Whole genome sequencing and molecular epidemiology of the 2021 African swine fever virus outbreak in the Dominican Republic

Steven M. Lakin[1], Vivian O'Donnell[1], Lizhe Xu[1], Roger W. Barrette[1], Jamie Barnabei[1], Rafael Nunez[2], Robin Holland[1], Bonto Faburay[1], and Suelee Robbe-Austerman[3]

[1]Foreign Animal Disease Diagnostic Laboratory National Veterinary Services Laboratories Veterinary Services Animal and Plant Health Inspection Service US Department of Agriculture Plum Island Animal Disease Center Plum Island New York USA
[2]Dirección General de Ganadería Santo Domingo Dominican Republic
[3]USDA National Veterinary Services Laboratories

July 27, 2022

Abstract

African swine fever (ASF) is a high-consequence transboundary disease of domestic and wild swine often characterized by high case mortality rates. On July 29, 2021, the Dominican Republic announced the African swine fever virus (ASFV) had been detected in samples collected in early- to mid-July 2021. Retrospective testing of samples collected as part of a collaborative surveillance project between the United States and the Dominican Republic identified ASFV in samples collected as early as May 13, 2021. These detections represent a new outbreak of ASF in the Dominican Republic, which had been declared free of the disease since 1981. Overall, 73 whole genomes of ASFV were sequenced from clinical samples received during the outbreak across 18 provinces. The genomic sequence data have been deposited in public databases to support and expand global data sharing on this impactful disease. While the sequences show a high degree of nucleotide identity to publicly available ASFV genomes from Europe and Asia, they are genetically distant from genomes in the public repositories by at least 8 previously undescribed single nucleotide polymorphisms.

Whole genome sequencing and molecular epidemiology of the 2021 African swine fever virus outbreak in the Dominican Republic

Steven M. Lakin[1], Vivian K. O'Donnell[1], Lizhe Xu[1], Roger W. Barrette[1], Jamie Barnabei[1], Rafael Núñez[2], Robin Holland[1], Bonto Faburay[1], Suelee Robbe-Austerman[3]

[1] Foreign Animal Disease Diagnostic Laboratory, National Veterinary Services Laboratories, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Plum Island Animal Disease Center, Plum Island, New York, USA.

[2] Dirección General de Ganadería, Santo Domingo, Dominican Republic.

[3] National Veterinary Services Laboratories, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Ames, Iowa, USA.

Running Head: Genomic characterization of African swine fever virus outbreak in the Dominican Republic

Correspondence

Steven Lakin , Foreign Animal Disease Diagnostic Laboratory, National Veterinary Services Laboratories, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Plum Island Animal Disease Center, Plum Island, New York, USA. Steven.Lakin@usda.gov

Abstract

African swine fever (ASF) is a high-consequence transboundary disease of domestic and wild swine often characterized by high case mortality rates. On July 29, 2021, the Dominican Republic announced the African swine fever virus (ASFV) had been detected in samples collected in early- to mid-July 2021. Retrospective testing of samples collected as part of a collaborative surveillance project between the United States and the Dominican Republic identified ASFV in samples collected as early as May 13, 2021. These detections represent a new outbreak of ASF in the Dominican Republic, which had been declared free of the disease since 1981. Overall, 73 whole genomes of ASFV were sequenced from clinical samples received during the outbreak across 18 provinces. The genomic sequence data have been deposited in public databases to support and expand global data sharing on this impactful disease. While the sequences show a high degree of nucleotide identity to publicly available ASFV genomes from Europe and Asia, they are genetically distant from genomes in the public repositories by at least 8 previously undescribed single nucleotide polymorphisms.

Keywords

African swine fever virus, pigs, disease outbreak, genome sequencing, epidemiology

1. Introduction

African swine fever (ASF) is a high-consequence transboundary disease of domestic and wild swine that typically presents as a fatal hemorrhagic fever, often with high case mortality in domestic populations. Due to its potential high mortality rate, ASF has serious implications for international trade, the socio-economic welfare of swine producers, and the health and management of domestic and wild swine populations. African swine fever virus (ASFV) is the causative agent of ASF and is an enveloped, double-stranded DNA virus with a genome size of approximately 175,000-195,000 base pairs. Over twenty genotypes have been described in the last 70 years based on characterization of the p72 (*B646L*) gene C-terminal region (Bastos et al., 2003).

ASF was first discovered in Kenya in 1921 (Sánchez-Vizcaíno et al., 2019). The first international outbreak of ASF occurred in 1957, affecting Portugal, and a second outbreak occurred in 1960, which impacted Europe, Brazil, Cuba, the Dominican Republic, and Haiti for the next several decades (Sánchez-Vizcaíno et al., 2019). Following these outbreaks, ASF was eradicated in the affected countries outside of Africa, except on the island of Sardinia, Italy (Mur et al., 2016). A third international outbreak of ASFV occurred in Georgia during June 2007, involving a p72 genotype II strain. Since 2007, p72 genotype II ASFV strains have been the most widely circulating ASFV isolates globally (Dixon et al., 2019). The 2007 Georgia outbreak affected Russia and subsequently Eastern and Central Europe in 2012-2018. China reported its first outbreak in 2018, with subsequent outbreaks reported in Southeast Asia. Here, we describe a recently identified outbreak of a p72 genotype II ASFV strain in the Dominican Republic (DR) first detected in samples dating to May 2021.

In the current investigation of the DR ASF outbreak, the National Veterinary Services Laboratories' Foreign Animal Disease Diagnostic Laboratory (USDA NVSL FADDL) initially received a batch of samples on July 26, 2021. The samples were obtained through a surveillance project initiated in 2018 between the United States Department of Agriculture (USDA) and the Dirección General de Ganadería (DIGEGA) to support early detection of ASF through surveillance for ASF and classical swine fever (CSF) in the DR. The initial shipment received at USDA NVSL FADDL included specimens collected from February through July 2021. PCR testing at USDA NVSL FADDL confirmed several samples positive for ASFV, the earliest of which was collected on May 13, 2021; and on July 29, 2021, the Dominican Republic reported ASFV detections and an outbreak in four provinces. To investigate the molecular epidemiology and characterize the genetic diversity of ASFV in the DR, the PCR-positive samples were subjected to whole genome sequencing. USDA NVSL FADDL continued to sequence samples collected by the DR through October 2021. This report details the analysis of the whole genome sequencing data and the molecular epidemiology of the DR 2021 ASFV outbreak that had been previously reported (Gonzales et al., 2021).

2. Materials and Methods

2

2.1 Samples

Samples were initially collected and metadata recorded by DR response officials. Samples were sent to USDA NVSL FADDL for ASFV PCR testing. Of the PCR-positive samples submitted to USDA NVSL FADDL, 119 tissue and whole blood samples were characterized by whole genome sequencing, resulting in 73 complete genomes.

2.2 DNA extraction

Nucleic acid extraction was performed using two commercial kits: MagMAX Pathogen RNA/DNA Kit (ThermoFisher) and Nanobind CBB Big DNA Kit (Circulomics). 200 microliters (μl) of whole blood or homogenized tissue sample were extracted following the manufacturer's instructions. The extracted DNA from the MagMax kit was sequenced on the Illumina MiSeq or Oxford Nanopore Technologies GridION platforms, while the high molecular weight DNA extracted with Circulomics kit was sequenced on the Nanopore PromethION platform.

2.3 Nanopore sequencing

On the GridION platform, sequencing libraries were constructed using the Oxford Nanopore Technologies Rapid Barcoding Sequencing kit (SQK-RBK004) and Rapid PCR Barcoding Kit (SQK-RPB004). On the PromethION platform, sequencing libraries were prepared using the Ligation Sequencing kit (SQK-LSK109) when run individually and combined with the Native Barcoding Expansion kit (EXP-NBD104 or EXP-NBD114) when multiplexed. All sequencing libraries were prepared according to the manufacturer's protocols. Libraries were quantified and samples were either individually run or multiplexed (up to 4 samples) on MinION SpotON flow cells (R9.4.1 FLO-MIN106) when run on GridION or PromethION flow cells (FLO-PRO002) when run on PromethION. GridION runs utilized host depletion via adaptive sampling available in the MinKNOW software, which removed host reads with respect to the *Sus scrofa* genome (Sscrofa 11.1, accession GCF_000003025.6).

2.4 Illumina sequencing

For sequencing on the Illumina MiSeq platform, libraries were constructed using the Nextera XT library preparation kit (Illumina), following the manufacturer's protocol. The final libraries were multiplexed and sequenced on the MiSeq System with a 500-cycle v2 sequencing kit (Illumina).

2.5 Bioinformatics analysis

Samples were sequenced on the Oxford Nanopore GridION and PromethION instruments until a sufficient average depth of coverage (minimum 8x, with >20x preferred) was reached for variant calling. For the few samples that did not reach this coverage threshold, individual review was performed; and samples were included in the analysis if informative variant calls were present that could be manually confirmed by inspection of the alignment files. Nanopore data were aligned to the ASFV Georgia 2007/1 reference genome (GenBank accession NC_044959.2) using Minimap2 (v2.18-r1015) with the options "-N 1000 -a –eqx -x map-ont" (Li, 2018). Illumina data were aligned to the same reference genome using the Burrows-Wheeler Aligner (v0.7.17) with options "-a -h 2 -Y -M" (Li & Durbin, 2012). Insertions and deletions were called for the subset of samples characterized with Illumina data using Freebayes parallel (v1.3.4) with the option "–standard-filters" (Garrison & Marth, 2012). SNPs for the epidemiological analysis were called using a custom, open-source SNP caller (https://github.com/lakinsm/simple-snp). Variants were required to meet the following thresholds to be considered a true variant: a minimum depth of 10 observed alleles at a given genomic location across the population of samples (DP > 10), a minimum observed alternate allele count of 7 at a given genomic location across the population of samples (AO > 7), and an alternative allele frequency greater than or equal to 70% at a given site within a given sample. Additionally, all single nucleotide polymorphisms described in the data were visually verified to be present in the alignment files by a subject matter expert, and final variant calls were manually corrected to match visual inspection if necessary. Low-quality SNPs located in the 5,000 base pairs flanking the 5' and 3' terminal regions of the genome were not included in the analysis.

All publicly available raw data labelled as African Swine Fever Virus whole genome sequence were downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA). Genome assemblies labelled as African Swine Fever Virus were downloaded from the NCBI GenBank repository. Genotype II ASFV sequences were selected from the NCBI SRA and GenBank data for comparison against samples from the DR. The selected SRA and GenBank data were evaluated for quality. Sequences that were of questionable quality based on the locations of mutations and degree of relatedness via comparison using multiple pairwise alignment were removed. A total of 54 ASFV genomes from public databases were included in the final analysis (Supplementary File 1) (Farlow et al., 2018; Gallardo et al., 2015; Olesen et al., 2009; Kovalenko et al., 2019; Mazur-Panasiuk et al., 2020; Gilliaux et al. 2018; Xuexia et al., 2019; Olasz et al., 2019; Hakizimana et al., 2021; Mazloum et al., 2021; Jia et al., 2020; Xiong et al., 2019).

NCBI SRA raw data retrieved from NCBI was aligned to the ASFV Georgia 2007/1 reference genome (GenBank accession NC_044959.2) using either the Burrows-Wheeler Aligner (v0.7.17, short-read data) or Minimap2 (v2.18-r1015, long-read data) and variant-called as described above. Consensus sequences including the SNP variants were produced for all DR samples and external NCBI SRA data. The resulting consensus sequences were multiple pairwise aligned against the whole genome sequences from NCBI GenBank using MAFFT (v7.487) (Katoh et al., 2002). Phylogenetic tree construction was performed using RAxML (v8.2.12) with the GTRGAMMA model argument (as determined by model selection using likelihood maximization) and visualized using FigTree (v1.4.4) (Kozlov et al., 2019). A subset of 45 nodes was selected using the Treemmer software (v0.3) to display on the phylogenetic tree in Figure 2 (Menardo et al., 2018). SNP tables were visualized using the vSNP pipeline developed by the USDA.

2.6 Epidemiological information

To characterize the molecular epidemiology and better understand the ASF outbreak in DR, observational data and information regarding the disease situation, the spectrum of ASF clinical manifestations, morbidity and mortality rates, and occurrence of the disease in the different pig production systems (backyard and commercial), were collected by DR officials using a passive surveillance strategy.

3. Results

3.1 Field epidemiological observations

During the OIE reporting period of January to December 2021, 223 outbreaks were detected in the DR for which there were 16,373 susceptible animals, 9,617 cases, 8,336 "killed or disposed of", and 7,449 deaths. Premises experiencing outbreaks represented a variety of production systems, including commercial and backyard smallholder farms. The population on each premises ranged between 1 and 1,700 pigs. While insufficient data were available for a detailed epidemiological analysis, morbidity and mortality were able to be calculated for this OIE reporting period. Overall morbidity was 58.7%, mortality 45.5%, and case mortality 77.5%; morbidity on premises varied widely from 1-100%. We note that these metrics were calculated from passive surveillance data and may be biased due to the lack of a statistically sound sampling strategy.

Several farms from which samples were positive by PCR between May and July had reported high mortality rates to DR officials. However, field observations included a variety of clinical presentations, ranging from the typical peracute presentation of Georgia 2007/1 ASFV to subclinical and chronic presentations, more indicative of an attenuated ASFV strain. In acute cases, pigs were described as presenting with some combination of ataxia, anorexia, nasal and ocular discharge, vomiting, abortion, cyanosis, pyrexia, lethargy, diarrhea, and diffuse dermal changes typical of hemorrhagic fevers. Swine with subacute presentations reportedly exhibited non-specific signs, such as depression, weight loss, and anorexia, and either died after a long period of time or survived at a higher rate than is typical for virulent strains of ASFV (Gonzales et al., 2021). A small number of infected animals presented with abortion as the only symptom or were asymptomatic at the time of sampling.

3.2 ASFV genome characterization from the Dominican Republic 2021 outbreak

As of November 2021, 73 ASFV samples from the 2021 outbreak in the DR have been characterized by

whole genome sequencing on the GridION and PromethION platforms (Oxford Nanopore Technologies) (Supplementary File 1). These 73 samples were obtained from 18 provinces in the DR (Figure 1). All 73 samples contained ASFV sequences that were characterized by p72 as genotype II and appeared to share a high degree of genetic similarity with European strains isolated from 2016-2018 (Figure 2). All sequences from the DR contained four single nucleotide polymorphisms (SNPs) relative to the Georgia 2007/1 genome (NC_044959.2) at positions NC_044959.2:7059 (C->T, MGF 110-1L Trp197Leu), NC_044959.2:44576 (A->G, MGF 505-9R Lys323Glu), NC_044959.2:134514 (T->C, NP419L Asn414Thr), and NC_044959.2:170862 (T->A, I267L Ile195Ile) (Supplementary File 2). Additionally, sequences from the DR did not contain a SNP at position NC_044959.2:26425 (T->C, MGF 360-10L Asn329Thr) that was characteristic of the Asiatic genomes.

Although the DR sequences were most closely related by sequence identity to European ASFV sequences, they appeared to have diverged by some distance from the publicly described European ASFV genomes. The DR sequences did not contain additional SNPs described in European ASFV sequences since 2016 (other than the 4 indicated above), and sequences from the DR contained at least 8 distinct SNPs that were not found in any publicly available ASFV sequence (Supplementary File 2). Additionally, the 73 sequences from the DR clustered into two genetically distinct groups that were differentiated by the presence of a SNP at position NC_044959.2:90280 (G->A, C962R Glu73Glu) (Supplementary File 3). We refer to the two distinct groups within the DR sequences as genetic cluster 1, containing the variant allele at NC_044959.2:90280, and genetic cluster 2, containing the reference allele at position NC_044959.2:90280.

Genetic cluster 1 was the largest cluster (68/73 samples), spanning 17/18 provinces characterized by whole genome sequencing from May 14 to October 4, 2021 (Figure 1). Genetic cluster 2 (5/73 samples) was characterized only in Santiago and Elías Piña provinces early in the outbreak from May 13 to July 20 and had not been detected since in the DR by USDA NVSL FADDL. The two genetic clusters differed by at least 5 SNPs but shared the common genetic backbone present in all sequences obtained from the DR. This suggests that these two clusters were genetically distinct from one another by descent but shared a common ancestor. By the time the first genomes were sequenced from the DR dating to May 13 and May 14, 2021, 1 SNP in cluster 1 and 4 SNPs in cluster 2 had developed away from the putative common ancestral genome. The putative common ancestor was not sequenced in samples obtained from the outbreak.

To verify the accuracy of the variant calls from Nanopore sequencing data, a subset of 14 samples from genetic cluster 1 were additionally sequenced on an Illumina MiSeq platform (Supplementary File 1). All 11 SNPs defining the mutational backbone from the DR sequences plus the defining SNP for genetic cluster 1 were characterized identically between Oxford Nanopore and Illumina platforms. For the samples characterized by both platforms, there were no discrepancies between any SNP calls, and Illumina sequencing did not identify additional SNPs beyond those identified in the Nanopore sequencing data. Based on these results, the variant calls seen in the DR were considered verified by two independent sequencing platforms, and all characterization for additional samples was performed using the Oxford Nanopore GridION and PromethION platforms.

Additionally, within the subset of genetic cluster 1 samples characterized by high accuracy Illumina data, several insertions and deletions were identified in one or more ASFV genomes from the DR (Table 1). Sequence data from this work has been deposited into the NCBI SRA database under BioProject accession PRJNA768333.

4. Discussion

The primary goals of viral genome characterization during the 2021 ASFV outbreak in the Dominican Republic were to describe geographical and local transmission and to track virus variation over time to inform disease control and future surveillance efforts. We successfully sequenced 73 whole ASFV genomes and characterized their genetic variations within 24-72 hours of receipt using Oxford Nanopore and Illumina sequencing technologies. We anticipate that the genomic data derived from this study, in conjunction with other epidemiological data, including cases reported to the World Organization for Animal Health (OIE),

will provide a clearer picture of ASFV transmission and geographical spread during the 2021 DR outbreak.

In addition to resolving patterns of local viral transmission, whole genome sequencing has been used to place an outbreak in the context of global viral isolates for virus tracing. The DR sequences shared 4 ancestral SNPs with European and Asiatic isolates but lacked a SNP present in the Asiatic isolates, suggesting that the DR sequences were most closely related to historically characterized European isolates (Supplementary File 2). These 4 ancestral SNPs appeared to have been acquired sometime between the Georgia 2007 outbreak and the subsequent outbreaks described in Europe and Asia from 2016-2018, according to when they were first described in public ASFV genomes. This, together with the fact that sequences from the DR did not contain a particular SNP that was characteristic of the Asiatic genomes, suggest that the DR isolates were more closely related to Eastern European ASFV sequences from 2016-2018 than those from the 2018 Asia outbreak, which was reflected in the phylogenetic tree (Figure 2).

However, the DR sequences contained at least 8 SNPs that had not been publicly described in other ASFV genomes; this suggests that the DR sequences were divergent by some time from the most recently characterized European isolates. Since no genomes were available with intermediate variations to link the public isolates to the DR sequences, it was difficult to accurately determine a divergence time or rate. It may be that the 8 or more SNPs present in the DR genomes are also present in ASFV genomes circulating elsewhere and have yet to be described in the public databases. Determining where the most recent common ancestor of the DR sequences was circulating prior to its introduction into the DR will require increased ASFV surveillance and whole or targeted genome characterization efforts.

Furthermore, our analysis revealed some insertions and deletions in the genome of the DR ASFV isolates (Table 1). It was previously established that some insertions and deletions have been associated with attenuation of virulence in ASFV Georgia 2007 strains (Li et al., 2021). Such insertions and deletions may provide a plausible explanation for the reduced virulence and subacute clinical manifestations of ASF observed among domestic pig populations during the DR outbreak. However, there was insufficient information obtained from this outbreak to correlate specific insertions and deletions to ASFV virulence, and further studies will be required to establish such a correlation.

Considering the field observations of reduced virulence and subacute clinical manifestation of ASF in the DR outbreak, it is possible that ASFV was not detected by initial surveillance efforts and has been circulating in this region for a longer time than anticipated. If so, this demonstrates that reliance on detection of increased swine mortality alone may not be a sufficient indicator of an ASFV outbreak or introduction. Serology and molecular detection may be valuable surveillance tools to use in conjunction with epidemiological measures for continued surveillance of ASFV in the DR and neighboring regions.

Within the ASFV genomes from the DR, we described two distinct genetic clusters that were divergent from one another by at least 5 SNPs (Supplementary File 3). Genetic cluster 1 contained the majority of ASFV genomes sequenced from the DR (68/73 genomes). Based on phylogenetic results, we hypothesized that the prototypical sequence from genetic cluster 1 was the strain that initially spread across the DR provinces. This "prototypical" sequence contained only the 11 SNP backbone characteristic of all DR sequences plus the single SNP defining genetic cluster 1 at position NC_044959.2:90280 G->A (e.g. strain DR/Duarte/2021/9682773, Supplementary File 3). Of the 73 genomes sequenced, 35 had this prototypical genetic profile from cluster 1 and were identified across 12 of the 18 provinces with dates ranging from July 5 to September 30, 2021.

The remainder of the genetic cluster 1 genomes (33/73 sequences) contained SNPs in addition to the prototypical cluster 1 sequence, and the additional SNPs groupings were more often found within the same province. This accumulation of mutations by geographic region was suggestive of localized transmission and evolutionary adaptation, at least within the provinces from which they were sequenced. It is possible that reduced animal movement following the initial ASFV detection resulted in more localized transmission patterns and therefore SNP profiles clustered by geographic region.

Additional data are needed to determine the origin and emergence of these distinct ASFV genetic clusters in the DR; however, sufficient time must have passed for these genomes to diverge from those characterized

6

in public genome databases. Whether this variation occurred within the DR or elsewhere over a longer timeframe with subsequent introduction into the DR is currently unclear given the genomic data available. Sequencing of additional samples from the DR and a thorough epidemiological investigation could help address this question.

Conflict of Interest

The authors declare no conflicts of interest.

Ethics Statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required for this study, as samples were taken by public authorities as a part of an official outbreak response in accordance with the laws and authority of the Dominican Republic.

Data Availability

Data from the 2021 ASFV outbreak in the DR will be linked to NCBI BioProject accession PRJNA768333. BioSample accessions for each DR sample are listed in Supplementary File 1. Raw Nanopore and Illumina FASTQ files are published in NCBI SRA under the above BioProject.

Supplementary File 1: Data for Dominican Republic and external sequences used in this analysis

Supplementary File 2: SNP tables including external ASFV Genotype II sequences from GenBank

Supplementary File 3: SNP tables for the Dominican Republic sequences

Hosted file

SupplementaryFile1_SampleMetadata.xlsx available at https://authorea.com/users/496323/articles/577762-whole-genome-sequencing-and-molecular-epidemiology-of-the-2021-african-swine-fever-virus-outbreak-in-the-dominican-republic

Hosted file

SupplementaryFile2_GenotypeII_SNP_Table.xlsx available at https://authorea.com/users/496323/articles/577762-whole-genome-sequencing-and-molecular-epidemiology-of-the-2021-african-swine-fever-virus-outbreak-in-the-dominican-republic

Hosted file

SupplementaryFile3_DR_Sample_SNP_Table.xlsx available at https://authorea.com/users/496323/articles/577762-whole-genome-sequencing-and-molecular-epidemiology-of-the-2021-african-swine-fever-virus-outbreak-in-the-dominican-republic

Acknowledgements

References

Bastos, A., Penrith, M.L., Crucière, C., et al. (2003). Genotyping field strains of African swine fever virus by partial p72 gene characterisation. *Arch Virol* , 148, 693–706.*doi* :10.1007/s00705-002-0946-8

Dixon, L.K., Sun, H., Roberts, H. (2019). African swine fever.*Antiviral Res* , 165, 34-41.*doi* :10.1016/j.antiviral.2019.02.018

Farlow, J., Donduashvili, M., Kokhreidze, M., Kotorashvili, A., Vepkhvadze, N.G., Kotaria, N., Gulbani, A. (2018). Intra-epidemic genome variation in highly pathogenic African swine fever virus (ASFV) from the country of Georgia. *Virol J* , 15, 190.*doi* :10.1186/s12985-018-1099-z

Gallardo, C., Soler, A., Nieto, R., Cano, C., Pelayo, V., Sanchez, M.A., Pridotkas, G. (2015). Experimental Infection of Domestic Pigs with African Swine Fever Virus Lithuania 2014 Genotype II Field Isolate.*Transbound Emerg Dis* , 64, 300-304. *doi* :10.1111/tbed.12346

Garrison, E., Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv* Preprint, 1207.3907.

Gilliaux, G., Garigliany, M., Licoppe, A., Paternostre, J., Lesenfants, C., Linden, A., Desmecht, D. (2019). Newly emerged African swine fever virus strain Belgium/Etalle/wb/2018: Complete genomic sequence and comparative analysis with reference p72 genotype II strains.*Transbound Emerg Dis* , 66, 2566-2591.*doi* :10.1111/tbed.13302

Gonzales, W., Moreno, C., Duran, U., Henao, N., Bencosme, M., Lora, P., Reyes, R., Nunez, R., De Gracia, A., Perez, A.M. (2021). African swine fever in the Dominican Republic. *Transbound Emerg Dis* , 1-2.*doi* :10.1111/tbed.14341

Hakizimana, J.N., Ntirandekura, J.B., Yona, C., Nyabongo, L., Kamwendo, G., Chulu, J.L.C., Ntakirutimana, D., Kamana, O., Nauwynck, H., Misinzo, G. (2021). Complete genome analysis of African swine fever virus responsible for outbreaks in domestic pigs in 2018 in Burundi and 2019 in Malawi. *Trop Anim Health Prod* , 53, 438.*doi* :10.1007/s11250-021-02877-y

Jia, L., Jiang, M., Wu, K., Hu, J., Wang, Y., Quan, W., Hao, M., Liu, H., Wei, H., Fan, W., Liu, W., Hu, R., Wang, D., Li, J., Chen, J., Liu, D. (2020). Nanopore sequencing of African swine fever virus. *Sci China Life Sci* , 63, 160-164. *doi* :10.1007/s11427-019-9828-1

Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* , 30, 3059-3066.*doi* :10.1093/nar/gkf436

Kovalenko, G., Ducluzeau, A., Ishchenko, L., Sushko, M., Sapachova, M., Rudova, N., Solodiankin, O., Gerilovych, A., Dagdag, R., Redlinger, M., Bezymennyi, M., Frant, M., Lange, C.E., Dubchak, I., Mezhenskyi, A.A., Nychyk, S., Bortz, E., Drown, D.M. (2019). Complete Genome Sequence of a Virulent African Swine Fever Virus from and Domestic Pig in Ukraine.*Microbiol Resour Announc* , 8, e00883-19.*doi* :10.1128/MRA.00883-19

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* , 35, 4453-4455.*doi* :10.1093/bioinformatics/btz305

Li, D., Liu, Y., Qi, X., Wen, Y., Li, P., Ma, Z., Liu, Y., Zheng, H., Liu, Z. (2021). African Swine Fever Virus MGF-110-9L-deficient Mutant Has Attenuated Virulence in Pigs. *Virol Sin* , 36, 187-195.*doi* :10.1007/s12250-021-00350-6

Li, H., Durbin, R. (2012). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences.*Bioinformatics* , 34, 3094-3100.*doi* :10.1093/bioinformatics/bty191

Mazloum, A., van Schalkwyk, A., Shotin, A., Igolkin, A., Shevchenko, I., Gruzdev, K.N., Vlasova, N. (2021). Comparative Analysis of Full Genome Sequences of African Swine Fever Virus Isolates Taken from Wild Boars in Russia in 2019. *Pathogens* , 10, 521.*doi* :10.3390/pathogens10050521

Mazur-Panasiuk, N., Walczak, M., Juszkiewicz, M., Wozniakowski, G. (2020). The Spillover of African Swine Fever in Western Poland Revealed Its Estimated Origin on the Basis of O174L, K145R, MGF 505-5R and

8

IGR I73R/I329L Genomic Sequences. *Viruses* , 12, 1094.*doi* :10.3390/v12101094

Menardo, F., Loiseau, C., Brites, D., Coscolla, M., Gygli, S.M., Rutaihwa, L.K., Trauner, A., Beisel, C., Borrell, S., Gagneux, S. (2018). Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics* , 19, 164.*doi* : https://doi.org/10.1186/s12859-018-2164-8

Mur, L., Atzeni, M., Martínez-López, B., Feliziani, F., Rolesu, S., Sanchez-Vizcaino, J.M. (2016). Thirty-Five-Year Presence of African Swine Fever in Sardinia: History, Evolution and Risk Factors for Disease Maintenance. *Transbound Emerg Dis* , 63, e165-e177.*doi* :10.1111/tbed.12264

Olasz, F., Meszaros, I., Marton, S., Kajan, G.L., Tamas, V., Locsmandi, G., Magyar, T., Balint, A., Banyai, K., Zadori, Z. (2019). A Simple Method for Sample Preparation to Facilitate Efficient Whole-Genome Sequencing of African Swine Fever Virus. *Viruses* , 11, 1129.*doi* :10.3390/v11121129

Olesen, A.S., Lohse, L., Dalgaard, M.D., Wozniakowski, G., Belsham, G.J., Botner, A., Rasmussen, T.B. (2009). Complete genome sequence of an African swine fever virus (ASFV POL/2015/Podlaskie) determined directly from pig erythrocyte-associated nucleic acid. *J Virol Methods* , 261, 14-16. *doi* :10.1093/bioinformatics/btp324

Sánchez-Vizcaíno, J.M., Laddomada, A., Arias, M.L. (2019). African Swine Fever Virus. In *Diseases of Swine* (eds J.J. Zimmerman, L.A. Karriker, A. Ramirez, K.J. Schwartz, G.W. Stevenson and J. Zhang).*doi* :10.1002/9781119350927.ch25

Xiong, D., Zhang, X., Yu, J., Wei, H. (2019). Rapid phylogenetic analysis of African swine fever virus from metagenomic sequences.*Virus Res* , 297, 198357. *doi* :10.1016/j.virusres.2021.198357

Xuexia, W., Xijun, H., Xiang, Z., Xianfeng, Z., Liling, L., Yuntao, G., Ying, Z., Zhigao, B. (2019). Genome sequences derived from pig and dried blood pig feed samples provide important insights into the transmission of African swine fever virus in China in 2018. *Emerg Microbes Infect* , 8, 303-306. *doi* :10.1080/22221751.2019.1565915

Table 1. ASFV sequences obtained from the DR contain several insertion and deletions relative to the Georgia 2007/1 genome.

Genomic Location	Type[1]	Length (bp)	Gene
NC_044959.2:1382	Insertion	1	Non-coding
NC_044959.2:1382	Deletion	1	Non-coding
NC_044959.2:4069	Insertion	1	Non-coding
NC_044959.2:6775	Deletion	1	Non-coding
NC_044959.2:10618	Insertion	1	MGF 110-7L
NC_044959.2:14225	Deletion	3-5	MGF 110-10-L/MGF110-14L fusion (Poly C region)
NC_044959.2:14626	Insertion	1	ASFV G ACD 00240
NC_044959.2:17838	Insertion	1	Non-coding
NC_044959.2:17845	Insertion	1	Non-coding
NC_044959.2:19792	Deletion	1	Non-coding
NC_044959.2:19993	Deletion	7	ASFV G ACD 00350 (Poly G region)
NC_044959.2:44471	Deletion	4	MGF 505-9R
NC_044959.2:73263	Deletion	1	Non-coding
NC_044959.2:129718	Deletion	1	O174L
NC_044959.2:155267	Deletion	1	H108R
NC_044959.2:170445	Deletion	1	Non-coding
NC_044959.2:171562	Deletion	1	Non-coding
NC_044959.2:173382	Insertion	10	Non-coding
NC_044959.2:188972	Deletion	1	MGF 360-21R

[1]These insertions and deletions were characterized in at least one sample using an Illumina sequencing platform. Insertions and deletions in the multi-gene family (MGF) genes have been previously associated with virus attenuation, however there were insufficient data in this study to correlate these insertions and deletions to virulence (Li et al., 2021).
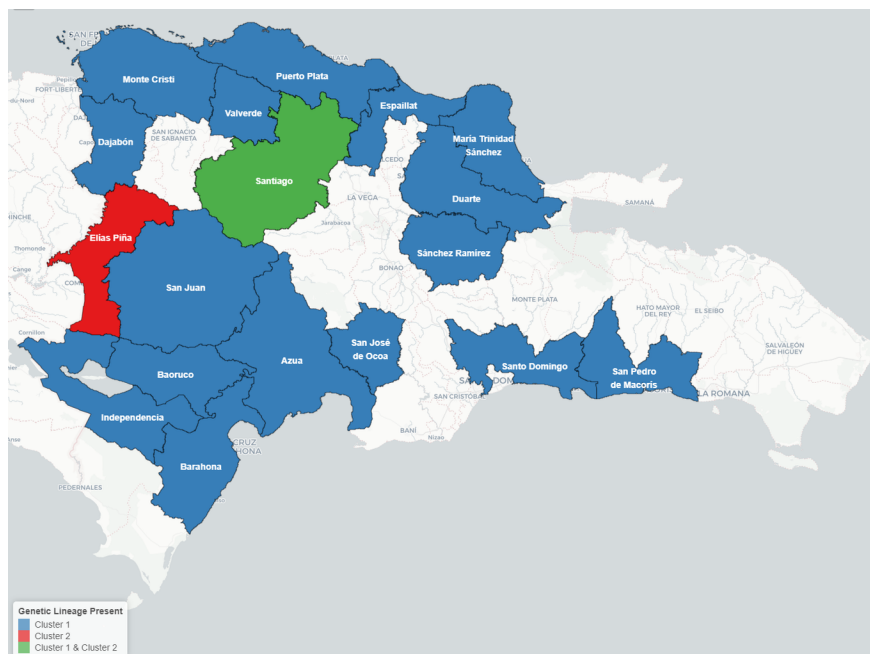

Figure 1. Geospatial distribution of the ASFV genetic clusters from the 2021 outbreak in the Dominican Republic. Genetic cluster 1 (blue) was the most widespread (17/18 provinces sequenced) and most frequently sequenced (68/73 samples) genetic cluster. Genetic cluster 2 (red) was localized to Santiago and Elías Piña provinces (5/73 samples) early in the outbreak and was not subsequently detected. Santiago was the only province where both genetic clusters (green) were detected by USDA NVSL FADDL. The densest swine production region in the DR is in the North-central region near Santiago province, with smaller operations spread throughout the country.

Figure 2. Phylogenetic tree showing clustering of the Dominican Republic sequences with other ASFV p72 genotype II genomes. The sequences generally cluster by the origin of their respective outbreaks across all European, Asiatic, and DR genomes. The DR sequences group within their respective genetic clusters as previously described (blue: genetic cluster 1, red: genetic cluster 2). A subset of unique

10

sequences from the DR was selected using Treemmer for this phylogenetic tree for readability, however all samples and associated SNPs are shown in Supplementary File 2.