits/sample), and *Aggregatibacter* spp. (GM 16 hits/sample),



**Figure 2.** Outline of the process of sampling and high-throughput sequencing protocols performed at 3 research institutes in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015. FLI, Friedrich-Loeffler-Institut; IMV, Institute of Microbiology and Virology.

**Table 3.** Main results of the k-mer–based approach on the random amplification metatranscriptomic dataset used in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015\*

Organism	No. reads (% total reads)					
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Total no. reads	231,907	773,835	272,102	300,807	235,255	187,049
Total no. reads passing QC	109,302 (47.1)	553,163 (71.48)	174,613 (64.17)	233,888 (77.8)	171,409 (72.86)	138,292 (73.93)
Unclassified/nonmicrobial	50,478 (46.18)	343,404 (62.08)	86,165 (49.35)	133,456 (57.06)	57,812 (33.73)	50,731 (36.68)
Virus	47 (0.03)	141 (0.03)	33 (0.03)	174 (0.05)	73 (0.04)	51 (0.04)
<i>Pasteurellaceae</i>	53,097 (48.58)	129,337 (23.38)	69,251 (39.66)	63,817 (27.29)	93,378 (54.48)	72,799 (52.64)
<i>Pasteurella multocida</i>	49,844 (45.6)	115,231 (20.83)	60,504 (34.65)	56,775 (24.27)	86,664 (50.56)	67,406 (48.74)
<i>Alteromonadales</i>	1303 (1.19)	2690 (0.49)	35 (0.02)	499 (0.21)	32 (0.02)	51 (0.04)
<i>Enterobacteriaceae</i>	52 (0.05)	208 (0.04)	153 (0.09)	112 (0.05)	77 (0.04)	48 (0.03)
<i>Haemophilus</i>	23 (0.02)	77 (0.01)	31 (0.02)	36 (0.02)	63 (0.04)	37 (0.03)
<i>Betaproteobacteria</i>	86 (0.08)	13160 (2.38)	30 (0.02)	86 (0.04)	18 (0.01)	22 (0.02)
<i>Mannheimia</i>	10 (0.0)	18 (0.0)	12 (0.01)	18 (0.01)	19 (0.01)	9 (0.01)
<i>Aggregatibacter</i>	8 (0.0)	37 (0.0)	13 (0.0)	18 (0.0)	16 (0.0)	18 (0.01)
<i>Klebsiella</i>	11 (0.0)	58 (0.01)	16 (0.0)	54 (0.01)	16 (0.0)	14 (0.0)

\*Only organisms that were identified in all samples and with >10 reads are listed. Samples 2 and 5 were also tested at Friedrich-Loeffler-Institut by using an RNA sequencing protocol. QC, quality control.

albeit to a lower level than the top 2 results. All the other organisms tested had matches of <10 hits and were not considered statistically significant.

The de novo analysis approach also did not identify any homologies with unexpected viral genomes. Several bacteria were identified by both k-mer and de novo analysis protocols, including *P. multocida* and *M. haemolytica*, although these bacteria had smaller contigs (182 bp and 85 bp, respectively) and were thus not included in the results (Table 4). We were unable to identify 195 contigs produced by the de novo assembly despite using several BLAST databases (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). We subjected these contigs to an extended analysis in which they were first aligned to BLAST databases with tblastx to find similarities at the protein level. That analysis generated matches for 35 contigs; the distribution of the matches in terms of species mirrors quite closely the one found by nucleotide BLAST (Appendix Tables 2–4). Separately, we also translated and subjected the unknown contigs to a search using SUPERFAMILY (<http://supfam.org>). This approach returned hits for 87 contigs (Appendix Tables 2–4), of which most appeared to be homologs of bacterial

proteins. No further pathogens could be conclusively identified using this analysis. Further analysis of the assembled sequences attributed to *P. multocida* did not permit accurate conclusions to be drawn because of the fragmented nature of the contigs.

#### RNA Sequencing Protocol

By using the RIEMS analysis pipeline, we performed taxonomic analysis of the sequencing reads obtained from libraries generated from RNA extracted from 4 blood spots from FTA cards that had been transcribed into cDNA using random hexamer priming followed by shotgun library preparation. Of the samples tested at FLI, samples 2 and 5 were also tested at Pirbright. Overall, these analyses classified 77.9%–93.5% of the reads as *P. multocida* (Table 5). The remainder mainly represented host sequences (0.9%–16.2%). With a few exceptions for phage reads, no reads were classified as being of viral origin, which was concordant with findings of the Pirbright dataset. In all samples, the proportion of reads remaining unclassified after analysis of the nucleic acid sequences was low (0.42%–0.45%); these unclassified reads had median lengths of

**Table 4.** Main results obtained using a de novo approach on the random amplification meta-transcriptomic dataset used in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015\*

Read area	No. contigs†	Total length, bp	Attribution‡	Comment
796	2	271	<i>Pasteurella bettyae</i> CCUG 2042	
1,758,115	162	27,999	<i>Ovis canadensis canadensis</i>	Host
1,676,355	23	5,780	<i>Capra hircus</i> (goat)	
36,795	7	1,287	<i>Bubalus bubalis</i> (water buffalo)	
30,763	14	1,959	<i>Bos taurus</i> (cattle)	
3,252	6	1,366	<i>Saiga tatarica</i>	
2,625	8	1,283	<i>Ovis aries</i> (sheep)	
2,414	5	969	<i>Bos indicus</i>	
1,650	2	317	<i>Eudorcas thomsonii</i> (Thomson's gazelle)	
14,221,307	6,641	2,103,430	<i>Pasteurella multocida</i>	Other
69,009	195	27036	Unknown sequence	
35,246	1	401	Uncultured eukaryote	
796	2	271	<i>Pasteurella bettyae</i> CCUG 2042	

\*Equivalent to the contig length × the average read coverage.

†Number of contigs with the same attribution.

‡As determined by the best blastn hit.

**Table 5.** Summary of the most relevant results obtained by RIEMS analyses of the datasets (sequenced from shotgun libraries generated from random primed cDNA) used in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015

Organism	No. (%) reads			
	Sample 2* (lib01416)	Sample 5* (lib01417)	Sample 7 (lib01418)	Sample 8 (lib01419)
Input reads	411,640 (100)	376,210 (100)	372,387 (100)	354,958 (100)
Quality filtered reads†	12,786 (3.1)	10,793 (2.9)	11,559 (3.1)	10,895 (3.1)
Unclassified reads‡	1,776 (0.43)	1,520 (0.40)	1,626 (0.44)	1,494 (0.42)
Classified reads†	397,078 (96.5)	363,897 (96.7)	359,202 (96.5)	342,569 (96.5)
Host‡	64,618 (16.3)	4,770 (1.3)	3,414 (1.0)	4,784 (1.4)
<i>Pasteurellaceae</i> ‡	317,009 (79.8)	345,893 (95.1)	339,484 (94.5)	324,770 (94.8)

\*Samples 2 and 5 were also tested by using high-throughput sequencing at the Pirbright Institute.

†Percentage is of the number of input reads.

‡Percentage is of the number of classified reads.

35 bp (interquartile range 25–45 bp) and accounted for 0.068%–0.082% of the total bases. Therefore, the information content of the unclassified portion of the datasets was too low to provide additional information even by additional analyses on the basis of the amino acid sequences deduced from these reads.

To conduct a detailed analysis of the numerous *P. multocida* organisms detected, we mapped the complete datasets along the *P. multocida* genome sequence (GenBank accession no. NC\_002663.1). We then performed blastx (7) analyses of the resulting contigs for a basic function prediction of the expressed genes. Besides detecting genes encoding proteins of gene expression, general metabolism, and cell division, these analyses detected several proteins associated with pathogenicity. For example, proteins facilitating active iron uptake (iron ABC transporter permease [GenBank accession no. WP\_010906625], iron ABC transporter substrate binding protein [accession no. WP\_005715971.1], iron binding protein [accession no. WP\_005726096.1], and iron permease [accession no. WP\_010906655.1]) or proteins of the oxidative stress response (catalase [accession no. WP\_010906440], superoxide dismutase [accession no. WP\_005750998], peroxiredoxin [accession no. WP\_005716614.1]). These analyses also revealed expression of genes encoding stress- and starvation-induced proteins (stringent starvation protein A homologue [accession no. WP\_005726291.1]) and the virulence factor SrfB (accession no. WP\_005755436.1).

### 16S Metagenomic Sequencing Protocol

We applied a metagenomics workflow for classifying organisms from the V3 and V4 regions of the 16S rRNA gene by using a Greengenes database (<http://greengenes.lbl.gov>) to test tissue taken from 2 animals (lung tissue from animal X and kidney tissue from animal Y) (Table 6). Among the variable regions of 16S gene, V3 is a highly variable region that can distinguish bacteria to the genus level. V4 is also efficient but less so than V3 (8). The output of the workflow classified the reads at the primary taxonomic levels (kingdom, phylum, class, order, family, genus, and species).

Sequencing statistics revealed the number of total reads to be 63,508 for lung tissue and 15,422 for kidney

tissue. The number of reads passing quality filtering was 58,161 for lung tissue and 14,291 for kidney tissue. The percentage of reads passing quality filtering was 91.6% for lung tissue and 92.7% for kidney tissue.

Of all reads generated, 86.80%–89.05% of all short reads were from bacteria of the genus *Pasteurella*, of which 44.06%–48.32% were identified as *P. multocida*. Other species were *Pasteurellaceae* (10.75%–12.21%), *P. pneumotropica* (4.06%–5.67%), and those unclassified at species level (34.91%–36.53%) (Table 5). More than 80% of unclassified reads at the species level belonged to the *Pasteurella* genus.

### Discussion

Saiga antelopes are a critically endangered species (1), and the population is increasingly fragmented and vulnerable to stochastic events such as disease epidemics. The mass die-off in Kazakhstan and the small population of  $\approx 10,000$  in Mongolia recently devastated by peste des petits ruminants (PPR) virus in 2017 illustrates this point (3). The saiga antelopes undertake large-scale seasonal migrations between their summer and winter ranges because of the extreme variation in climate conditions and the need for pastures offering sufficient forage. The calving sites are highly variable from year to year and depend on plant phenology, environmental factors, and anthropogenic effects (9). The analysis of available data showed that the number of saiga antelopes in Kazakhstan over the past 60 years has fluctuated widely, from  $\approx 2$  million in the 1970s to  $\approx 50,000$  animals in the early 21st century because of poaching and other factors, including a series of mass die-offs (10,11). A few incidences of infectious disease, including foot-and-mouth disease, have been confirmed (12), but most events were attributable to pasteurellosis; *M. haemolytica* and *P. multocida* were isolated on occasion (13). However, diagnoses are lacking comprehensive clinical, pathological, epidemiologic, and environmental investigation and remain tentative in all cases outside the 2015 event. Diagnosis of wildlife deaths is constrained by the fact these populations are not managed nor always monitored regularly, meaning die-offs occur frequently and investigators often do not have access to fresh carcasses. In the 2015

**Table 6.** Top 8 of 94 species classification results after 16S bacterial metagenome sequencing in an investigation of a mass die-off of saiga antelopes, Kazakhstan, 2015\*

Classification	No. reads		% Total reads	
	Lung (animal X)	Kidney (animal Y)	Lung (animal X)	Kidney (animal Y)
<i>Pasteurella multocida</i>	25,625	6,907	44.06	48.32
Unclassified at species level	21,246	4,990	36.53	34.91
<i>Pasteurellaceae</i>	7,101	1,536	12.21	10.75
<i>Pasteurella pneumotropica</i>	3,298	580	5.67	4.06
<i>Mannheimia caviae</i>	462	78	0.79	0.55
<i>Serratia entomophila</i>	50	17	0.09	0.12
<i>Bacillus horneckiae</i>	49	16	0.08	0.11
<i>Vagococcus teuberi</i>	39	13	0.07	0.09
<i>Sporolactobacillus putidus</i>	50	0	0.09	0
<i>Acinetobacter gerneri</i>	49	0	0.08	0
<i>Gallibacterium melopsittaci</i>	39	0	0.07	0

\*Total species-level taxonomic categories identified: 94 for lung sample (animal X) and 68 for kidney sample (animal Y).

saiga antelope event, a monitoring team was in place in 2 of the 15 die-off locations and were equipped for general diagnostic work. This situation was unusual and provided a unique opportunity, but the unpredictability of such an event happening limited the extent of the outbreak investigation. Sampling was necessarily strategic, and because all of the animals in the population were affected by the same syndrome and died, the sample size did not need to be large or statistically representative. Each case would have an equal chance of providing the result, and failure to diagnose would be more likely a product of insufficient material per case or loss of viability of organisms because of cold chain and storage issues.

Nevertheless, the findings obtained from this work are representative of the population for a few reasons. First, the clinical syndrome was uniform in both the adult and calf populations. We observed no statistically significant variation in the temporal progression once symptoms were noted, and clinical signs and gross pathology were highly consistent. Second, the 100% mortality rate among the herds indicates a universal effect, and no samples taken were likely to be nonrepresentative or attributable to an alternate etiology. In addition, the rapidity of the syndrome precluded large numbers of cases being investigated by our relatively small team because necropsy and sampling for each case took several hours to complete.

Microbiologic and virologic diagnostic methods showed that in samples from >90% of saiga antelopes that died in 2015, the cause was pasteurellosis (3,6). Previous studies had demonstrated the absence of other potential causative agents by diagnostic PCR, including bacteria (e.g., anthrax bacillus, *Coxiella burnetti*, *Erysipelothrix rhusiopathiae*, and *Listeria* spp.), mycoplasma (e.g., *Mycoplasma ovipneumoniae*), and virus diseases (e.g., foot-and-mouth disease, bluetongue, PPR, epizootic hemorrhagic disease, sheep pox, Akabane, Aujeszky's disease, bovine viral diarrhea, visna-maedi, and malignant catarrhal fever) (3,6). Furthermore, these studies used capsular typing with specific primers to show that strains of *P. multocida* from saiga antelopes belonged to serogroup B (3,6). Our 16S

analysis also showed that the *P. multocida* isolated from saiga antelopes in 2015 in Akmolra and Kostanay oblasts were genetically identical to the bacteria isolated from saiga antelopes in 1988, 2010, 2011, and 2012, as well as *P. multocida* subsp. *multocida* PM30 strain (GenBank accession no. AY299312) isolated from ill cattle with hemorrhagic septicemia in 2004 (6).

Three different HTS protocols were used in parallel to identify unknown microbial pathogens that played a potential role in disease pathogenesis: a protocol based on random amplification using Illumina sequencing; an RNA-based analysis without amplification combined with sequencing on an Ion Torrent (ThermoFisher, <https://www.thermofisher.com>); and a bacterial 16S sequencing pipeline using Illumina technology. Each of these workflows demonstrated the potential for different experimental challenges in obtaining metagenomic datasets (e.g., biases in amplification-based protocols and the use of low-input starting material in no-amplification protocols) (14). The high sensitivity of such methods to detect small amounts of nucleic acids also poses challenges in terms of prevention of contamination and false-positive results. Caution should be exercised in drawing conclusions from such datasets without appropriate validation. In addition to blood spots, other tests, including bacteriologic and virologic tests on various tissues and samples taken, were conducted locally in Kazakhstan at government laboratories and reported elsewhere (3).

Despite the high sensitivity of the methodologies we used, our study is somewhat limited by the sample type (FTA cards), which precludes the detection of pathogens in lymphoid tissues and other organs. The use of FTA cards might also introduce biases in the testing protocols, which can favor or hinder the detection of certain types of viruses (15).

Both metagenomic protocols conclusively identified *Pasteurella* spp. in large numbers of reads compared with other pathogens; these findings were then confirmed in a third pipeline using 16S bacterial ribosomal RNA sequencing. Further analysis of *P. multocida* bacterial sequences

suggested that the expression of metabolic- and stress-related proteins might suggest that the bacteria were actively growing and in active competition with the host organism for essential nutrients, especially iron, as shown by the expression of the genes coding for iron uptake systems.

Our de novo assembly approach also identified 195 short contigs that could not be attributed to any sequence present in several BLAST databases; of those, only 47 were identified using tblastx. The subsequent analysis based on SUPERFAMILY was only able to find protein homologies for 6 contigs, with most of them having homologies to bacterial genes. Whether this result is important is unclear; our unknown contigs might belong to  $\geq 1$  uncultured bacteria that have not been sequenced before. Previously published metagenomic studies have resulted in as many as 50% unidentifiable reads (16,17); the figures for our work are reduced in comparison, in particular when considering the results of our de novo approach ( $\approx 72\%$  of our reads map to our assembly and  $< 1\%$  of the reads map to the contigs that we are unable to identify) (Appendix Table 2). Overall, the amount of unexplained sequence seems relatively small, in particular when considering the substantial number of species of bacterial, viral, and eukaryotic genome that remain either to be discovered or characterized. The simple fact that not all organisms have been sequenced or are available on central sequence repositories will always contribute to a percentage of unidentifiable reads.

*P. multocida* is a ubiquitous organism, most probably widely present in the saiga antelope population in its latent form. The potential pathogenicity is inherent in the organism and can be triggered opportunistically at any time in response to environmental triggers. The epidemiology of and observations on the spatiotemporal distribution of ill animals and carcasses in this study suggests that transmission of bacteria from animal to animal did not occur in most cases (except from mothers to calves, which occurred through infected milk). The near synchronous events in discrete subpopulations, with large distances between aggregations of many hundreds of kilometers, further precluded an infective process spreading across the population.

Research to date suggests that environmental conditions in the 10 days leading up to a die-off are critical and significantly associated with increased heat and humidity (3). The trends in climate in the region are for warmer and wetter conditions, which might have been an important factor in these recent events that have occurred irregularly over the last few decades. Immunocompetence was not thought to be a factor in the pathogenesis because the population was behaving normally, was unstressed, and was in apparent good health and body condition with large fat reserves observed postmortem. In addition, genetic analysis of the saiga antelope population shows them to be the most heterogeneous of any mammal species on record (S.

Zuther, unpub. data), thus excluding inbreeding as a factor, despite the potential bottlenecking of the population in recent times.

The mechanism behind the mass die-off might be an environmentally triggered bacterial proliferation that overwhelmed the mucosal immunity of the upper respiratory and gastrointestinal tracts. This hypothesis is further supported by the observation that calves, which are unlikely to be infected with the commensal bacteria in the first couple days of life, died some hours or longer after their mothers, most likely from suckling infected milk from ill or dead mothers, activity that was observed by investigators (3).

In this study, HTS was used to identify pathogens that might have predisposed or contributed to the severity of the saiga antelope die-off in 2015. In previous studies, *P. multocida* type B was identified by culture, and viruses of veterinary importance (foot-and-mouth, PPR, and bluetongue viruses) were ruled out by using pathogen-specific diagnostic tests. In our study, 3 laboratories using 3 distinct HTS analytic approaches failed to identify additional pathogens. These findings, combined with clinical, necroscopic, microbiologic, and histopathologic investigations, indicate hemorrhagic septicemia caused by *P. multocida* serotype B is the proximate cause, and possibly the only cause, of this die-off. Environmental factors might have triggered nearly simultaneous bacterial proliferation and subsequent virulence in affected aggregations.

Comprehensive field monitoring and additional experimental studies of *P. multocida* infection in saiga antelopes are necessary to evaluate the potential co-factors triggering the virulence of bacteria. These recurrent mass die-offs could cause extinction of saiga antelope populations in just 1 event, especially if, in future outbreaks, additional pathogens in combination with *P. multocida* affect the population.

This study was undertaken with the approval of the Royal Veterinary College Ethical Review Board (unique reference no. 2015 1435), the Biosafety Institute of the Ministry of Science and Education, the Committee for Forestry and Wildlife, and the Veterinary Reference Laboratory Astana, Ministry of Agriculture, Republic of Kazakhstan.

This study was funded by the National Environment Research Council of the United Kingdom urgency research grant (no. NE/N007646/1), the Association for the Conservation of Biodiversity of Kazakhstan, the Flora and Fauna International, the People's Trust for Endangered Species, the Saiga Conservation Alliance, the Royal Society for the Protection of Birds, and the Frankfurt Zoological Society. G.L.F. was funded through a research project (EpiSeq) by the European and Major Infectious Disease of Livestock–European Research Area. S.F., D.H., and M.B. received funding from the European Union Horizon 2020 Research and Innovation Program COMPARE

(grant no. 643476). J.F. and D.P.K. were funded through grants from the Department of the Environment, Fisheries, and Rural Agriculture. P.R. received support from the Biotechnology and Biological Sciences Research Council of the United Kingdom (project nos. BB/E/1/00007035, BB/E/1/00007036, and BBS/E/1/00007039).

## About the Author

Dr. Fereidouni is a senior scientist at the Research Institute of Wildlife Ecology at the University of Veterinary Medicine of Vienna. His primary research interests include wildlife conservation and health, wildlife diseases and emerging pathogens, One Health, and eco-immunology.

## References

- Mallon DP. *Saiga tatarica* ssp. In: IUCN red list of threatened species. Version 2011.2. Cambridge: International Union for Conservation of Nature Global Species Programme Red List Unit; 2011.
- Kock R, Grachev Y, Zhakypbayev A, Usenbayev A, Zuther S, Klimanova O, et al. G. A retrospective assessment of saiga antelope *Saiga tatarica* die-off in Western Kazakhstan 2010–2011. *Saiga News*. 2011;14:1–4.
- Kock R, Orynbayev M, Robinson S, Zuther S, Singh N, Beauvais W, et al. Saigas on the brink: multi-disciplinary analysis of the factors influencing mass mortality events. *Sci Adv*. 2018;4:eaa02314.
- Zuther S. The saiga antelope mass die-off in the Betpak-Dala population in May 2015. *Saiga News*. 2016;20:6–8.
- Abdrakhmanov A. Overview of the 2015 saiga mass mortality event in Kazakhstan. Presented at: Technical Workshop of the Third Meeting of Signatories of the Memorandum of Understanding Concerning Conservation, Restoration, and Sustainable Use of the Saiga Antelope (*Saiga* spp.), October 2015, Tashkent, Uzbekistan [in Russian] [cited 2018 Jan 7]. <https://www.cms.int/en/document/technical-workshop-2610-overview-2015-saiga-mass-mortality-event-kazakhstan>
- Orynbayev M, Sultankulova K, Sansyrbay A, Rystayeva R, Shorayeva K, Namet A, et al. Biological characterization of *Pasteurella multocida* present in the Saiga population. *BMC Microbiol*. 2019;19:37. <http://dx.doi.org/10.1186/s12866-019-1407-9>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402. <http://dx.doi.org/10.1093/nar/25.17.3389>
- Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods*. 2007;69:330–9. <http://dx.doi.org/10.1016/j.mimet.2007.02.005>
- Singh NJ, Grachev IA, Bekenov AB, Milner-Gulland EJ. Saiga antelope calving site selection is increasingly driven by human disturbance. *Biol Conserv*. 2010;143:1770–9. <http://dx.doi.org/10.1016/j.biocon.2010.04.026>
- Milner-Gulland E, Kholodova M, Bekenov A, Bukreeva O, Grachev I, Amgalan L, et al. Dramatic declines in saiga antelope populations. *Oryx*. 2001;35:340–5. <http://dx.doi.org/10.1017/S0030605300032105>
- Grachev Y. Results of the 2013 aerial survey in Kazakhstan. *Saiga News*. 2013;17:5.
- Orynbayev MB, Beauvais W, Sansyrbay AR, Rystayeva RA, Sultankulova KT, Kerimbaev AA, et al. Seroprevalence of infectious diseases in saiga antelope (*Saiga tatarica tatarica*) in Kazakhstan 2012–2014. *Prev Vet Med*. 2016;127:100–4. <http://dx.doi.org/10.1016/j.prevetmed.2016.03.016>
- Bekenov AB, Pole SB, Khakhin GV, Grachev YA. Mortality from diseases and parasitic invasions. In: Sokolov VE and Zhirnov LV, editors. *The saiga antelope*. Moscow: Russian Academy of Sciences; 1998. p 247–52.
- Höper D, Mettenleiter TC, Beer M. Metagenomic approaches to identifying infectious agents. *Rev Sci Tech*. 2016;35:83–93. <http://dx.doi.org/10.20506/rst.35.1.2419>
- Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci Rep*. 2015;5:16532. <http://dx.doi.org/10.1038/srep16532>
- Yoosuf S, Andrews-Pfannkoch C, Tenney A, McQuaid J, Williamson S, Thiagarajan M, et al. A metagenomic framework for the study of airborne microbial communities. *PLoS One*. 2013;8:e81862. <http://dx.doi.org/10.1371/journal.pone.0081862>
- Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, et al. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst*. 2015;1:72–87. <http://dx.doi.org/10.1016/j.cels.2015.01.001>

---

Address for correspondence: Richard Kock, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire AL9 7TA, UK; email: rkock@rvc.ac.uk

# Mass Die-Off of Saiga Antelopes, Kazakhstan, 2015

## Appendix

### Random Amplification–Based Sequencing Protocol

A random amplification-based sequencing protocol designed to detect both RNA and DNA based pathogens was employed at Pirbright to further investigate the blood spots. For random amplification-based sequencing, total nucleic acids were extracted by excising 1cm diameter discs of the blood spot and incubating these in 200µl of RLT buffer (with 1% β-mercaptoethanol) (Qiagen) at 56°C for 15 minutes, before supernatant was decanted off and nucleic acids extracted using the RNeasy mini kit (Qiagen). All samples were processed individually to minimise cross-contamination with periods of at least 30 minutes separating the handling of subsequent samples. Extracted nucleic acid was validated for further investigations for the presence of the highly-conserved housekeeping gene (GAPDH) using an established qualitative real-time RT-PCR assay (1). Total extracted nucleic acid was also tested for the presence of notifiable transboundary ruminant viral pathogens using previously established diagnostic real-time RT-PCR assays for foot-and-mouth disease virus (2), blue-tongue virus (3), peste des petits ruminants virus (4), capripox viruses (5) and epizootic haemorrhagic disease virus (6).

All 12 samples that were positive for the presence of GAPDH, suggestive of intact nucleic acid, were subjected to further investigations. Random amplification of nucleic acid of 6 GAPDH positive samples (samples 1–6) was performed using the SeqPlex kit (Sigma). Briefly, 3.3µL of extracted nucleic acid was subjected to random amplification using manufacturer's protocols. Amplification products were run in a 1% agarose gel at 60V for 1 hour and a 300bp band was excised and purified using a Qiagen MinElute kit (Qiagen) and eluted in 10uL. The elution was quantified using a nanodrop, qubit dsDNA BR kit (Life technologies) and Bioanalyser dsDNA kit (Agilent). Five hundred nanograms of amplicon were used to prepare

sequencing libraries using a Kapa Hyper prep kit (KapaBiosciences). Constructed libraries were purified and quantified using a Kapa Illumina library quantification kit (Kapa Biosciences). Libraries were diluted to 15pM and loaded onto a MiSeq 2x300 version 3 reagent cartridge and run on an Illumina MiSeq with a 5% PhiX spike-in control. Files were run as fastq only and transferred onto a high-performance computing cluster for further analysis.

Two separate analysis protocols were employed to interrogate the data; a k-mer based approach and a de-novo approach (7,8). Before both protocols adaptors were removed, and reads quality trimmed (using a q-score threshold of 30), with trim-galore (version 0.4.3) (9). For the k-mer based protocol, any reads less than 100 bases in length were removed. All remaining reads were then analyzed and classified taxonomically using the Kraken mapping software (version 0–15-β) using a local database for identification of all viruses, bacteria and archaeal genomes extracted from the NCBI refseq database (downloaded 15/04/2017). Results were visualized using Kronatools (10). In the case of the *de-novo* protocol, the sequencing reads from all samples were initially pooled together and then assembled using SPAdes (11). Only contigs with length 200 nt or more were kept, to filter out possible false positives due to short sequences. The resulting contigs were scanned against several BLAST (12) pre-generated databases downloaded from NCBI (*nt*, *tsa\_nt*, *ref\_prok\_rep\_genomes*, *ref\_viruses\_rep\_genomes*, and *vector* – downloaded 15/04/2017). The results were filtered keeping only the best hits and discarding hits having <80% similarity with the query, and length <50% than the length of the query. Hits were subsequently accumulated by species. As most BLAST databases containing genomes of organisms assembled with Illumina technologies also contain the phiX genome due to it being present as positive sequencing control, we manually fished out the phiX contigs before running BLAST. The contigs for which no hit from any of the 5 BLAST databases was recorded were classified as unknown. The unknown contigs were subsequently scanned against the same BLAST databases using TBLASTX (13), to find similarities at protein level. Only hits with e-value <10<sup>-5</sup>, and only the best hit per contig, were kept, to reduce the number of spurious matches. Separately, the unknown contigs were also translated using all possible open reading frames and subsequently processed with SUPERFAMILY version 1.75 (13) to identify potential homologies with known proteins. The raw paired fastq reads generated by this protocol and subsequent assemblies (identifiable and unknown) were submitted to the European nucleotide archive (ENA) archive under accession no. (to be confirmed).

A local contamination by an unrelated laboratory adapted avian coronavirus (Infectious Bronchitis Virus IBV M41-CK) was detected by both methods. To exclude the possibility that IBV was present in the original RNA samples, we tested remaining stocks of original RNA using an IBV-specific 5'UTR RT-qPCR (14), which was negative in 6/6 samples tested. IBV reads/contigs were then excluded from subsequent analysis and conclusions, as they were not considered to have any impact on the investigation.

## **RNA Sequencing Protocol**

This protocol was applied by FLI, and RNA was extracted from blood spots on FTA cards. To this end, before extraction, from each FTA card representing one individual, five 5mm punches were ground in a 2ml tube with a 5mm stainless steel bead in 1ml Trizol (Invitrogen) using a TissueLyser (Qiagen) set at 20 Hz for 3 min. Thereafter, the tubes were spun in a standard table-top centrifuge at 13,000 rpm and the supernatant transferred to a fresh tube. Subsequently, the published protocol for the extraction of RNA was applied (15). In brief, the aqueous phase was mixed with ethanol (40% v/v) and this mixture transferred to a Qiagen RNeasy spin column and all further steps, including the optional on-column DNase treatment, carried out according to the manufacturer's instructions. The extracted RNA was quantified using a Nanodrop ND1000 instrument (Peqlab, Erlangen, Germany), and 500 ng were used for cDNA synthesis and library preparation as described (16). Briefly, after the addition of random hexamer primers, the RNA was denatured at 95°C for 2 min, immediately followed by snap-freezing. This RNA-primer mix was used as input for reverse transcription and second strand synthesis with the cDNA synthesis system kit (Roche, Mannheim, Germany). The obtained double-stranded cDNA was fragmented to a peak size of approx. 500 bp using the M220 Focused-ultrasonicator (Covaris, Brighton, United Kingdom) and used as input for library preparation with a GeneRead DNA Library L Core Kit (Qiagen) and Ion Xpress Barcode Adapters (Life Technologies, Darmstadt, Germany). After quality control with an Agilent Bioanalyzer 2100 DNA HS kit (Agilent, Waldbronn, Germany) and quantification with the KAPA Library Quantification Kit - Ion Torrent Universal (Roche), the resulting libraries (libraries lib01416, lib01417, lib01418, lib01419; corresponding to samples 2, 5, 8, 11) were sequenced using the Ion Torrent PGM (Life Technologies) with 400 bp HiQ reagents following the manufacturer's instructions. The obtained datasets were analyzed using the software pipeline

RIEMS (17). In addition, the datasets were mapped along the available *P. multocida* genome sequence (NC\_002663.1) using the Roche/454 software suite (v3.0; Roche) and the generated contigs analyzed using BLASTX (v2.2.26+) (12). The raw reads generated by this protocol were submitted to the ENA archive under accession no. PRJEB28164.

## **16S Metagenomic Sequencing Protocol**

This protocol was applied by IMV, and microbial DNA was extracted from lung and kidney tissues separately using Trizol Reagent (Thermo Fisher Scientific, USA) according to manufacturer's recommendations.

Library preparation was conducted according to Illumina 16S Metagenomic Sequencing Library Preparation Workflow (Illumina, USA). This protocol combined with a benchtop sequencing system, on-board primary analysis, and secondary analysis using MiSeq Reporter or BaseSpace, provides a comprehensive workflow for 16S rRNA amplicon sequencing. Briefly, the 16S Amplicon PCR forward and reverse primers (recommended by Illumina) were used to amplify the bacterial V3 and V4 regions (with  $\approx$ 460 bp length). PCR products were purified using magnetic AMPure XP beads (Beckman Coulter, USA) to remove free primers and primer dimers to avoid interference with the sequencing process. Then amplification of the V3 and V4 region using a limited cycle PCR with simultaneous addition of Nextera XT (Illumina, USA) sequencing adapters and dual indexed barcodes to the amplicon target was conducted. AMPure XP beads were used to clean up the final library before quantification on Qubit 2.0 spectrophotometer (Thermo Fisher Scientific, USA). Fragments were visualized on an agarose gel to check quality and average nucleotide length.

Sequencing was performed on an Illumina MiSeq using Illumina v.3 reagent kit with a 7.5% PhiX (Illumina, USA) spike-in control. Data were analyzed locally by on-board MiSeq Reporter software (Illumina, USA). Taxonomic classification was performed using the Greengenes database showing genus or species level classification in a graphical and table format. The 16S sequencing metagenome dataset file was submitted to Genbank under accession no. PRJNA486600.

## References

1. King DP, Burman A, Gold S, Shaw AE, Jackson T, Ferris NP. Integrin sub-unit expression in cell cultures used for the diagnosis of foot-and-mouth disease. *Vet Immunol Immunopathol.* 2011;140:259–65. [PubMed http://dx.doi.org/10.1016/j.vetimm.2011.01.008](http://dx.doi.org/10.1016/j.vetimm.2011.01.008)
2. Callahan JD, Brown F, Osorio FA, Sur JH, Kramer E, Long GW, et al. Use of a portable real-time reverse transcriptase-polymerase chain reaction assay for rapid detection of foot-and-mouth disease virus. *J Am Vet Med Assoc.* 2002;220:1636–42. [PubMed http://dx.doi.org/10.2460/javma.2002.220.1636](http://dx.doi.org/10.2460/javma.2002.220.1636)
3. Hofmann M, Griot C, Chaignat V, Perler L, Thür B. Bluetongue disease reaches Switzerland [in German]. *Schweiz Arch Tierheilkd.* 2008;150:49–56. [PubMed http://dx.doi.org/10.1024/0036-7281.150.2.49](http://dx.doi.org/10.1024/0036-7281.150.2.49)
4. Batten CA, Banyard AC, King DP, Henstock MR, Edwards L, Sanders A, et al. A real time RT-PCR assay for the specific detection of Peste des petits ruminants virus. *J Virol Methods.* 2011;171:401–4. [10.1016/j.jviromet.2010.11.022 PubMed http://dx.doi.org/10.1016/j.jviromet.2010.11.022](http://dx.doi.org/10.1016/j.jviromet.2010.11.022)
5. Bowden TR, Babiuk SL, Parkyn GR, Copps JS, Boyle DB. Capripoxvirus tissue tropism and shedding: A quantitative study in experimentally infected sheep and goats. *Virology.* 2008;371:380–93. [PubMed http://dx.doi.org/10.1016/j.virol.2007.10.002](http://dx.doi.org/10.1016/j.virol.2007.10.002)
6. Maan NS, Maan S, Potgieter AC, Wright IM, Belaganahalli M, Mertens PPC. Development of real-time RT-PCR assays for detection and typing of epizootic haemorrhagic disease virus. *Transbound Emerg Dis.* 2017;64:1120–32. [10.1111/tbed.12477 PubMed http://dx.doi.org/10.1111/tbed.12477](http://dx.doi.org/10.1111/tbed.12477)
7. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46. [10.1186/gb-2014-15-3-r46 PubMed http://dx.doi.org/10.1186/gb-2014-15-3-r46](http://dx.doi.org/10.1186/gb-2014-15-3-r46)
8. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods.* 2012;9:1185–8. [PubMed http://dx.doi.org/10.1038/nmeth.2221](http://dx.doi.org/10.1038/nmeth.2221)
9. Babraham Bioinformatics. Trim Galore [cited 15 April 2017]. [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore)
10. Ondov B. KronaTools 2.7 [cited 2017 Apr 13]. <https://github.com/marbl/Krona/wiki/KronaTools>

11. Bankevich A, Nurk S, Antipov D, Guevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77. [PubMed http://dx.doi.org/10.1089/cmb.2012.0021](http://dx.doi.org/10.1089/cmb.2012.0021)
12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402. [PubMed http://dx.doi.org/10.1093/nar/25.17.3389](http://dx.doi.org/10.1093/nar/25.17.3389)
13. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY—comparative genomics, datamining and sophisticated visualisation. *Nucleic Acids Res.* 2009;37:D380–6. [PubMed http://dx.doi.org/10.1093/nar/gkn762](http://dx.doi.org/10.1093/nar/gkn762)
14. Callison SA, Hilt DA, Boynton TO, Sample BF, Robison R, Swayne DE, et al. Development and evaluation of a real-time Taqman RT-PCR assay for the detection of infectious bronchitis virus from infected chickens. *J Virol Methods.* 2006;138:60–5. [PubMed http://dx.doi.org/10.1016/j.jviromet.2006.07.018](http://dx.doi.org/10.1016/j.jviromet.2006.07.018)
15. Höper D, Hoffmann B, Beer M. A comprehensive deep sequencing strategy for full-length genomes of influenza A. *PLoS One.* 2011;6:e19075. [PubMed http://dx.doi.org/10.1371/journal.pone.0019075](http://dx.doi.org/10.1371/journal.pone.0019075)
16. Juozapaitis M, Aguiar Moreira É, Mena I, Giese S, Riegger D, Pohlmann A, et al. An infectious bat-derived chimeric influenza virus harbouring the entry machinery of an influenza A virus. *Nat Commun.* 2014;5:4448. [PubMed http://dx.doi.org/10.1038/ncomms5448](http://dx.doi.org/10.1038/ncomms5448)
17. Scheuch M, Höper D, Beer M. RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. *BMC Bioinformatics.* 2015;16:69. [PubMed http://dx.doi.org/10.1186/s12859-015-0503-6](http://dx.doi.org/10.1186/s12859-015-0503-6)

**Appendix Table 1.** Viruses and bacteria identified with the Kraken analysis protocol from the data produced using the random amplification meta-transcriptomic protocol. Numbers in parentheses represent % of total reads

Organism	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
<b>Viruses</b>						
Enterobacteria phage phiX174 sensu lato *	18742 (7.79)	46303 (4.9)	16755 (4.8)	22694 (4.85)	19518 (5.69)	13204 (4.77)
Infectious Bronchitis virus (M41-CK) *	18 (0.01)	8394 (0.55)	6 (0)	52 (0.01)	19 (0.01)	36 (0.01)
Enterobacteria phage M13	2 (0)	–	–	–	–	–
Enterobacteria phage ID2 Moscow/ID/2001	–	3 (0)	–	–	–	–
Haemophilus phage SuMu	41 (0.02)	48 (0.01)	22 (0.01)	40 (0.01)	45 (0.01)	32 (0.01)
Primula malacoides virus China/Mar2007	1 (0)	34 (0)	1 (0)	104 (0.02)	11 (0)	9 (0.01)
Hyposoter fugitivus ichnovirus	1 (0)	–	–	–	–	–
Torque teno midi virus 2	–	3 (0)	1 (0)	2 (0)	–	–
Elephantid herpesvirus 1	1 (0)	2 (0)	–	6 (0)	1 (0)	–
Parvovirus NIH-CQV	–	2 (0)	–	–	2 (0)	–
Jingmen tick virus	–	10 (0)	–	4 (0)	2 (0)	2 (0)
Solenopsis invicta virus 3	–	3 (0)	–	–	–	–
Carp picornavirus 1	–	6 (0)	–	–	3 (0)	3 (0)
Eel picornavirus 1	–	1 (0)	–	–	–	–
Mouse astrovirus M-52/USA/2008	1 (0)	4 (0)	–	–	–	1 (0)
Spring beauty latent virus	–	2 (0)	–	–	1 (0)	–
Hepatitis C virus	–	1 (0)	–	2 (0)	1 (0)	–
Dolphin morbillivirus	–	4 (0)	–	–	–	–
Dickeya phage RC-2014	–	3 (0)	2 (0)	6 (0)	1 (0)	–
Cynomolgus macaque cytomegalovirus (Ottawa)	–	1 (0)	–	–	–	–
Grapevine Syrah virus 1	–	–	–	–	1 (0)	–
Bacillus phage SPO1	–	–	2 (0)	–	1 (0)	2 (0)
Cercopithecine herpesvirus 2	–	2 (0)	–	–	–	1 (0)
Ictalurid herpesvirus 1	–	2 (0)	–	–	–	–
Cyprinid herpesvirus 1	–	2 (0)	–	–	–	–
Pandoravirus salinus	–	1 (0)	2 (0)	4 (0)	–	–
Hyposoter fugitivus ichnovirus	–	2 (0)	–	–	1 (0)	–
Glypta fumiferanae ichnovirus	–	–	1 (0)	–	–	–
Cotesia congregata bracovirus	–	1 (0)	–	–	–	–
Maruca vitrata nucleopolyhedrovirus	–	1 (0)	–	–	–	–
Orgyia leucostigma NPV	–	–	–	2 (0)	–	–
Phaeocystis globosa virus	–	1 (0)	–	4 (0)	2 (0)	–
Hop trefoil cryptic virus 2	–	1 (0)	–	–	1 (0)	–
Y73 sarcoma virus	–	1 (0)	–	–	–	–
Solenopsis invicta virus 3	–	–	1 (0)	–	–	–
Red clover cryptic virus 2	–	–	1 (0)	–	–	–
Dulcamara mottle virus	–	–	–	–	–	1 (0)
<b>Bacteria</b>						
Unclassified	14019 (58.34)	810950 (73.3)	219326 (62.8)	266912 (57.06)	254164 (54.02)	205048 (54.81)
Pasteurella multocida	52771 (28.56)	121145 (14.4)	62719 (23.23)	88348 (24.27)	91558 (25.45)	71673 (24.99)
Alteromonas macleodii str. 'Ionian Sea U8'	1366 (0.57)	2724 (0.25)	46 (0.01)	980 (0.21)	4350 (0.92)	7737 (2.03)
Achromobacter xylosoxidans	2 (0)	13247 (1.2)	22 (0.01)	82 (0.02)	3423 (0.72)	6036 (1.61)
Dickeya dadantii Ech703	69 (0.03)	95 (0.01)	103 (0.03)	78 (0.02)	76 (0.01)	42 (0.02)
Haemophilus spp.	23 (0.01)	95 (0.01)	43 (0.01)	72 (0.02)	76 (0.01)	42 (0.02)
Klebsiella variicola At-22	11 (0)	58 (0.01)	16 (0)	54 (0.01)	16 (0)	14 (0)
Mannheimia haemolytica	10 (0)	13 (0)	12 (0)	18 (0)	22 (0)	7 (0)
Mycoplasma spp.	13 (0.01)	312 (0.02)	32 (0.01)	42 (0.01)	29 (0.01)	13 (0)
Rickettsia africae ESF-5	1 (0)	5 (0)	1 (0)	18 (0)	29 (0)	33 (0.01)
Campylobacter spp. (C.jejuni subsp.)	3 (0)	39 (0)	25 (0.01)	18 (0.01)	5 (0)	8 (0)
Aggregatibacter	8 (0)	6 (0)	7 (0)	18 (0)	3 (0)	4 (0)
Candidatus Riesia pediculicola USDA	2 (0)	13 (0)	9 (0)	8 (0)	3 (0)	4 (0)
Vibrio spp. (inc. V.Cholerae)	2 (0)	23 (0)	13 (0)	14 (0)	11 (0)	8 (0)
Dichelobacter nodosus VC51703A	80 (0.03)	6 (0)	3 (0)	2 (0)	1 (0)	1 (0)
Histophilus somni	1 (0)	14 (0)	5 (0)	6 (0)	3 (0)	1 (0)

\*Reads aligning to PhiX and avian Coronavirus Infectious Bronchitis Virus M41-CK (IBV) were present in the final datasets. These were attributed to the positive sequencing control (PhiX) and a local contamination by a lab adapted strain of IBV, respectively (the presence of IBV in the original samples was excluded by RT-qPCR, as explained in the methods).

**Appendix Table 2.** Results from the *de-novo* assembly protocol applied to samples sequenced using the random amplification meta-transcriptomic protocol. The attribution as determined by the best BLASTN hit; the number of contigs with the same attribution; and their total length, are listed in columns 4, 2, and 3, respectively. In column 1 the read area (equivalent to the contig length times the average read coverage) is listed.

Area (length*coverage)	Contigs	Total length	Attribution
21640705	6013	2025938	<i>Pasteurella multocida</i>
12470305	1	5421	Enterobacteria phage phiX174 sensu lato
462023	602	108406	UNKNOWN SEQUENCE
397727	47	26577	Infectious bronchitis virus
351922	235	43608	<i>Gallus gallus</i> (chicken)
103710	353	55287	<i>Ovis canadensis canadensis</i>
71394	5	590	<i>Escherichia coli</i>
53466	1	230	uncultured <i>Pasteurella</i> sp.
49238	3	397	<i>Lasius turcicus</i>
29395	2	293	Cloning vector lambda EMBL3 SP6/T7, left arm
29395	2	293	Enterobacteria phage HK630
21658	5	2901	<i>Antidorcas marsupialis</i> (springbok)
12173	9	2176	<i>Ovis aries musimon</i> (mouflon)
11830	29	5197	<i>Capra hircus</i> (goat)
5984	16	2081	<i>Ovis aries</i> (sheep)
5973	2	531	<i>Antilope cervicapra</i> (blackbuck)
4833	4	967	<i>Nanger dama</i> ( <i>Dama gazelle</i> )
4271	6	2113	<i>Eudorcas thomsonii</i> (Thomson's gazelle)
3527	14	2579	<i>Bubalus bubalis</i> (water buffalo)
2687	8	1544	<i>Saiga tatarica</i>
2331	8	1020	<i>Numida meleagris</i> (helmeted guineafowl)
2119	24	6912	<i>Bos taurus</i> (cattle)
1779	9	1684	<i>Bos indicus</i>
1246	1	600	<i>Madoqua kirkii</i> (Kirk's dik-dik)
1033	4	513	<i>Apteryx australis mantelli</i>
893	4	530	<i>Meleagris gallopavo</i> (turkey)
854	1	330	uncultured bacterium
729	3	361	<i>Sus scrofa</i> (pig)
484	1	249	<i>Gazella leptoceros</i> (Rhim gazelle)
446	1	207	<i>Phascolarctos cinereus</i> (koala)
446	3	303	<i>Odocoileus virginianus texanus</i>
391	2	317	<i>Pantholops hodgsonii</i> (chiru)
376	1	313	<i>Odocoileus hemionus</i> (mule deer)
372	2	255	<i>Chinchilla lanigera</i> (long-tailed chinchilla)
146	1	206	<i>Onchocerca flexuosa</i>
0	1	256	<i>Schmidtea mediterranea</i>

**Appendix Table 3.** Results from the *de-novo* assembly protocol applied to samples sequenced using the random amplification meta-transcriptomic protocol. TBLASTX results for the contigs unassigned by BLASTN (*Unknown sequence*).

Area (length*coverage)	Contigs	Total length	Attribution
121760	70	33252	<i>Pasteurella multocida</i>
17650	69	10509	<i>Ovis canadensis canadensis</i>
15395	10	2379	<i>Eudorcas thomsonii</i> (Thomson's gazelle)
7154	8	2158	<i>Antidorcas marsupialis</i> (springbok)
4491	1	566	<i>Gazella leptoceros</i> (Rhim gazelle)
3574	1	486	<i>Haemophilus influenzae</i>
3020	2	999	Infectious bronchitis virus
2425	2	311	<i>Nanger dama</i> (Dama gazelle)
2019	12	2547	<i>Gallus gallus</i> (chicken)
1494	2	869	<i>Avibacterium paragallinarum</i> JF4211
1302	5	651	<i>Apteryx australis mantelli</i>
1102	1	213	<i>Pasteurella bettyae</i> CCUG 2042
1087	3	562	<i>Meleagris gallopavo</i> (turkey)
1067	1	241	<i>Avibacterium paragallinarum</i>
1015	4	663	<i>Bubalus bubalis</i> (water buffalo)
966	6	1133	<i>Capra hircus</i> (goat)
942	1	270	<i>Raphicerus sharpei</i> (Sharpe's grysbok)
756	1	180	<i>Bos indicus</i>
726	1	253	<i>Coturnix japonica</i> (Japanese quail)
701	1	209	<i>Actinobacillus succinogenes</i> 130Z
552	3	512	<i>Bos taurus</i> (cattle)
543	1	173	<i>Nanger granti</i> (Grant's gazelle)
542	1	313	<i>Haemophilus parasuis</i> SH0165
540	2	347	<i>Pseudorca crassidens</i> (false killer whale)
519	1	593	<i>Haemophilus somnus</i> 129PT
518	1	295	<i>Ficedula albicollis</i> (collared flycatcher)
482	1	237	<i>Haemophilus influenzae</i> 2019
476	2	238	<i>Ovis aries</i> (sheep)
445	1	171	<i>Odocoileus virginianus texanus</i>
444	3	646	<i>Antelope cervicapra</i> (blackbuck)
312	1	133	<i>Numida meleagris</i> (helmeted guineafowl)
292	1	184	<i>Bibersteinia trehalosi</i> USDA-ARS-USMARC-189
248	1	124	groundwater metagenome
233	1	222	<i>Catelicoccus marimammalium</i> M35/04/3
89	1	218	<i>Lepidonotothen nudifrons</i> (yellowfin notie)ID

**Appendix Table 4.** Results from the *de-novo* assembly protocol applied to samples sequenced using the random amplification meta-transcriptomic protocol. SUPERFAMILY hits for the contigs unassigned by BLASTN (*Unknown sequence*). Columns list contig name with frame and peptide counter appended; peptide region matched; SUPERFAMILY score; and SUPERFAMILY attribution.

Contig name	Peptide region	Score	Attribution
NODE_1869_length_403_cov_0.699752_g1587_i0_6_1	23468	2.59E-12	Actin/HSP70
NODE_477_length_778_cov_3.95861_g348_i0_5_1	33512	5.50E-23	FKBP immunophilin/proline isomerase
NODE_477_length_778_cov_3.95861_g348_i0_5_1	91–127	5.49E-07	TF C-terminus
NODE_477_length_778_cov_3.95861_g348_i0_6_2	27–136	1.31E-25	TF C-terminus
NODE_570_length_733_cov_1.83356_g423_i0_5_1	13–101	1.70E-14	Band 7/SPFH domain
NODE_570_length_733_cov_1.83356_g423_i0_6_2	30–83	1.16E-08	Band 7/SPFH domain
NODE_579_length_724_cov_6.30041_g430_i0_1_1	6–103	3.17E-30	Pseudouridine synthase II TruB
NODE_579_length_724_cov_6.30041_g430_i0_2_3	34–97	2.88E-12	Pseudouridine synthase II TruB
NODE_579_length_724_cov_6.30041_g430_i0_2_3	101–139	4.37E-03	PUA domain
NODE_2811_length_299_cov_5.0903_g2487_i0_6_1	12510	5.10E-04	NadC C-terminal domain-like
NODE_2367_length_343_cov_2.73105_g2045_i0_5_1	15950	1.88E-03	Porin chaperone SurA
NODE_2430_length_336_cov_7.68899_g2106_i0_1_1	43160	1.32E-03	Rubredoxin
NODE_2431_length_336_cov_3.47619_g2107_i0_2_1	23346	3.79E-14	LemA-like
NODE_2461_length_334_cov_2.71307_g2137_i0_3_1	18872	5.56E-10	GlnE-like domain
NODE_284_length_936_cov_3.66346_g202_i0_5_1	80–142	6.10E-21	FtsK C-terminal domain-like
NODE_284_length_936_cov_3.66346_g202_i0_5_2	32–148	1.96E-37	Outer-membrane lipoproteins carrier protein LolA
NODE_5168_length_140_cov_2_g4844_i0_1_2	43922	7.78E-03	B-box zinc binding domain
NODE_1234_length_521_cov_1.29175_g1000_i0_2_1	28430	5.30E-19	Lambda integrase-like
NODE_1287_length_506_cov_4.32016_g1049_i0_4_2	36–63	1.05E-05	Fumarate reductase/Succinate dehydrogenase iron-sulfur protein
NODE_1287_length_506_cov_4.32016_g1049_i0_5_1	35247	4.71E-11	Fumarate reductase/Succinate dehydrogenase iron-sulfur protein
NODE_4140_length_209_cov_3.35407_g3816_i0_6_1	19603	9.84E-07	Phage repressors
NODE_1037_length_567_cov_3.37449_g824_i0_2_1	34912	2.09E-06	HlyD-like secretion proteins
NODE_1104_length_550_cov_1.47091_g885_i0_4_2	22–66	3.27E-04	Mitotic arrest deficient-like 1
NODE_1422_length_476_cov_5.28361_g1175_i0_6_1	40–70	8.76E-05	Multidrug efflux transporter AcrB
NODE_1431_length_474_cov_1.90443_g1183_i0_4_1	24351	1.37E-12	TolC docking domain
NODE_1431_length_474_cov_1.90443_g1183_i0_6_2	16–75	4.02E-13	GHMP Kinase
NODE_1457_length_470_cov_0.501064_g1109_i1_5_1	32448	6.97E-15	GHMP Kinase
NODE_1520_length_457_cov_2.34792_g1268_i0_1_3	45323	9.61E-03	glucose-1-phosphate thymidyltransferase
NODE_1520_length_457_cov_2.34792_g1268_i0_2_1	45323	9.61E-03	DinB-like
NODE_88_length_1421_cov_2.96129_g36_i0_2_7	11–102	8.50E-10	Kelch motif
NODE_217_length_1020_cov_3.88725_g142_i0_4_2	27150	8.89E-16	Trp repressor
NODE_5735_length_120_cov_2_g5411_i0_4_1	19725	1.50E-11	Nitrogenase iron protein-like
NODE_5526_length_128_cov_2_g5202_i0_1_1	12451	4.19E-05	FAD-dependent thiol oxidase
NODE_5657_length_123_cov_2_g5333_i0_2_1	46419	3.36E-03	B-box zinc binding domain
NODE_5665_length_123_cov_2_g5341_i0_1_1	43983	7.45E-03	TM1622-like
NODE_5682_length_122_cov_2_g5358_i0_1_1	12086	3.57E-03	Variant RING domain
NODE_5427_length_131_cov_1.67939_g5103_i0_2_1	13119	5.49E-03	DNA binding domain of intron-encoded endonucleases
NODE_6013_length_108_cov_2.32407_g5689_i0_2_1	45717	9.16E-03	Myotoxin
NODE_188_length_1067_cov_2.89972_g117_i0_1_2	43466	8.89E-03	HIT zinc finger
NODE_188_length_1067_cov_2.89972_g117_i0_2_1	46–215	2.88E-47	Phosphoribosylpyrophosphate synthetase-like
NODE_188_length_1067_cov_2.89972_g117_i0_2_1	27668	4.47E-22	Phosphoribosylpyrophosphate synthetase-like
NODE_188_length_1067_cov_2.89972_g117_i0_5_3	46023	1.57E-03	Regulatory protein AraC
NODE_2577_length_322_cov_1.49596_g2253_i0_1_1	24716	1.06E-10	GHMP Kinase
NODE_1552_length_451_cov_2.82927_g1295_i0_5_1	25538	1.47E-10	Decarboxylase
NODE_5795_length_117_cov_1.46154_g5471_i0_6_1	13820	8.26E-05	GpdQ-like
NODE_5279_length_136_cov_2_g4955_i0_6_1	44256	4.38E-03	Aspartate/glutamate racemase
NODE_1577_length_446_cov_0.946188_g1320_i0_5_1	30621	7.33E-05	Acetyl-CoA synthetase-like
NODE_1577_length_446_cov_0.946188_g1320_i0_6_3	19–70	2.75E-09	Acetyl-CoA synthetase-like
NODE_1670_length_432_cov_1.3287_g1408_i0_5_2	19–67	9.61E-10	RNase P protein
NODE_1706_length_426_cov_15.2986_g1442_i0_1_1	20486	3.14E-15	N-acetylmuramoyl-L-alanine amidase-like
NODE_1706_length_426_cov_15.2986_g1442_i0_6_2	45078	9.16E-03	PMP inhibitors
NODE_1706_length_426_cov_15.2986_g1442_i0_6_3	16–68	5.72E-07	Exostosin
NODE_1754_length_419_cov_0.699284_g1487_i0_2_1	28734	9.27E-18	Formate dehydrogenase/DMSO reductase
NODE_1754_length_419_cov_0.699284_g1487_i0_6_1	26634	7.84E-08	Cold shock DNA binding domain-like
NODE_6253_length_99_cov_2_g5929_i0_3_1	47150	7.54E-03	Alcohol dehydrogenase-like

NODE_6276_length_98_cov_0_g5952_i0_4_1	47119	4.64E-03	RING finger domain
NODE_4661_length_185_cov_1.33514_g4337_i0_3_1	20821	9.73E-05	Saposin B
NODE_4708_length_181_cov_10.3481_g4384_i0_6_1	19815	4.12E-09	Lambda integrase-like
NODE_2774_length_303_cov_1.51155_g2450_i0_4_1	24-63	7.55E-06	Extended AAA-ATPase domain
NODE_695_length_676_cov_0.761834_g529_i0_6_2	28-101	1.84E-10	Extended AAA-ATPase domain
NODE_797_length_639_cov_2.37715_g617_i0_1_1	4-103	5.90E-32	Phosphoribosyltransferases (PRTases)
NODE_797_length_639_cov_2.37715_g617_i0_2_2	66-131	4.85E-07	Phosphoribosyltransferases (PRTases)
NODE_812_length_633_cov_5.69217_g632_i0_4_3	6-102	1.23E-23	Arylsulfatase
NODE_861_length_616_cov_4.54708_g673_i0_2_4	25263	4.47E-16	Leukotriene A4 hydrolase catalytic domain
NODE_1422_length_476_cov_5.28361_g1175_i0_4_3	25-80	4.84E-10	Multidrug efflux transporter AcrB TolC docking domain
NODE_2887_length_293_cov_4.83276_g2563_i0_4_1	32690	3.40E-12	DAK1
NODE_3070_length_279_cov_2.09498_g2746_i0_5_1	12236	2.09E-06	Inositol monophosphatase/fructose-1
NODE_3070_length_279_cov_2.09498_g2746_i0_5_2	19511	2.96E-06	Inositol monophosphatase/fructose-1
NODE_3257_length_265_cov_1.22642_g2933_i0_5_1	46539	3.01E-03	Chorismate synthase
NODE_592_length_718_cov_7.62535_g440_i0_5_2	39-138	1.37E-32	Phosphate binding protein-like
NODE_596_length_717_cov_1.13529_g444_i0_4_1	8-167	6.00E-24	Extended AAA-ATPase domain
NODE_217_length_1020_cov_3.88725_g142_i0_5_3	97-265	1.09E-32	GABA-aminotransferase-like
NODE_217_length_1020_cov_3.88725_g142_i0_5_3	27-104	3.32E-19	RecA protein-like (ATPase-domain)
NODE_222_length_1013_cov_6.8924_g147_i0_4_1	5-282	7.06E-88	Aconitase iron-sulfur domain
NODE_255_length_976_cov_3.78586_g179_i0_5_2	69-202	2.05E-27	Glycerol kinase
NODE_255_length_976_cov_3.78586_g179_i0_6_1	43-139	3.04E-15	Glycerol kinase
NODE_284_length_936_cov_3.66346_g202_i0_1_3	12206	3.79E-03	Interleukin 8-like chemokines
NODE_3601_length_241_cov_4.42739_g3277_i0_3_1	25812	1.29E-04	Lambda integrase-like
NODE_877_length_614_cov_0.467427_g684_i0_5_3	15827	1.96E-06	ABC transporter transmembrane region
NODE_877_length_614_cov_0.467427_g684_i0_6_1	16834	7.33E-04	Neurotransmitter-gated ion-channel transmembrane pore
NODE_914_length_601_cov_10.9867_g715_i0_1_1	9-189	2.43E-26	TrmB-like
NODE_621_length_705_cov_0.92766_g466_i0_4_1	20302	1.98E-15	Extended AAA-ATPase domain
NODE_621_length_705_cov_0.92766_g466_i0_4_2	18-91	3.45E-04	Extended AAA-ATPase domain
NODE_631_length_700_cov_4.50571_g474_i0_2_4	27791	6.93E-20	UDP N-acetylglucosamine acyltransferase
NODE_631_length_700_cov_4.50571_g474_i0_3_1	8-141	1.05E-26	FabZ-like
NODE_632_length_700_cov_2.6_g475_i0_2_2	31-142	3.21E-20	Phosphate binding protein-like
NODE_693_length_676_cov_3.63757_g527_i0_5_3	25-84	2.39E-09	Thioltransferase
NODE_693_length_676_cov_3.63757_g527_i0_6_1	5-125	9.39E-19	Glutathione peroxidase-like

\*Reads aligning to PhiX and avian Coronavirus Infectious Bronchitis Virus M41-CK (IBV) were present in the final datasets. These were attributed to the positive sequencing control (PhiX) and a local contamination by a lab adapted strain of IBV, respectively (the presence of IBV in the original samples was excluded by RT-qPCR, as explained in the methods).

**Appendix Table 5.** Summary of the most relevant results obtained by random primed cDNA shotgun sequencing

Organism	Sample 2	Sample 5	Sample 8	Sample 11
Total number reads	411,640	376,210	372,387	354,958
Number high quality reads	398,854	365,417	360,828	344,063
Number classified reads	397,078	363,897	359,202	342,569
Number unclassified reads	1,776	1,520	1,626	1,494
Number host reads	64,618	4,770	3,414	4,784
Percentage host reads	16.2	1.3	0.9	1.4
Number Pasteurellaceae reads	317,009	345,893	339,484	324,770
Percentage Pasteurellaceae reads	79.5	94.7	94.1	94.4
Number <i>P. multocida</i> reads	310,837	341,845	334,221	319,905
Percentage <i>P. multocida</i> reads	77.9	93.5	92.6	93.0

**Appendix Table 6.** Total species-level taxonomic categories identified. The table shows the top 8 of 68 classifications

Classification	Number of Reads	% Total Reads
<i>Pasteurella multocida</i>	6,907	48.32%
Unclassified at Species level	4,990	34.91%
Pasteurellaceae	1,536	10.75%
<i>Pasteurella pneumotropica</i>	580	4.06%
<i>Mannheimia caviae</i>	78	0.55%
<i>Serratia entomophila</i>	17	0.12%
<i>Bacillus horneckiae</i>	16	0.11%
<i>Vagococcus teuberi</i>	13	0.09%